Understand and use the estimation of soil organic carbon persistence by Rock-Eval® thermal analysis

Eva Kanari

Sorbonne Université

École doctorale 398 Géosciences, Ressources Naturelles et Environnement

Institut des Sciences de la Terre de Paris / Paléoenvironnemements, Paléoclimats, Bassins

Understand and use the estimation of soil organic carbon persistence by Rock-Eval® thermal analysis

Par Eva Kanari

Thèse de doctorat de Biogéosciences

Dirigée par François Baudin, Pierre Barré et Lauric Cécillon

Présentée et soutenue publiquement le 25 Mars 2022

Devant un jury composé de :

Claire CHENU	Prof. AgroParisTech, DR INRAe (France)	Examinatrice
Josette GARNIER	DR CNRS (France)	Examinatrice
Daniel RASSE	Senior Researcher NIBIO (Norvège)	Rapporteur
Eric VERRECCHIA	Prof. Université de Lausanne (Suisse)	Rapporteur
Thomas EGLIN	IR ADEME (France)	Invité
François BAUDIN	Prof. Sorbonne Université (France)	Directeur de thèse
Pierre BARRÉ	CR CNRS (France)	Co-directeur de thèse
Lauric CÉCILLON	CR INRAe (France)	Co-directeur de thèse

Dédicace

To my grandfather for teaching me the joy of learning amongst so many other things

I. Foreword

Continuously rising global temperatures led to 2011-2020 being the warmest decade including the six warmest years on record, with 2020 climbing on the first place with an average global temperature of 1.2°C above the pre-industrial level. Extreme climatic phenomena, such as sustained heat waves, wildfires, droughts, floods and hurricanes, tormenting every corner of the world are only some of the most obvious events caused by climate change. Apart from the evident environmental aspect, the ongoing crisis has significant societal and economic impacts as well.

This realization has forced governments and international organizations to pay attention to climate change. Efforts are intensifying at every level, from private stakeholders and small-scale initiatives to global organizations such as the United Nation's Framework Convention on Climate Change. Important milestones towards fighting climate change include the establishment of the Paris Agreement and the launch of the United Nation's Sustainable Development Goals, but also the shift towards green consciousness observed amongst the young generation sparking initiatives such as Fridays for Future.

It is widely understood that immediate coordinated global action is imperative and that the climate is right for reforming the future.

Nature-based solutions are of immense value and are thus in the centre of this effort with the spotlight turned on soils. To the French Agency for Ecological Transition and to the ANR project StoreSoilC, the two parties funding my research during the last three years — I am grateful to have played a small part in this universal effort through our collaboration.

II. Acknowledgments

First and foremost, for their kindness and support, I would like to say the biggest thank you to my supervisors. Thank you for taking me by the hand since day one and making me part of your team. Pierre, thank you for the genuine care and support, and for your jokes. Lauric, thank you for the ideas, the inspiration, the resourcefulness. François, thank you for always being there when needed and for always backing me up. I never imagined perusing a PhD before meeting you, yet I cannot think of a better way to have spent the last 3 years. I can't wait to see where this trip will take me next, but I hope to cross paths with you again and again in the future.

I would like to thank the ANR project StoreSoilC and the ADEME for the financial support during the last three years, but also for the privilege to be part of such active and dynamic networks of the French research world.

For their willingness to participate in the final act of this project I would like to thank the members of my thesis jury.

The list goes on with all the incredibly kind and brilliant people I had the chance to meet and work with. Laure, Florence, Hugues, Bruno, Fabien, Nicolas, thank you all for sharing your knowledge and your passion for what you do.

To office-mates and lab-mates who became friends, for all the beers that were supposed to be "just one", for all the pots de thèse, the missed deadlines, the stressful weeks. Lavinia, Thomas, Arnaud, Cedric, Arefeh, Manon, Ariel, Elenora, Francesco, Martin, Amin, Sun-Ping, Juan, Louise, Camille, Dia. It was an absolute joy sharing these past 3 years with all of you.

A mes colocataires, ma famille en France, Oliver, Natachou, Claire, Remi, Elo, Milou, Enael, Adele, Sysy, Marie P, Magelan, Clems, Cindy, Fanny, Aziz, Giuliano, un grand merci.

Manon, I could thank you a million times. You know you are the one behind all this.

I would like to say a big warm thank you to my friends, my dearest Daphne, my closest Aristia, my oldest Evita and all the others whose support I have constantly felt, Savvaki mou, Maraki, meine liebe Jentel.

To my brothers, thank you for the challenge and the reason to care.

Τέλος, πάνω που ούλλους ευχαριστώ τους γονιούς μου. Για την στήριξη σας σε ότι θκιαλέω να κάμω τζαι για την αγάπη σας. Χωρίς εσάς εν θα ήμουν δαμέ σήμερα, κυριολεκτικά.

III. Acronyms

AIAL	Arvalis, INRA, Agro-Transfert-RT, LDAR
AMG	Andriulo, Mary, Guérif
C	Carbon
COP	Conference of the Parties
DOC	Dissolved organic carbon
GHG	Greenhouse gases
EA	Elemental analysis
ESM	Earth system models
FEU-US	Universal Ecological Fund (Fundación Ecológica Universal)
FID	Flame ionization detector
HC	Hydrocarbons
HI	Hydrogen index
IPCC / GIEC	Intergovernmental panel on climate change / Groupe d'experts intergouvernemental sur l'évolution du climat
IPBES	Intergovernmental science-policy Platform on Biodiversity and Ecosystem Services
IR	Infrared
LTBF	Long-term bare fallow
LTE	Long-term experiment
NBS	Nature based solutions
NET	Negative emission technologies
OI_{RE6}	Oxygen index
PC	Pyrolyzed carbon
POM	Particulate organic matter
RMQS	Réseau de mesures de la qualité des sols / French soil quality monitoring network
SOC	Soil organic carbon
SOERE-PRO	Observatoire de recherche en environnement sur les produits résiduaires organiques / French research observatory on the recycling of organic residues in agriculture
SOM	Soil organic matter
SMN	Soil monitoring network
TOC	Total organic carbon
WWF	World Wide Fund for Nature

I. Table of contents

II. List of figures

III. List of tables

IV. Résume étendu en langue française

1. Le défi climatique et le rôle des sols

Le rapport spécial du GIEC[1] (2018) sur "Le changement climatique et les terres émergées" souligne de façon explicite les conséquences causées par le changement climatique sur la Terre, ainsi que les risques associés pour la biodiversité, la santé humaine et écosystémique, et les systèmes alimentaires. Dans son dernier rapport, le GIEC (2021) prévient que seule une réaction immédiate menant à la neutralité carbone au niveau mondial avant 2050 pourrait permettre de rester sous la limite des 1.5–2 °C d'augmentation de la température globale par rapport aux niveaux pré-industriels. Pour atteindre cet objectif fixé par l'Accord de Paris (UNFCCC, 2015) afin d'éviter les conséquences sur le bien-être des humains et des écosystèmes, des technologies à émissions négatives sont nécessaires (GIEC, 2018; Anderson et al., 2019). En raison du fait que les solutions fondées sur la nature (SFN), comme la gestion durable des forêts, des zones humides et des prairies, sont des « solutions sans regret » mais aussi en raison de leur potentiel de réduction des émissions estimé à un niveau considérable de 37% de l'objectif de 2030, les scientifiques et les activistes environnementaux appellent à leur mise en œuvre extensive (Griscom et al., 2017).

[1] Groupe d'experts intergouvernemental sur l'évolution du climat

L'accent est mis sur l'importance des pratiques appropriées de gestion des sols, qui constituent l'une des solutions les plus pratiques et les plus réalisables. De nombreuses politiques nationales et internationales (par exemple, l'initiative 4pour1000 et le pacte vert pour l'Europe) s'engagent à séquestrer du C dans les sols afin d'inverser la perte historique de carbone causée par des millénaires d'appropriation et d'utilisation non durable des terres par l'homme (Sanderman et al., 2017). Des pratiques appropriées de gestion des terres pourraient ralentir la dégradation des sols et contribuer à assurer la sécurité alimentaire et à lutter contre le changement climatique. Lors de la récente COP 26, 148 pays ont soumis des contributions déterminées au niveau national nouvelles ou actualisées. Dans l'ensemble, 61% des pays ont fait référence au carbone organique des sols (COS) ou à des mesures liées au COS. Les priorités pour les mesures d'atténuation ou d'adaptation liées au SOC comprenaient la gestion des zones humides (43%), l'agroforesterie (34%) et la gestion des prairies (22% ; Rose et al., 2021). L'urgence de cette situation, mais aussi les avantages socio-économiques qui peuvent être générés par une gestion correcte des sols m'ont incité à développer mes travaux de recherche autour d'eux.

2. Les sols dans le cycle de carbone et l'importance de la matière organique du sol

Sur une profondeur de 2 mètres les sols contiennent deux fois plus de carbone (1500–2400 GtC) que l'atmosphère (860 GtC) et la végétation (600 GtC) réunies (Friedlingstein et al., 2020). La taille ainsi que la position critique de ce réservoir, à la frontière entre la terre solide et l'atmosphère, soulignent son importance pour la régulation du climat (Batjes, 1996; Lal, 2004a). Des variations faibles de la taille du réservoir du carbone organique du sol (COS) peuvent avoir des effets majeurs sur la concentration de CO_2 atmosphérique (Lal, 2004b). Cette idée a suscité le "mouvement de séquestration du COS" et, entre autres, l'initiative "4 pour 1000", selon laquelle une augmentation annuelle relativement faible de 0,4 % des stocks de COS est suffisante pour compenser l'augmentation du CO_2 atmosphérique causée par les émissions anthropiques (Minasny et al., 2017; Pellerin et al., 2019). Mais même en mettant de côté le débat sur le climat, les stocks du COS doivent être protégés et si possible augmentés, car la matière organique du sol (MOS) est essentielle pour sa santé et sa capacité à fournir de multiples services écosystémiques (Lal, 2004b; Fig.1).

En outre, la MOS est considérée comme le composant le plus important qui peut être géré efficacement pour améliorer la santé du sol et parvenir à son utilisation durable. L'augmentation de la teneur en COS est bénéfique pour de nombreuses fonctions telles que leur fertilité (Tiessen et al., 1994; Six et al., 2004), leur capacité de régulation du climat (Heimann and Reichstein, 2008), et leur potentiel de filtration et de rétention de l'eau (Doran and Parkin, 1994; Hudson, 1994).

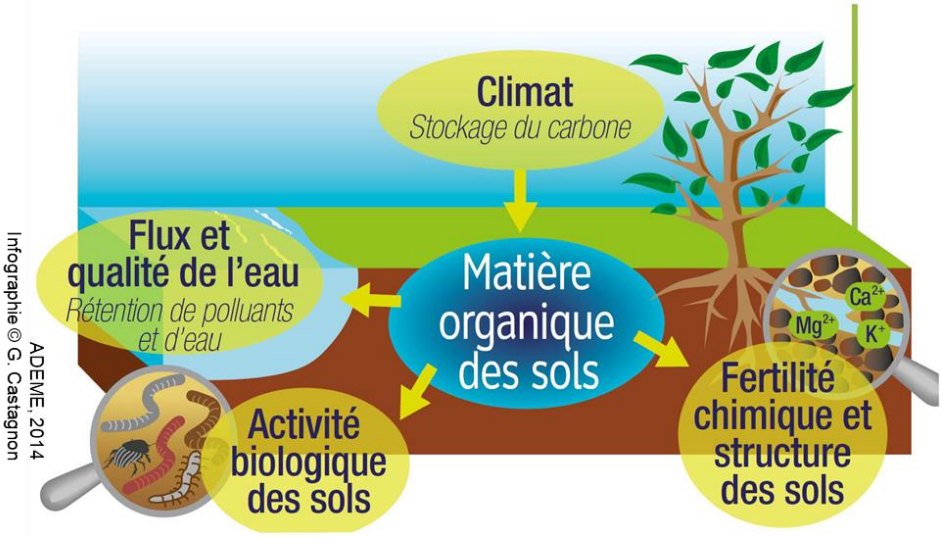


Figure 1 : Les services écosystémiques du sol qui dépendent son contenu en matière organique (source indiquée sur la figure).

Cependant, la réalisation de l'objectif noble de mieux gérer les sols et augmenter leurs stocks de carbone est actuellement entravée par notre compréhension limitée des processus contrôlant la persistance du COS et notre incapacité à prévoir son évolution.

3. Principaux processus et mécanismes contrôlant l'évolution du COS

3.1. Vue rapide des processus qui déterminent le COS

Le principal processus d'apport de carbone organique dans le sol est la fixation du CO_2 de l'atmosphère par les plantes par la photosynthèse, et le dépôt de ce carbone soit directement sous terre par les exsudats des racines, soit sous forme de résidus dans la couche de litière aérienne (FAO & ITPS, 2015). Dans le cas de sols gérées, d'autres ajouts de matière organique exogène peuvent se produire, comme l'apport de fumier ou de compost végétal.

La matière organique entre dans la voie métabolique de divers organismes hétérotrophes (verres de terre, champignons, bactéries ; Pellerin et al., 2019). Une partie du carbone est respirée et libérée dans l'atmosphère sous forme de CO_2 en quelques heures, alors que les composants plus persistants peuvent résider dans le sol pendant des centaines à des milliers d'années (Trumbore, 1997; Jenkinson and Coleman, 2008).

3.2. Mécanismes de stabilisation de la matière organique du sol

Jusqu'à récemment, trois catégories de mécanismes de stabilisation de la MOS étaient considérées : (1) augmentation de la récalcitrance de la MO par la minéralisation sélective des apports de litière plus biodégradables et la formation de macromolécules humiques stables (2) protection physique, par exemple par la formation d'agrégats, et (3) associations organo-minérales par adsorption qui limitent l'accessibilité aux enzymes (Sollins et al., 1996; Six et al., 2002; von Lützow et al., 2006).

Bien que leur importance relative ait été étudiée de manière intensive et qu'il a été démontré qu'elle varie d'un cas à l'autre, aucun des mécanismes mentionnés ci-dessus ne peut expliquer à lui seul la persistance de la MOS. Elle est plutôt considérée comme une propriété de l'écosystème (Schmidt et al., 2011) et, en tant que tel, ses associations avec les variables pédoclimatiques sont une autre façon de l'étudier.

3.3. Les facteurs de contrôle de temps de résidence de la matière organique du sol

Les facteurs qui influencent le renouvellement de la MOS ainsi que la distribution des stocks du COS comprennent la température, l'humidité du sol, le pH, la texture et la minéralogie, ainsi que le type d'apport de matière organique fraîche et la profondeur du sol (Balesdent et al., 2018; Luo et al., 2019; Luo and Viscarra-Rossel, 2020). Ces facteurs sont par essence liés aux mécanismes décrits ci-dessus (pour une description détaillée voir Pellerin et al., 2019), mais ils sont plus faciles à décrire et à inclure dans les approches de modélisation du COS.

4. Les différents types des modèles de dynamique du COS

Depuis les années 1980, une approche valorisant la simplicité mathématique et basée sur l'exemple donné par Hénin et Dupuis (1945) a conduit au développement des modèles très utilisés de la dynamique du COS. Dans ces modèles, le continuum de décomposition du COS est représenté par une série de compartiments cinétiques, caractérisés par des temps de renouvellement différents. Parmi les avantages les plus importants de ces modèles, intrinsèquement liés à leur structure simple, figurent leur transférabilité, leur efficacité de calcul et leur paramétrage adéquat. Cette catégorie comprend des modèles de renommée mondiale tels que Century (Parton et al., 1987) et RothC (Jenkinson and Rayner, 1977), et des modèles bien paramétrés, principalement utilisés au niveau national, tels que le modèle suédois ICBM (Andrén and Kätterer, 1997), le modèle français AMG (Andriulo et al., 1999) ou le modèle danois C-TOOL (Taghizadeh-Toosi et al., 2014).

Actuellement une nouvelle génération de modèles complexes, axés sur les mécanismes qui contrôlent le renouvellement des COS, est en train de gagner en popularité (Abramoff et al., 2018; Sulman et al., 2018; Waring et al., 2020; Zhang et al., 2021). Cependant, leurs structures complexes et le nombre important de paramètres inconnus à contraindre, rend ces modèles très difficiles à calibrer et à valider par rapport aux données d'observation (Sulman et al., 2018). Cela entrave particulièrement leur potentiel à être utilisés comme outils opérationnels pour prédire l'évolution du COS à grande échelle.

Les modèles multi-compartimentaux de la dynamique du COS sont la meilleure option dont nous disposons actuellement pour favoriser des actions de préservation et de séquestration du COS basées sur la science, étant donné la forte incertitude des modèles plus complexes (Cécillon, 2021a; Dangal et al., 2021; Lee et al., 2021; Shi et al., 2018; Crowther et al., 2019). Cependant, les prédictions d'évolution des stocks du COS fournies par ces modèles simples sont très sensibles à la distribution initiale du COS parmi les différents compartiments cinétiques (Luo et al., 2016; Smith and Falloon, 2000; Clivot et al., 2019). Cela fait de la question du partitionnement des compartiments cinétiques du COS une priorité pour améliorer leur précision (Luo et al., 2016; Taghizadeh-Toosi et al., 2020).

5. Les sites rares pouvant être utilisés pour estimer la persistance *in situ* du COS

Divers types d'expériences agronomiques à long terme offrent différentes possibilités d'étudier la persistance *in situ* du COS dans des conditions pédologiques, climatiques et biologiques réelles. Il s'agit notamment de chronoséquences avec un changement de végétation (C_3–C_4), de jachères nues et d'expériences agronomiques avec des apports de carbone connus et des pratiques contrôlées. Dans le premier cas, le suivi de l'abondance naturelle du ^{13}C depuis le moment du changement de végétation permet de simuler la décomposition du carbone natif idéalement sur plusieurs décennies et de calculer la taille de la fraction active du COS (Fig.2a ; Balesdent et al., 1987; Balesdent, 1991). Les sites en jachère nue de longue durée (LTBF), c'est-à-dire les sites maintenus sans végétation pendant plusieurs décennies, offrent une occasion unique d'observer la décomposition du carbone natif sans interférence de la matière organique fraîche (Fig. 2b ; Rühlmann, 1999) et de calculer la taille de la fraction stable à l'échelle du siècle (Barré et al., 2010). Dans le troisième cas, les apports de carbone connus, les données pédoclimatiques, les pratiques de gestion du sol sur de longues périodes permettent une modélisation inverse et une optimisation des conditions telles que la répartition du COS entre le compartiment « stable » et le compartiment « actif » (Fig. 2c ; Clivot et al., 2019).

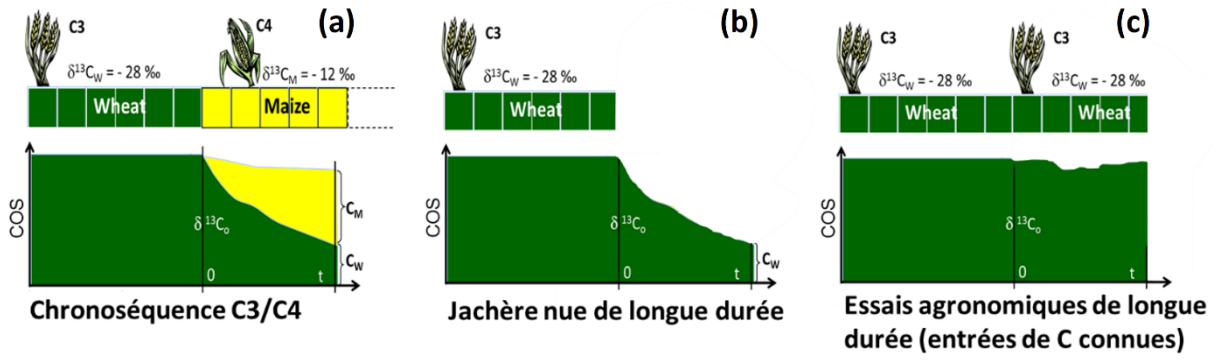


Figure 2 : Représentation schématique de l'évolution des stocks du COS théorique lors de divers types d'expériences agronomiques à long terme (Barré, 2020 ; Adapté d'après Balesdent, 1991)

A part ces sites d'expérimentation à long terme très spécifiques et très limités géographiquement, les moyens d'estimer efficacement la taille des compartiments cinétiques des modèles restent limités.

6. Méthodes actuellement utilisées pour estimer la taille des compartiments de modèles

Des procédures d'initialisation des modèles du COS ont été proposées, mettant en relation les résultats des méthodes de fractionnement du COS avec les tailles des compartiments cinétiques du COS (par exemple, Zimmermann et al. (2007) ou Skjemstad et al. (2004) pour le modèle RothC). Cependant, cette approche souffre d'importants inconvénients. Premièrement, les procédures de fractionnement du COS sont laborieuses et ne peuvent pas être mises en œuvre dans des études à grande échelle. Deuxièmement, leur reproductibilité est discutable (Poeplau et al., 2013, 2018), et troisièmement, leur utilisation pour initialiser les tailles des compartiments des modèles du COS n'a jamais été correctement validée.

Une validation correcte nécessiterait de montrer que (1) la taille des fractions du COS mesurées correspond à celle des bassins cinétiques du modèle, et que (2) les simulations de la dynamique du COS sont plus précises en utilisant cette stratégie d'initialisation, par rapport aux simulations par défaut (sur des sites de validation indépendants alors que les autres paramètres du modèle restent constants). Une correspondance raisonnablement bonne entre les fractions du COS mesurées et les pools conceptuels de COS modélisés a été rapportée dans un certain nombre d'études, avec toutefois quelques divergences notables (Zimmermann et al., 2007a; Leifeld et al., 2009a; Herbst et al., 2018).

Les études qui ont tenté d'initialiser les tailles des compartiments modèles à l'aide d'un schéma de fractionnement du COS n'ont généralement pas signalé d'amélioration de la précision des simulations de la dynamique du COS par rapport à une approche d'initialisation par défaut ou par *spin up* (Leifeld et al., 2009b; Nemo et al., 2016; Cagnarini et al., 2019). Seules deux études ont montré qu'une initialisation basée sur un schéma de fractionnement du COS donnait des simulations plus précises de la dynamique du COS observée, mais au prix d'une modification du taux de décomposition des pools cinétiques du COS (Skjemstad et al., 2004 ; Luo et al., 2014).

La question de la partition des compartiments du COS reste cruciale et sans réponse.

7. Analyse thermique et le modèle d'apprentissage automatique PARTY$_{SOC}$

Au cours de la dernière décennie, l'analyse thermique en rampe a montré un grand potentiel pour obtenir des marqueurs de la stabilité biogéochimique du COS (Plante et al., 2013; Gregorich et al., 2015; Soucémarianadin et al., 2018). Surtout lorsqu'elle est accompagnée d'informations sur la chimie globale du COS, comme l'appauvrissement ou l'enrichissement en hydrogène (Barré et al., 2016) ou en fractions spécifiques (Sanderman and Grandy, 2020), la stabilité thermique a montré une forte corrélation avec les outils classiques utilisés pour estimer l'âge des atomes de carbone (^{14}C ; Plante et al., 2013) et les estimations *in situ* de la stabilité biogéochimique du SOC (sites LTBF ; Barré et al., 2016).

En utilisant les données de cette dernière étude, y compris des échantillons de sol d'archives provenant des jachères nues de longue durée et leur caractérisation connue par analyse thermique, Cécillon et al. (2018) ont développé un modèle d'apprentissage automatique. Ce modèle utilise des paramètres thermiques obtenus par analyse Rock-Eval® et permet d'estimer la taille de la fraction persistante du COS à l'échelle du siècle dans un échantillon. Cette première version a été adaptée récemment en élargissant son ensemble de calibration pour former la version la plus récente du modèle, nommée PARTY$_{SOC}$ (Cécillon et al., 2021; ANNEXE 1).

Cependant, la question reste ouverte de savoir si le modèle PARTY$_{SOC}$ fonctionne sur des échantillons provenant de sites indépendants de son ensemble d'apprentissage et si les nouvelles informations fournies par l'analyse thermique Rock-Eval® et PARTY$_{SOC}$ peuvent réellement être utilisées pour initialiser la taille des compartiments cinétiques des modèles de dynamique du COS et si cela va améliorer la performance prédictive de ces modèles. Ce qui m'amène aux objectifs de cette thèse.

8. Mes questions de recherche et mon travail en trois chapitres

Mon travail de thèse peut être résumé en trois objectifs.

Comme décrit ci-dessus, la première série de questions de recherche est la suivante : Pouvons-nous appliquer le modèle $PARTY_{SOC}$ à des sites en dehors de son ensemble de calibration et si oui, quel est le potentiel de cette approche pour améliorer la précision des simulations du COS ?

Les deux parties suivantes se concentrent sur l'étude des limites et des possibilités de l'approche en utilisant des expériences de laboratoire. Plus précisément, dans la deuxième partie de cette thèse, nous demandons s'il est possible de comparer les paramètres Rock-Eval® d'échantillons provenant de différentes profondeurs d'un profil pédologique à des fins d'harmonisation des bases de données existantes.

La question de la troisième et dernière partie est de savoir s'il est possible de progresser vers une compréhension plus mécaniste du lien entre la stabilité thermique et biogéochimique. Plus précisément, peut-on étudier le rôle de l'adsorption sur les surfaces minérales — un mécanisme classique de protection du SOC — à l'aide de Rock-Eval® et, si oui, quelle est son influence sur la stabilité thermique ?

Utilisation des estimations de la persistance du SOC prédites par l'apprentissage automatique et l'analyse thermique Rock-Eval® pour améliorer la précision des simulations de la dynamique du COS

Mes objectifs étaient premièrement, de tester la dernière version du modèle PARTY$_{SOC}$ basé sur l'analyse thermique (fournissant des estimations du COS persistant) sur des échantillons indépendants et deuxièmement, d'utiliser les informations obtenues pour initialiser la taille des compartiments cinétiques du modèle AMG (actuellement le modèle le plus précis pour les grands cultures françaises) afin d'estimer l'amélioration potentielle des simulations du COS.

Nous avons émis l'hypothèse que le modèle PARTY$_{SOC}$ peut être appliqué efficacement sur de nouveaux échantillons de sol provenant de pédoclimats similaires aux sites utilisés pour sa calibration (Europe du Nord-Ouest). Deuxièmement, puisque le modèle choisi de la dynamique du COS (AMG) et le modèle PARTY$_{SOC}$ ont la même architecture de compartiments de SOC, nous nous attendions à ce que le partitionnement des compartiments obtenu avec ces deux approches soit directement comparable. Troisièmement, nous avons émis l'hypothèse que l'utilisation du partitionnement de la taille des compartiments cinétiques prédit par PARTY$_{SOC}$ améliorerait la précision des projections des stocks de SOC, par rapport à l'initialisation par défaut du modèle.

Ce travail a été mené sur des échantillons de sol d'archive et récents provenant de 9 expériences agronomiques françaises à long terme, incluant 32 traitements (Chapitre 1 de ce manuscrit ; Kanari et al., 2022). Nous avons montré que le modèle PARTY$_{SOC}$ prédit un partitionnement des compartiments de COS qui prend en compte les effets hérités de l'occupation du sol et des pratiques de gestion du sol et qui est optimal pour l'AMG. Le couplage des deux modèles, c'est-à-dire l'initialisation de la partition des compartiments d'AMG à l'aide de PARTY$_{SOC}$, nous a permis de reproduire les changements de COS observés à une échelle pluri-décennale dans les 32 traitements. Les simulations ont été particulièrement améliorées dans les sites avec des histoires complexes de gestion du sol, où les stocks de COS étaient loin de l'équilibre (Fig. 3). Comme c'est le cas pour la plupart des terres cultivées européennes, nous estimons que le potentiel net d'amélioration apporté par cette méthode est substantiel.

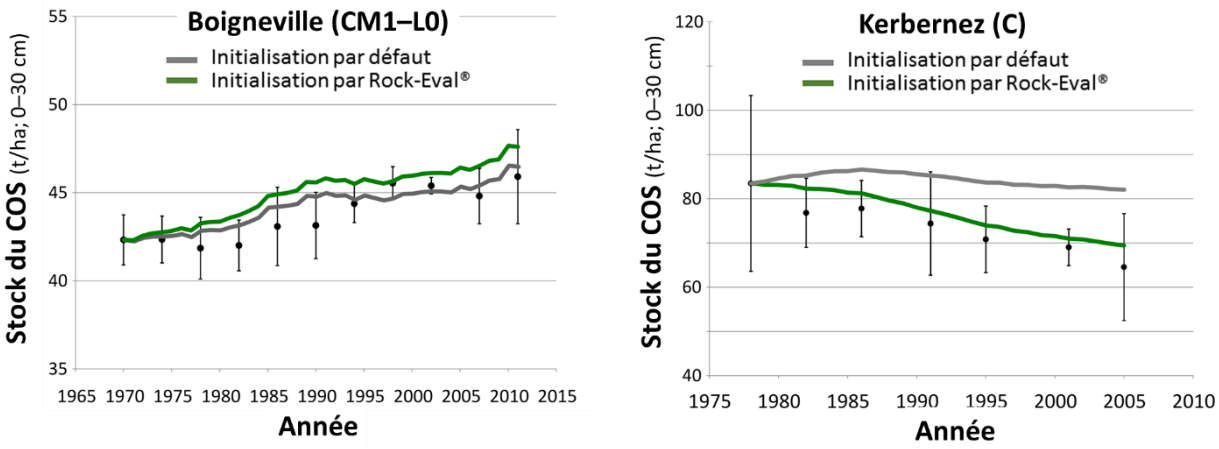


Figure 3 : Deux exemples de la performance du modèle AMG après deux méthodes d'initialisation différentes. A droite, un exemple de site (Kerbernez) avec un historique complexe de gestion du sol.

De plus, cette méthode d'initialisation est rapide, a un faible coût et est facile à mettre en œuvre. En tant que telle, elle présente un grand potentiel pour obtenir des prédictions précises de l'évolution à grande échelle des stocks de COS au cours des prochaines décennies.

Un autre message important que nous soulignons dans ce travail est la valeur des modèles multi-compartimentaux simples de la dynamique du COS en tant qu'outil pour prédire avec précision l'évolution des stocks du COS. Nous démontrons que lorsqu'il est correctement initialisé, un modèle simple fournit des simulations non biaisées et précises à l'échelle pluri-décennale. Contrairement à une nouvelle génération de modèles mécanistes qui sont mathématiquement plus complexes et nécessitent un travail de paramétrisation et de validation supplémentaire, les modèles simples en revanche, une fois que leurs compartiments cinétiques sont correctement initialisés, peuvent fournir des projections fiables du SOC à l'échelle de la parcelle, de l'écosystème et du globe.

Sur l'harmonisation des données Rock-Eval®

L'augmentation exponentielle du nombre d'études axées sur la MOS, lancée par la prise de conscience de son importance pour l'atténuation du changement climatique, a entraîné la production d'une énorme quantité de données obtenues sur des échantillons de sol collectés à l'aide de différentes stratégies d'échantillonnage. L'harmonisation des données disponibles est plus impérative que jamais pour l'incorporation correcte des informations dans des études à grande échelle, notamment la modélisation et la cartographie des stocks et du potentiel de séquestration du COS.

Le chapitre 2 aborde la question de la comparabilité des données entre les échantillons de sol obtenus à différentes profondeurs (Kanari et al., 2021). En utilisant des échantillons provenant de 10 parcelles, situées sur 8 sites forestiers en France, nous avons mené une expérience de mélange de sols, visant à trouver une méthode appropriée pour calculer les paramètres Rock-Eval® d'un profil de sol (0–50 cm) en combinant les résultats Rock-Eval® enregistrés sur ses sous-couches (0–30 et 30–50 cm).

Nous avons émis l'hypothèse que, dans les sols tempérés, la plupart des paramètres Rock-Eval® seront indépendants de la stratégie d'échantillonnage (c'est-à-dire que les paramètres Rock-Eval® caractéristiques de la SOM d'une couche de sol donnée pourront être obtenus à partir des paramètres Rock-Eval® mesurés dans ses sous-couches), car nous nous attendons à ce que l'effet de matrice minérale soit faible dans ces sols. Cependant, il était également prévu que le mélange de matériaux contrastés pourrait entraîner des modifications du signal Rock-Eval® des échantillons composites.

Les paramètres Rock-Eval® mesurés sur les échantillons composites étaient généralement en bon accord avec les paramètres calculés. Cependant, pour les paramètres dérivés du signal des hydrocarbures, la relation entre les valeurs mesurées et calculées n'était pas satisfaisante pour certains sites. Il s'agissait de sites avec une couche de sol profonde très riche en argile et une couche de surface avec une texture plus grossière, où le mélange a provoqué une addition d'argiles et donc une rétention des hydrocarbures par la matrice minérale pendant la pyrolyse. Par conséquent, dans le contexte du travail en cours pour l'harmonisation des bases de données, nous avons conclu que dans la plupart des sols tempérés, la caractérisation Rock-Eval® d'une couche de sol, y compris sa proportion de carbone persistante au cours du siècle, peut être déduite des caractéristiques des sous-couches.

En ce qui concerne les argiles et leur interférence avec le signal des hydrocarbures, nous fournissons une valeur seuil obtenue empiriquement de 20 % de différence de teneur en argile entre les sous-couches, au-dessus de laquelle le mélange des échantillons de sol pourrait entraîner des changements dans la forme du thermogramme des hydrocarbures et donc des paramètres liés (Fig. 4).

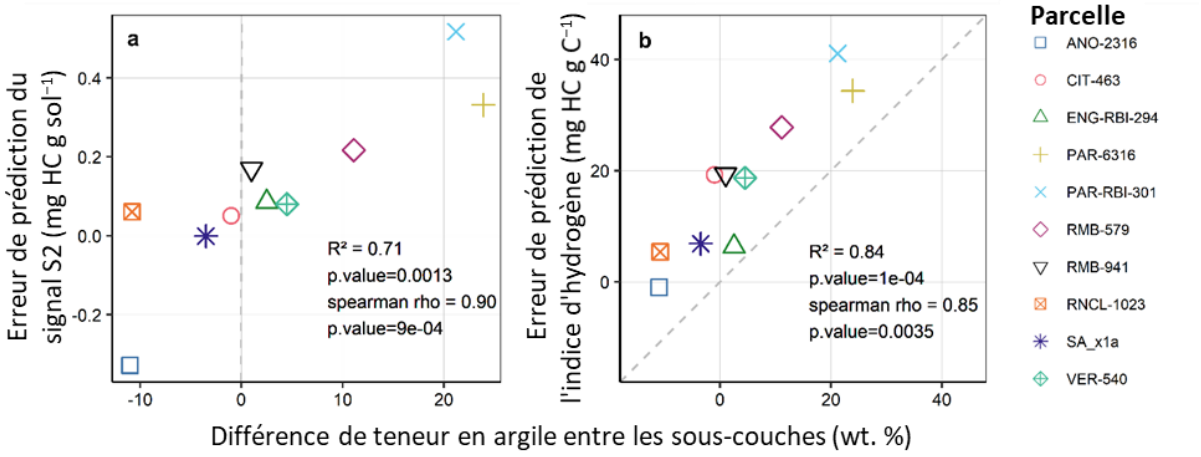


Figure 4 : Relation entre l'erreur de prédiction des paramètres Rock-Eval® et la différence de teneur en argile entre les sous-couches du sol (c.-à-d. la teneur en argile de la couche de surface − la teneur en argile de la couche souterraine) pour chaque parcelle. L'erreur de prédiction indiquée sur l'axe y est calculée comme la différence moyenne entre les valeurs calculées des paramètres Rock-Eval® et les valeurs mesurées par parcelle.

Enfin, une dernière directive a été tirée de ce travail, concernant une catégorie spécifique de paramètres : les paramètres de température. Ceux-ci sont par définition particulièrement sensibles aux changements de forme du thermogramme et doivent être calculés après la reconstruction du signal.

Ouverture de la boîte noire de Rock-Eval® : Pourquoi la stabilité thermique est-elle un bon indicateur de la persistance biogéochimique de la matière organique du sol ?

Bien que la valeur de l'analyse thermique Rock-Eval® pour l'estimation de la persistance et de la modélisation du COS soit désormais indiscutable, notre compréhension mécaniste de la relation entre la stabilité thermique et biogéochimique au centre de cette approche est encore très limitée.

Dans cette partie, sur la base d'un dispositif expérimental simple et d'un système modèle, nous nous sommes concentrés sur l'effet de l'adsorption comme mécanisme de protection potentiel de la matière organique vis-à-vis de la dégradation thermique. Nous avons analysé des composés biochimiques purs de plusieurs groupes (lignine, acide humique, protéines, glucides, lipides) pour obtenir leur signal Rock-Eval®. Nous avons préparé des mélanges organo-minéraux en suivant un protocole simple de sorption par lots pour étudier les changements qui en résultent sur la stabilité thermique et les paramètres Rock-Eval®. Comme étape intermédiaire, nous avons évalué l'effet de la présence de minéraux dans des mélanges simples à sec. Nous avons utilisé des composés biochimiques et des minéraux artificiels purs ainsi que de la matière organique et des matrices minérales de sols naturels.

Nous avons émis l'hypothèse que le degré et la force d'adsorption dépendraient des groupes fonctionnels des composés organiques ainsi que du pH, de la surface spécifique des minéraux et de leur réactivité. Nous nous attendions à ce qu'une association croissante entre la matière organique et les minéraux ait un effet plus fort sur les paramètres Rock-Eval® (précédemment corrélés empiriquement à la stabilité biogéochimique du COS). Nous avions également prévu des interactions entre les minéraux réactifs et les effluents de pyrolyse, ce qui se traduirait par un effet important de la présence de minéraux sur le signal Rock-Eval®, même dans les mélanges secs simples.

Le choix de travailler avec des composés purs pour des raisons de simplicité du modèle expérimental a révélé un inconvénient important de la méthode Rock-Eval® lorsqu'elle est appliquée à certains composés, à savoir une déficience du rendement en carbone des composés purs oxygénés. Ce problème est actuellement négligé dans la littérature car la technique est surtout utilisée pour les échantillons de sols et de sédiments naturels contenant de la matière organique, pour lesquels les analyses Rock-Eval® présentent de bons rendements en carbone (90–100 %). En outre, l'efficacité de la détection du carbone a été influencée par la présence de

goethite dans les mélanges pauvres en carbone (TOC≤1% en poids). Nous proposons que cette déficience est due au principe de fonctionnement du FID, et nous présentons des preuves obtenues en utilisant une configuration expérimentale de Rock-Eval® soutenant que le rendement en carbone manquant est associé à l'étape de pyrolyse. Pour des composés tels que les polymères organiques et la matière organique particulaire, le rendement en carbone de Rock-Eval® était très bon (rendement en carbone > 95 %).

Les expériences de sorption par lots ont été concluantes pour l'une des protéines, l'albumine de sérum bovin, et pour tous les minéraux (kaolinite, montmorillonite, goethite et trois matrices de sol naturel). Nous discutons des explications possibles de ce résultat, et du fait que pour les deux autres composés utilisés ici (acide humique et cystéine) l'efficacité d'adsorption était très faible.

Les mélanges secs d'albumine de sérum bovin avec des minéraux ont montré que certains minéraux purs peuvent avoir un effet important sur le signal Rock-Eval®. Nous avons observé une forte oxydation des effluents de pyrolyse par la goethite, et une rétention des effluents par la montmorillonite (Fig. 5a & b). Bien que beaucoup moins intense, une certaine rétention et oxydation des effluents a également été observée pour les sols naturels. Parmi les différents thermogrammes, le signal hydrocarbure était le plus sensible à la présence de minéraux (Fig. 5a & c). Une source importante d'incertitude remettant en cause l'origine de ces observations est la variabilité de la teneur en carbone entre les mélanges.

L'effet des mélanges adsorbés sur les paramètres Rock-Eval® était très similaire à celui des mélanges secs correspondants (Fig. 5a–c). L'effet cumulatif des changements causés sur le signal Rock-Eval® a conduit à des valeurs légèrement plus élevées du COS persistent prédit par $PARTY_{SOC}$ pour les mélanges adsorbés. Cependant, les mélanges secs ont également eu un impact significatif sur les prédictions fournies par $PARTY_{SOC}$ (Fig. 5d). Par rapport à la détection du composé pur, la différence causée par l'adsorption était beaucoup plus faible que celle causée par la présence de minéraux. Nous attribuons ce manque d'importance de l'adsorption au choix spécifique des composés utilisés dans ce système modèle et nous upposons que l'adsorption réussie de molécules organiques plus petites pourrait conduire à une protection plus efficace contre la dégradation thermique.

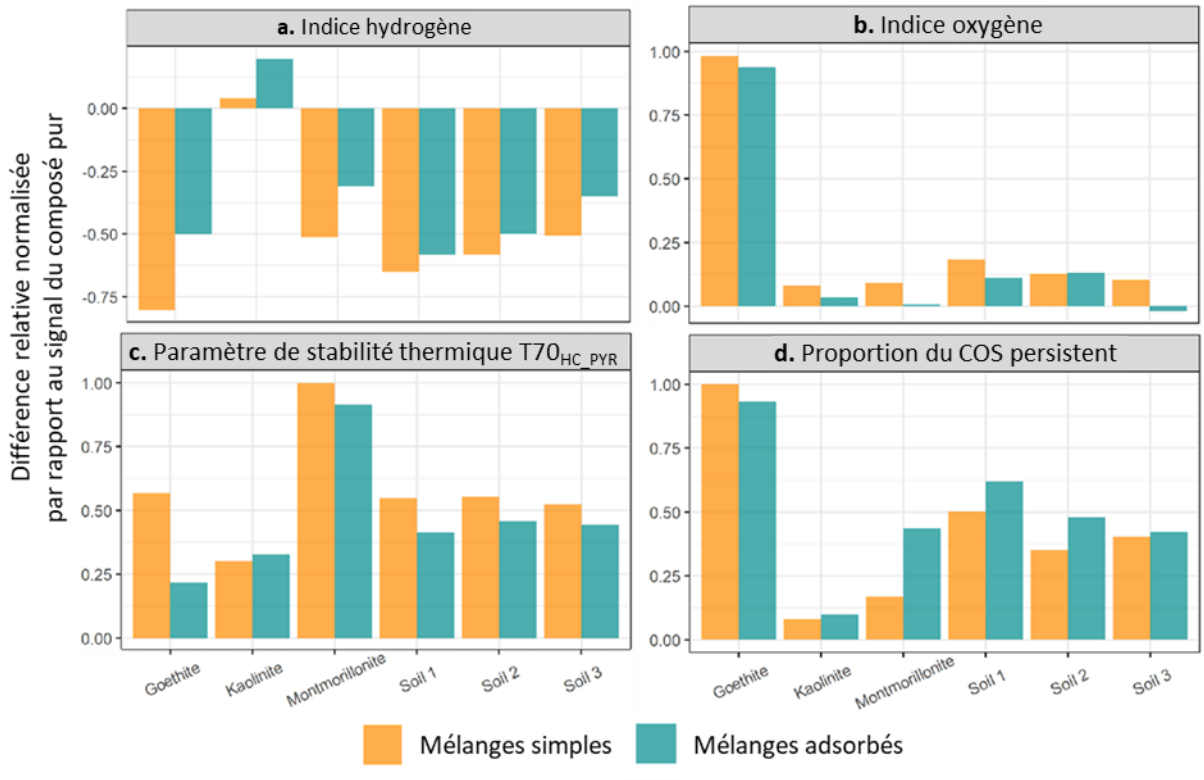


Figure 5 : Effet des interactions organo-minérales sur les paramètres de Rock-Eval®. Les barres oranges représentent des mélanges secs d'albumine de sérum bovin avec différents minéraux, tandis que les barres bleues représentent des mélanges adsorbés du même composé. Les valeurs des paramètres indiquées ici présentent le changement par rapport au signal du composé pur (sable) normalisé par la plage de chaque paramètre.

9. Conclusions

Ce travail présente plusieurs avancements concernant les objectifs principaux de cette thèse : « Comprendre et utiliser l'estimation de la persistance du carbone organique du sol par l'analyse thermique Rock-Eval® »

Nos résultats montrent que le modèle $PARTY_{SOC}$, une approche d'apprentissage automatique basée sur des données d'analyse thermique, peut fournir des informations précises sur la persistance du COS dans des échantillons inconnus. Nous soulignons la capacité de cette approche à prendre en compte les effets de l'histoire de l'utilisation des terres sur la dynamique du COS et nous montrons comment les informations qu'elle fournit peuvent être utilisées pour initialiser les simulations du COS et améliorer leur précision. La valeur de cette approche basée sur l'analyse thermique est claire, non seulement en raison de son pouvoir prédictif, mais aussi pour des raisons pratiques telles que la rapidité de l'analyse et son faible coût.

Le besoin actuel d'estimer la persistance du COS satisfait par cette approche soutient l'expansion de cette méthode à une plus grande échelle. En outre, l'une de nos principales conclusions concerne la valeur d'AMG, le modèle simple de dynamique du COS utilisé dans ce travail. En raison de la précision optimale des simulations fournies par ce modèle, lorsqu'il est correctement initialisé, nous recommandons qu'il soit utilisé comme un outil de prédiction opérationnel pour développer des pratiques de gestion des terres appropriées. Nous pensons que la simplicité et l'efficacité du modèle AMG est un avantage important par rapport aux modèles du COS plus récents et plus complexes qui sont très difficiles à contraindre et à évaluer. L'approche proposée devrait aboutir à des prédictions plus précises de l'évolution des stocks du COS à l'échelle nationale, voire continentale, et pourrait soutenir la mise en œuvre de politiques efficaces d'atténuation du changement climatique.

Dans le contexte de l'application de cette méthode à plus grande échelle, nous avons considéré la question de l'application de cette technique d'analyse thermique sur des échantillons de sol déjà disponibles mais obtenus à différentes profondeurs comme c'est souvent le cas dans différents projets de surveillance des sols. Nous montrons qu'il est possible de déduire la

caractérisation Rock-Eval® (y compris la persistance du COS) d'une couche de sol à partir des caractéristiques de ses sous-couches car les paramètres obtenus avec cette méthode sont linéairement additifs pour la plupart des sols de forêts tempérées testés ici. Nous soulignons que la méthode de calcul doit être adaptée au type de paramètre et nous fournissons des directives concernant les procédures les plus appropriées. Nous attirons l'attention sur un effet minéral qui entrave la prédiction des paramètres Rock-Eval® liés aux hydrocarbures dans les sols présentant une différence prononcée de teneur en argile entre les sous-couches du sol (Δ_{CLAY}>20 % wt.). Notre travail peut être considéré comme un guide empirique ouvrant la voie à l'harmonisation des données obtenues avec cette technique.

Enfin, dans un effort pour ouvrir la boîte noire de Rock-Eval®, nous avons tenté de mieux comprendre les interactions entre la matière organique et les minéraux qui influencent le signal de Rock-Eval® et les estimations associées de la stabilité thermique. Nous avons utilisé un système modèle composé de matières organiques et de minéraux purs ainsi que de la matière organique particulaire naturelle et de matrices de sol. Nous fournissons des informations sur les limites de la méthode Rock-Eval® lorsqu'elle est utilisée pour analyser des composés oxygénés et nous invitons à une révision des études antérieures utilisant des composés similaires et ignorant cet effet. Nous soulignons le fort effet des minéraux purs tels que la goethite et la montmorillonite sur le signal Rock-Eval®, cependant cette observation n'est pas pertinente pour l'environnement du sol en raison des grandes différences dans la nature des matériaux utilisés et des conditions occurrentes. L'analyse de la matière organique particulaire confirme la capacité de Rock-Eval® à détecter la SOM avec une grande précision.

Une légère augmentation des prédictions de la proportion persistante de SOC est observée lorsque des mélanges organo-minéraux adsorbés sont analysés au lieu de simples mélanges secs de même composition. Cependant, il n'y a pas de tendance cohérente parmi les paramètres thermiques, ce qui permet de conclure que l'effet de l'adsorption produite dans cette étude sur le signal Rock-Eval® n'est pas significatif. De plus, cette observation est éclipsée par la forte influence de la matrice minérale et les fortes différences de teneur en carbone entre les mélanges. Nous discutons des options possibles pour favoriser la formation d'adsorption et nous suggérons que des travaux supplémentaires sont nécessaires pour déchiffrer le lien entre la stabilité thermique et biogéochimique.

10.Perspectives

Une perspective immédiate concernant l'utilisation du modèle $PARTY_{SOC}$ à l'échelle nationale est son application sur un ensemble d'échantillons plus large. Un effort déjà en cours est l'utilisation de $PARTY_{SOC}$ sur des échantillons du réseau RMQS (Réseau de Mesures de la Qualité des Sols) comprenant un total de 2240 sites, répartis de façon homogène sur toute la France métropolitaine et la Corse selon une grille de 16 km². Ceci permettrait de générer une carte de persistance du COS qui pourrait contribuer à l'amélioration des simulations d'évolution des stocks du COS à l'échelle nationale, avec une importance claire en tant qu'outil d'aide à la décision (projet de thèse d'Amicie Delahaie).

De plus, une priorité élevée est d'inclure plus de sites dans l'ensemble d'apprentissage du modèle $PARTY_{SOC}$ pour augmenter sa robustesse et sa généricité. Sur la base du bon accord entre les prédictions de $PARTY_{SOC}$ et la partition des compartiments post- optimisée par AMG observée dans cette étude, nous suggérons que la plupart des LTE agricoles avec des simulations AMG précises pourraient être utilisées comme sites de référence pour une future version du modèle $PARTY_{SOC}$. Ceci pourrait améliorer la précision du modèle $PARTY_{SOC}$ dans des conditions pédoclimatiques tempérées et pourrait éventuellement lever une importante limitation technique à son expansion géographique.

Avant la mise en œuvre continentale ou mondiale de $PARTY_{SOC}$ dans les études de modélisation du COS, ses performances devront être validées sur une gamme plus large de pédoclimats. De plus, comme dans cette étude, son potentiel pour améliorer la précision des simulations devra être démontré pour les nouveaux domaines d'application. Pour continuer à utiliser le modèle AMG pour ces étapes d'évaluation, il faudra disposer de LTE provenant d'une variété de lieux géographiques et de données de suivi à long terme (climat, couverture du sol, apports de C, pratiques de gestion, caractéristiques du sol) pour exécuter les simulations.

Une autre possibilité serait de profiter de la littérature sur les travaux de modélisation déjà réalisés avec d'autres modèles multi-compartimentaux (par exemple, Century, RothC) et de tester le potentiel de $PARTY_{SOC}$ pour améliorer leurs simulations en initialisant leurs compartiments du COS "stable".

L'avenir à long terme de cette approche pourrait ressembler à un cycle constant d'adaptation, d'expansion et d'amélioration pour tirer le meilleur parti de ce que la caractérisation du COS a à offrir. À terme, d'autres paramètres (par exemple, le pH du sol) pourraient être ajoutés à l'ensemble d'étalonnage du modèle $PARTY_{SOC}$ et sa structure pourrait évoluer en testant d'autres algorithmes d'apprentissage automatique.

Nos travaux sur la linéarité des paramètres de Rock-Eval® suggèrent qu'il est possible de déduire la caractérisation d'une couche de sol à partir des sous-couches qui la composent. Bien que cette observation soit en faveur de la mise en œuvre du modèle $PARTY_{SOC}$ sur des échantillons existants, il est important de ne pas dépasser les limites pédoclimatiques pour lesquelles cela est vrai.

Nous suggérons de mener des expériences similaires mais plus étendues couvrant une plus grande variabilité de sols présentant une hétérogénéité texturale et minéralogique verticale. Des seuils mieux définis de l'effet minéral dans les sols naturels pourraient être utiles pour prédire les limites de la méthode.

Une perspective intéressante liée de manière générale à la méthode Rock Eval® serait d'estimer l'incertitude associée à chaque paramètre. Des indications similaires existent mais uniquement pour des matériaux standards. Compte tenu de la forte hétérogénéité du matériau sol, une estimation de l'erreur correspondant aux différents types ou fractions de sol semblerait utile. Ceci permettrait la propagation de l'incertitude dans des calculs tels que celui présenté dans ce travail mais aussi dans d'autres cas où les paramètres d'acquisition de Rock-Eval® sont utilisés pour déduire des termes plus élaborés.

En ce qui concerne notre compréhension des processus reflétés par le signal Rock-Eval® qui en font un indicateur approprié de la stabilité biogéochimique, des travaux supplémentaires sont nécessaires.

Un inconvénient important des résultats présentés dans cette étude qui devrait être amélioré immédiatement est la variabilité de la teneur en carbone dans les mélanges produits.

Toutefois, au cas où cette perspective s'avérerait trop difficile, une approche intéressante consisterait à étudier l'effet de la concentration en carbone sur l'intensité de l'effet minéral, pour différents composés organiques et minéraux.

Des efforts supplémentaires devraient être consacrés à la création d'associations plus fortes et à l'examen plus approfondi de leur influence sur la stabilité thermique. Le pH de la solution doit toujours être étroitement surveillé et contrôlé, car il peut avoir un impact sur le type de liaisons qui peuvent être générées. L'influence de la taille moléculaire des composés utilisés sur l'efficacité de la formation d'adsorption est un aspect intéressant à examiner.

Bien qu'il semble trivial que la dissolution complète des composés dans une solution mère soit confirmée, soit par une mesure directe, soit par l'utilisation d'un contrôle, ceci n'est pas toujours pris en compte par les études de sorption par lots trouvées dans la littérature. La comparabilité des résultats obtenus avec ces expériences bénéficierait de procédures plus standardisées et communément acceptées.

Il convient également d'étudier d'autres mécanismes qui pourraient influencer le lien entre la stabilité thermique et biogéochimique, tels que la récalcitrance et l'équilibre entre le gain d'énergie et l'énergie d'activation. L'utilisation de méthodes complémentaires telles que la calorimétrie différentielle à balayage et la microscopie électronique à transmission pourrait fournir des informations précieuses pour nous aider à mieux comprendre la formation des associations organo-minérales.

En conclusion, il serait trompeur de considérer que les sols sont la seule solution, ou qu'ils constituent à eux seuls une solution adéquate, pour compenser les émissions globales de gaz à effet de serre et lutter contre le changement climatique. Cependant, la gestion intelligente des sols à l'échelle globale reste une option pratique et avantageuse pour les humains ainsi que pour l’environnement, qui peut contribuer de manière significative à éviter des émissions supplémentaires et à garantir la sécurité alimentaire. La gestion des sols peut être une solution éthique et solidaire.

Intentionally blank page

V. Synopsis

Soil organic matter (SOM) is an important indicator of soil health, the ability of soils to provide various ecosystem services, such as water quality and storage, soil life support and biodiversity conservation. Moreover, this reservoir contains two times more carbon than the atmosphere and vegetation combined and it is located at a critical position, in direct exchange with the atmosphere. As a result, small changes in soil organic carbon (SOC) stocks have a major impact on atmospheric CO_2 concentration and climate regulation. Improving land-management practices to increase SOC stocks is amongst the most practical and affordable solutions contributing to climate change mitigation and food security (4per1000). However, achieving this goal is currently hindered by our limited understanding of the processes controlling SOC persistence and our disability to predict its evolution.

Part of SOC is characterized by a short turnover time counting only hours, whereas more persistent components can reside in the soil for hundreds to thousands of years. The amount of persistent carbon significantly varies between locations according to pedoclimatic conditions and land-use history. Models of SOC dynamics used to predict the evolution of SOC stocks divide SOC into conceptual stability pools and correct initialization of pool sizes is crucial for their accuracy. At the same time, apart from some very specific long-term experimentation sites (long-term bare fallow, C_3 to C_4 vegetation change chronosequences), means to efficiently estimate the size of model kinetic compartments remain limited. Recently, Rock-Eval® thermal analysis was proposed as an appropriate method for direct determination of SOC stability. Based on this technique, a machine learning model ($PARTY_{SOC}$) was developed, calibrated on long-term bare fallow data, which allows the estimation of the size of the centennially persistent SOC fraction in a sample. However, it still remains an open question whether the new information provided by Rock-Eval® thermal analysis and $PARTY_{SOC}$ can actually improve the predictive performance of SOC dynamics models.

The first objective of my thesis was to use this novel approach to estimate the size of SOC stability fractions in unknown samples in an effort to evaluate its performance. Second, I aimed at using the obtained information to initialize the SOC pool sizes of the [2]AMG model of SOC dynamics to test if this approach improved the accuracy of its simulations. Nine French long-term agronomical experiments from the [3]AIAL database and the [4]SOERE PRO network, including 32 treatments, were used to test this hypothesis and estimate the potential improvement brought by this initialization approach. Using $PARTY_{SOC}$ on unknown samples from independent sites resulted in predictions of optimal pool partitioning, accounting for legacy effects of soil management history. Initializing the SOC pool partitioning of AMG using this approach improved its overall accuracy when reproducing the observed SOC dynamics in the nine sites. These results indicate that the simple AMG model combined with a robust initialization approach can simulate observed SOC stock changes with precision. Harnessing the predictive power of simple models for development of science-based land-management strategies is recommended. Finally, the proposed approach is quick and fully automated and can be easily implemented on soil monitoring networks, paving the way towards precise predictions of SOC stock changes over the next decades.

Having obtained these encouraging results, for the second part of my thesis, I turned my focus back to the Rock-Eval® technique. My goal here was to understand more about the limits and the potential of this analytical method using two simple experimental designs.

The first experiment aimed at predicting the response of Rock-Eval® parameters to mixing of soil layers to answer the practical question: Is the comparison of available Rock-Eval® data obtained on samples from different depths possible? The main idea was that if we can predict the behavior of Rock-Eval® parameters according to the sampling strategy, this would allow the comparison of results of studies using different sampling methods.

[2] AMG is a French model of SOC dynamics used in this work and described in detail in Sect. 1.10 (Andriulo et al., 1999)

[3] AIAL is an extensive database including the majority of French long term agronomic experiments (Clivot et al., 2019)

[4] SOERE PRO is an “Organic residues” research observatory including a network of long term field experiments studying recycling of organic residues in agriculture

The initial hypothesis was that there would be no major effect, considering that this study was conducted on temperate soil samples. Natural soil samples from ten forest plots in France were used. Carbon-enriched surface soil samples (0–30cm), mineral-enriched deeper soil samples (30–50cm) and their mixtures were analyzed with Rock-Eval®. The linearity of the Rock-Eval® parameters of soil mixtures between theory and observation was investigated. For the calculation of theoretical values of Rock-Eval parameters® it was necessary to adapt the calculation method according to the type of parameter. In conclusion it was possible to predict Rock-Eval® parameters for the 0–50cm layer. The natural mineral matrix effect in temperate soils was indeed low. Only in cases where deep soil was much richer in clay than the surface layer (>20%), there was an interaction with pyrolysis effluents that affected the detection of hydrocarbons. This work provides a first empirically developed guide towards harmonization of databases of Rock-Eval® data.

The goal of the second experiment was to progress towards a more mechanistic understanding of the link between thermal and biogeochemical stability of SOC. Even though comparison of Rock-Eval® results with classical methods for estimation of carbon stability pools has shown a strong correlation, this link remains purely empirical. Here, the main goal was to understand the role of organo-mineral interactions. A simple model system was implemented involving pure organic compounds and minerals as well as particulate organic matter and natural soil mineral matrices. A series of batch sorption experiments were conducted with the objective of tracking the changes on the Rock-Eval® signal caused by the existence of adsorption between a compound and a mineral compared to pure compound detection and detection in dry simple mixtures. Preliminary results suggest that the apparatus in its current set-up has a limited ability to detect oxygenated pure organic compounds. The presence of dry reactive minerals such as goethite and montmorillonite can interfere strongly with the detection of organic effluents during the pyrolysis step. Moreover, our experiments showed that producing organo-mineral associations to a sufficient degree is quite challenging. Finally, although successfully formed, pre-existing adsorption between organic compounds (bovine serum albumin) and minerals had a non-significant increasing effect on the overall thermal stability. These results allow a first estimate of the part of carbon stability attributed to the interaction of organic matter with minerals and very importantly they reveal previously unknown limitations of the Rock-Eval® technique regarding detection of pure compounds.

Intentionally blank page

1. General introduction

1.1. Current environmental and climate challenge

Every year, human activities, such as burning fossil fuels and deforestation, release more than 10 Gt of carbon into the atmosphere, mainly in the form of CO_2. About 30% of the additional CO_2 produced each year is absorbed by terrestrial ecosystems, 28% by the oceans, whereas 49% is retained by the atmosphere (Friedlingstein et al., 2020), causing the atmospheric CO_2 concentration to rise. In 2019, the increase in CO_2 compared to pre-industrial levels (1750) reached 148% (Fig. 6; WMO, 2021). This radical change in the composition of the atmosphere is contributing to global warming and climate change. While discussions at the global level about reducing emissions of greenhouse gases are intensifying (e.g., COP 26), the trend is currently still positive with global emissions increasing by approximately 90-300 Mt C each year (Peters et al., 2020). Global organisations such as the IPCC have been sounding the alarm for decades. In its last report IPCC (2021) warns that unless immediate large-scale action is taken, the 1.5–2 °C warming limit agreed upon by the Paris Agreement (UNFCCC, 2015) will be beyond reach.

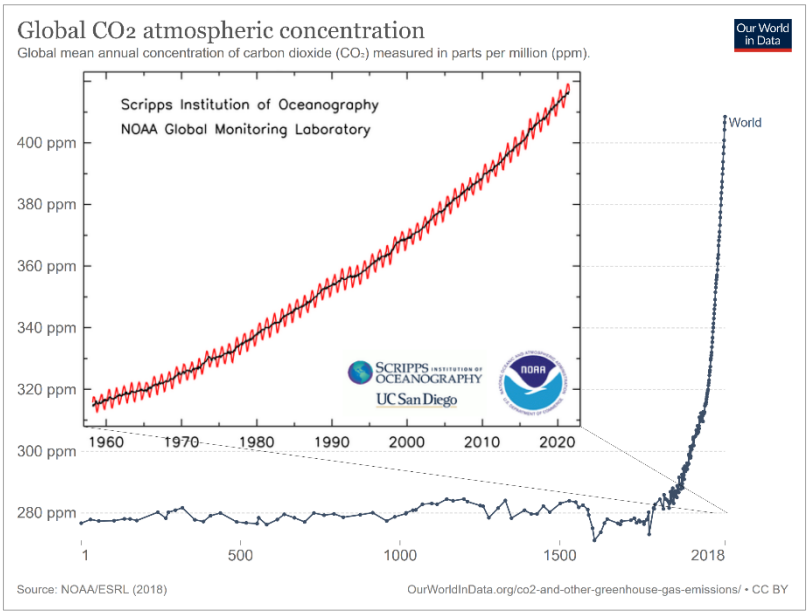


Figure 6: Atmospheric CO_2 concentration rise in the last two millennia. Values before 1958 are inferred from gases trapped in Antarctic ice cores (Bereiter et al., 2015). Data from 1959 onwards are direct measurements of atmospheric CO_2 concentration from the Mauna Loa observatory in Hawaii. Figure modified after sources mentioned on the figure.

Today, all areas of the world are already impacted in some way by climate change. This reality is not only an environmental crisis but it has serious economic and societal aspects as well. According to the Universal Ecological Fund's report (FEU-US, 2017) the economic cost of climate change including extreme weather events and health hazards due to air pollution was estimated at a yearly average of 240 billion for the U.S. economy only. Worldwide, the price of the loss of ecosystem services due to land degradation alone, caused mainly by millennia of human land use (Sanderman et al., 2017) and aggravated by climate change, corresponds to 10% of the annual global GDP (IPBES, 2018). Moreover, the World Meteorological Organization (WMO, 2021) estimated that extreme weather phenomena caused the displacement of about 10 million people just in the first half of 2021. The simultaneous climatic crisis, causing ecosystems to fall out of balance and contributing to land degradation, combined with the accelerating increase in human population expected to reach 10 billion by 2100 (UN, 2019), raises a serious issue for feeding the global population.

Governments and global organizations are called to take action including emission reductions and a shift to clean energy. In the frame of the European Green Deal the EU aims to become the first carbon-neutral economy by 2050 by reaching net-zero greenhouse gas emissions (European Commission, 2019). Programmes launched by the UN such as the Sustainable Development Goals, to be achieved by 2030, were designed to ensure a safer future for the global population (UN General Assembly, 2015). Nature-Based Solutions (NBS) and negative emission technologies (NET) are necessary to reach carbon neutrality, stabilize temperature and limit the effects of climate change (IPCC, 2018; Anderson et al., 2019). Due to the fact that NBS are no regret solutions but also because of their potential to reduce emissions estimated at a significant 37% of the 2030 goal, scientists and environmental activists urge for ecosystem conservation and restoration, and sustainable management of forests, wetlands and grasslands (Griscom et al., 2017) with the potential to sequester organic carbon in terrestrial ecosystems. Projecting future climate change and developing appropriate climate policies requires an extensive understanding of the global carbon cycle and the ability to predict shifts between its reservoirs, especially land and the atmosphere.

1.2. Soils in the global carbon cycle

The global carbon cycle (Fig. 7) refers to the distribution of carbon amongst various reservoirs and the flows that connect them. Shifts of carbon out of a reservoir increase its concentration into others and cause changes in the cycle balance. Because this cycle is inherently linked to climate, serious efforts have been made by the scientific community to acquire detailed knowledge to support policy development and action to mitigate climate change (Global Carbon Project, 2001). Although important amounts of carbon are found in reservoirs such as inorganic carbon in rocks or carbon dissolved in the deep ocean, exchanges between these reservoirs and the atmosphere are slow (Riebeek, 2011). The flows most relevant to climate regulation are the ones occurring faster, between the fossil fuel pool, soil, vegetation and atmosphere. Compared to the atmosphere, soils contain two to three times more carbon, in the form of soil organic matter (SOM), making them the largest terrestrial organic carbon reservoir (Jobbagy and Jackson, 2000; Houghton, 2007).

On a global scale the mass of soil organic carbon (SOC) is estimated to be 1500–2400 Gt, which corresponds to 3–4 times the total amount of carbon found in all biosphere (600 Gt). The size as well as the critical position of this reservoir, at the border between solid earth and atmosphere, underline its importance for climate regulation (Batjes, 1996; Lal, 2004a). Relatively small shifts of net SOC source or sink can have major effects on concentration of atmospheric CO_2 (Lal, 2004a). This idea sparked the "SOC sequestration movement" and among others the "4 per 1000" initiative promoting that a relatively small annual increase of 0.4% in SOC stocks is enough to compensate for the increase in atmospheric CO_2 caused by anthropogenic emissions (Minasny et al., 2017). But even setting aside the climate debate, SOC stocks must be protected and if possible increased, as SOM is essential for soil health and provision of multiple soil ecosystem services (Lal, 2004b).

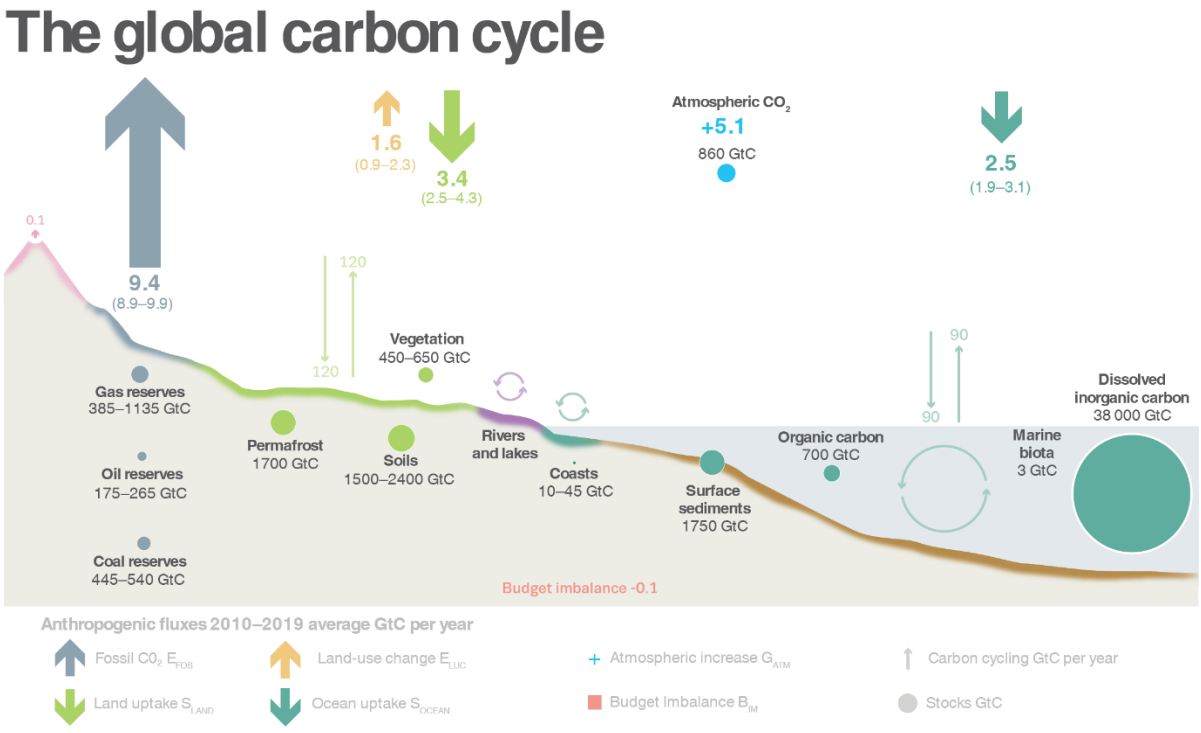


Figure 7: Diagram of the global carbon cycle including the natural fluxes and stocks in different reservoirs, as well as perturbations caused by anthropogenic activity, namely increase in fossil fuel emissions (E_{FOS}) and land-use change related emissions (E_{LUC}). Carbon emissions either retained in the atmosphere expressed as a CO_2 growth rate (G_{ATM}), adsorbed by land (S_{LAND}) or oceans (S_{OCEAN}) are averaged over the last decade (2010-2019; Friedlingstein et al., 2020).

1.2.1. Quick view of the processes that drive SOC

The main process of organic carbon input into soil is the fixation of CO_2 from the atmosphere by plants through the process of photosynthesis, and deposition of this carbon either directly belowground through root exudates or in the form of residues in the aboveground litter layer (Fig. 8). In cases of managed land other exogenous organic matter additions might occur such as manure or compost amendment. Organic matter in soils enters the metabolic pathway of various heterotrophic organisms. Some of the carbon is respired and released back into the atmosphere in the form of CO_2. The variety of microorganisms as well as the soil mineral substrate make soil a heterogeneous environment accommodating a myriad of biotic and abiotic interactions that will define the residence time (s. Sect. 1.8.) of organic carbon. Organic carbon can be lost from soils through erosion and leaching as well but the relative importance of these processes is low in most soils (FAO & ITPS, 2015).

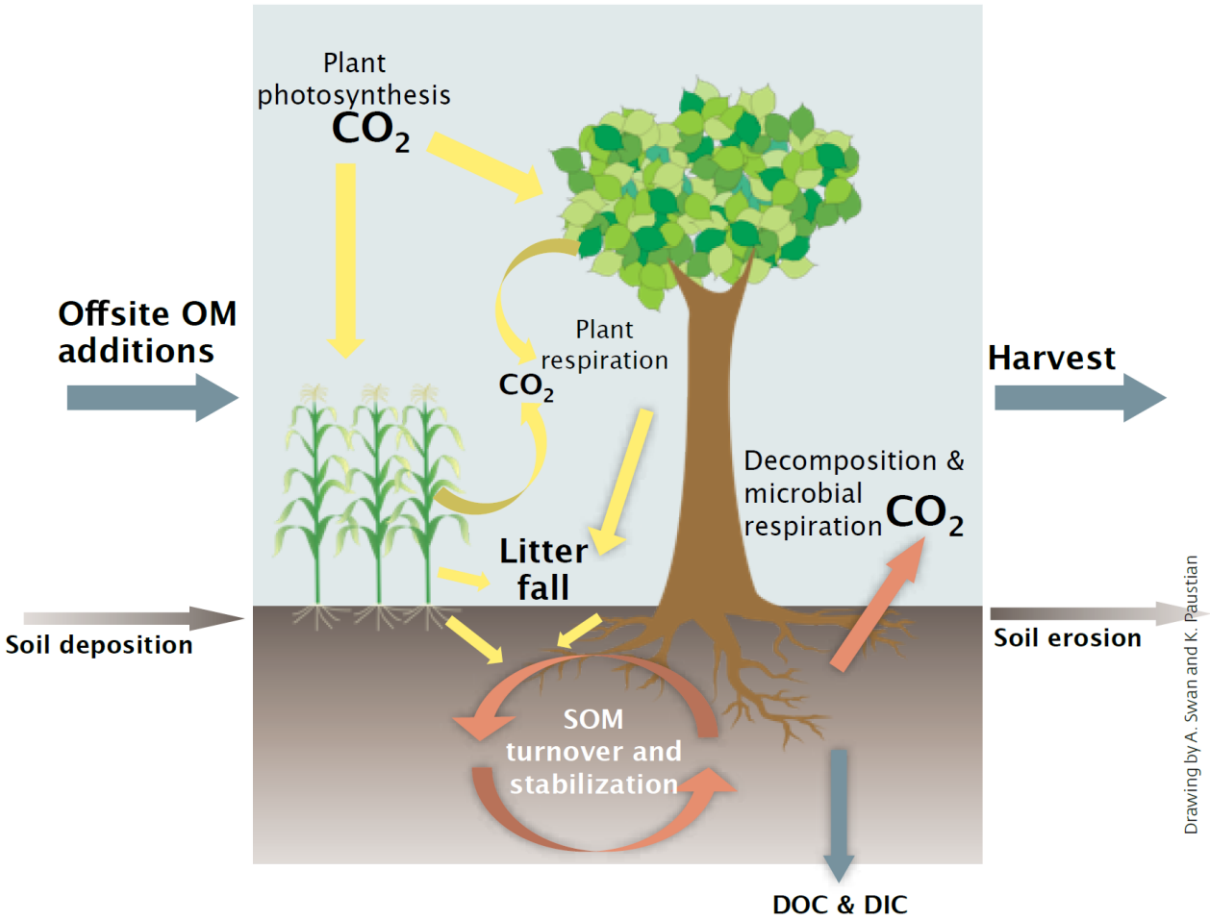


Figure 8: Schematic representation of the main SOC input and output processes controlling SOC stocks (FAO and ITPS, 2015).

1.3. Soil properties, functions and ecosystem services

The term "soil functions" often refers to processes that are taking place in soils because of their inherent physical and chemical properties (e.g., texture, structure, mineralogy…), whereas "soil ecosystem services" are the associated benefits to the ecosystem seen from a human perspective (Spangenberg et al., 2014; Vogel et al., 2019).

Soil functions include a wide range of processes, taking place under the unique conditions that exist only in the soil environment, *"making this thin layer of loose material covering the Earth's surface the difference between life and lifelessness"* (Berhe, 2019). Seven soil functions are considered key for human well-being and are underlined in the EU Soil Framework Directive (European Commision, 2006): (1) biomass production for agriculture or forestry, (2) storage filtering and transport of nutrients and water, (3) support of the biodiversity pool through provisioning of an appropriate habitat for multiple species, (4) provisioning of a physical and cultural environment for humans, (5) acting as a source of raw materials, (6) as a carbon pool, and (7) as an archive of geological and archaeological heritage.

The ability of a soil to host these functions requires that its properties remain intact. The inherent composition of a soil as a mixture of minerals, living organisms, soil organic matter and water gives it its characteristic texture, pH, bulk density, cation exchange capacity, electrical conductivity, porosity, aggregate stability and hydraulic conductivity (Hatfield et al., 2017) – all crucial to soil health.

1.3.1. The influence of SOC content on soil properties and soil functions

Soil organic matter is considered the most important component that can be efficiently managed to enhance soil health and achieve sustainable soil development. Increasing the organic carbon content of soils is beneficial to many functions such as their fertility (Tiessen et al., 1994; Six et al., 2004), their climate regulation capacity (Heimann and Reichstein, 2008), and their water filtration and retention potential (Doran and Parkin, 1994).

An increase in SOC leads to a multitude of complex interactions; a few examples are provided here taken mainly from Pellerin et al. (2019). As a chemical element soil organic carbon is expected to directly influence the overall chemistry and thus the chemical properties of soil. These include but are not limited to its nutrient availability and cation exchange capacity. SOM has a cation exchange capacity that is up to ten times higher than clays (per unit of mass), thus it can increase the retention of nutrients (such as Ca^{2+}, Mg^{2+}, or K^{+}) vital for plant growth, improving soil fertility.

Physical properties of soil such as its porosity, bulk density, and structural stability are also influenced by its SOC content. A combined effect of the improvement of soil physical properties is the increase in its capacity to store water (Tisdall and Oades, 1982) and the improvement of its fertility (as more space becomes available for root growth), enhancing the resistance of soil to erosion.

Moreover, biological properties of soil benefit from an increase in SOC content, namely through the provision of energy for microorganisms, supporting once again the plant soil interaction though an increase in biodiversity.

Soils constitute the medium in which all plants grow, including feed, fibre and fuel production. They are the source of building material such as earth, sand, clay or peat. They are the foundation for all structure and infrastructure support. They filter and store water and they are immensely important for flood regulation. They are a conserve of our cultural heritage. Soils might even be considered as part of the recreational landscape important for human physical and even mental health (Sustainable Soils Alliance, 2022).

1.4. Current status of global soil health

Despite their importance the ability of soils to support ecosystem well-being is currently compromised through serious land degradation caused by human activity and climate change. Deforestation or inappropriate forest management, expanding agriculture and intensive agricultural practices are some of the most important direct human-induced drivers of land degradation (Olsson et al., 2019). It is estimated that one third of the world's soils are currently under moderate or severe stress (Fig. 9; IPBES, 2018). The threat of land degradation is amplified by the positive feedback relationship with climate change through reduced rates of carbon uptake and increased GHG emissions in degrading soils (Olsson et al., 2019). With rates of soil erosion estimated to be 2–3 orders of magnitude greater than soil formation, soils are a limited and fragile resource that can be considered as non-renewable (JRC, 2016).

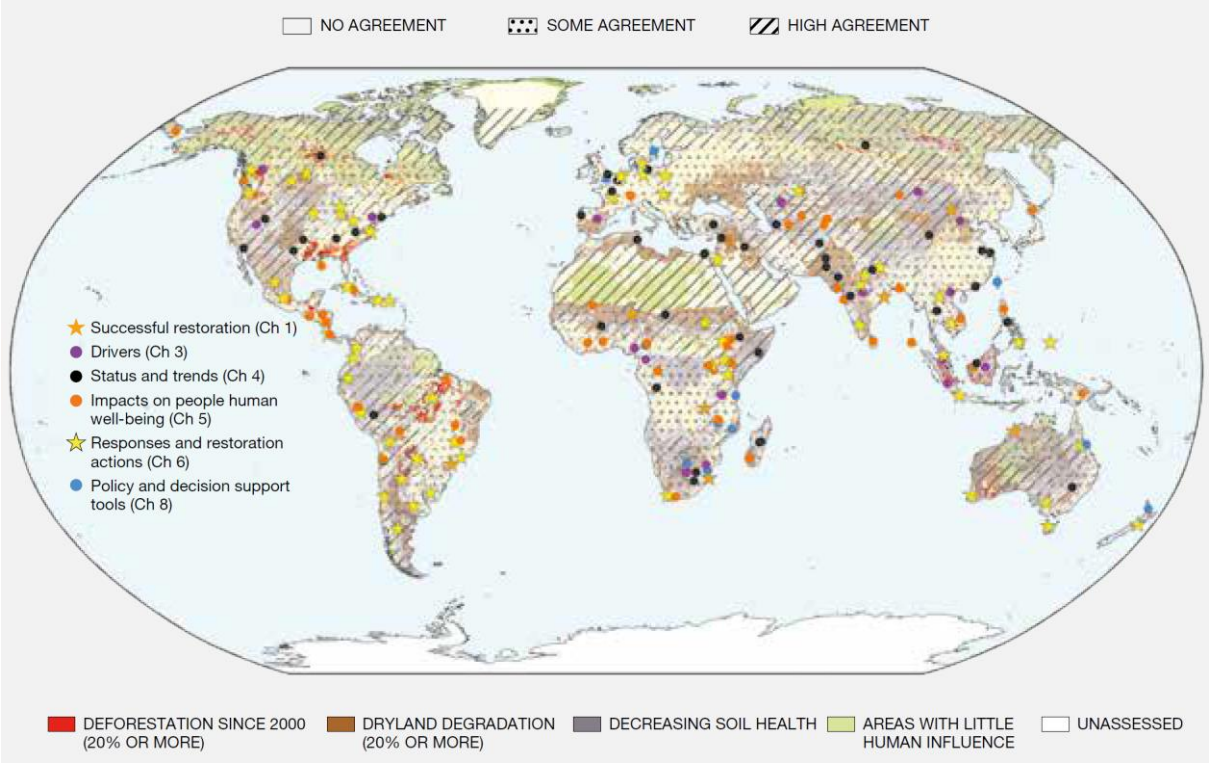


Figure 9: Level of loss of ecosystem services caused by land degradation in different parts of the world (IPBES, 2018).

Soil degradation however, does not only involve the loss of soil mass through erosion but also the depletion of soil nutrients and organic matter, sealing of the soil surface through urbanisation, soil compaction, salinization, acidification and increase in toxic elements concentration causing partial or complete failure of soil functions (IPBES, 2018).

Amongst EU members, 13 countries are reporting being affected at some degree by desertification (EEA, 2019), while erosion rates in Europe are expected to keep increasing in the next 30 years (Panagos et al., 2021) resulting in an estimated loss of 16% of the EU agronomic income by 2050 (EAA, 2019). Nevertheless, not all soils are susceptible to degradation to the same degree. Due to their sensitivity to rising temperatures, permafrost soils are a major concern. Even a small loss of carbon of their estimated current stocks could offset efforts for SOC sequestration (UNEP, 2019). Many peatland soils, intensively used for agriculture due to their high productivity, are also greatly affected by soil degradation and need to be carefully managed to maintain their potential and avoid additional emissions (Liu et al., 2012).

The current state of soils is predominantly a result of their exploitation by humans in past millennia heavily impacting their health and their and ability to stock carbon (Amundson et al., 2015; Sanderman et al., 2017). While inappropriate land management causing SOC depletion is a major threat for soils, increasing SOC is important both for soil conservation and climate regulation.

1.5. Historic trends in SOC stocks

Understanding the causes, magnitude and consequences of historic SOC trends is considered a valuable reference point for estimating the current potential of soils to sequester carbon (FAO & ITPS, 2015). Evidence suggests that since the last ice age human-induced land cover changes and particularly unsustainable land management caused significant SOC depletion (Amundson et al., 2015; Sanderman et al., 2017). Estimates of the magnitude of historic SOC loss vary widely, with values at the lower end of the spectrum at 40–55 Pg C (Houghton, 1995; IPCC, 1995) and maximal estimates in the range of 500 Pg C (Lal, 2001).

In an interesting study combining land use data from a historical database with environmental variables using machine-learning, Sanderman et al. (2017) showed that SOC loss has been ongoing for at least 12 000 years. The global SOC loss saw an exponential increase in the last 200 years (Sanderman et al., 2017; Fig. 10) coinciding with the industrial revolution and the expansion and intensification of agriculture. Most soils under agricultural use may have lost 25–75% of their initial SOC stocks due to intensive cultivation, with a cumulative SOC loss of 133 Pg C caused by agriculture alone (Lal, 2013; Sanderman et al., 2017). This loss showed spatial and temporal variability, whereas a strong correlation was found between degree of land degradation, including decrease in biomass and biodiversity, water and soil quality, and the SOC loss caused by land use change (Sanderman et al., 2017), indicating that not only land use change but also management has a significant impact.

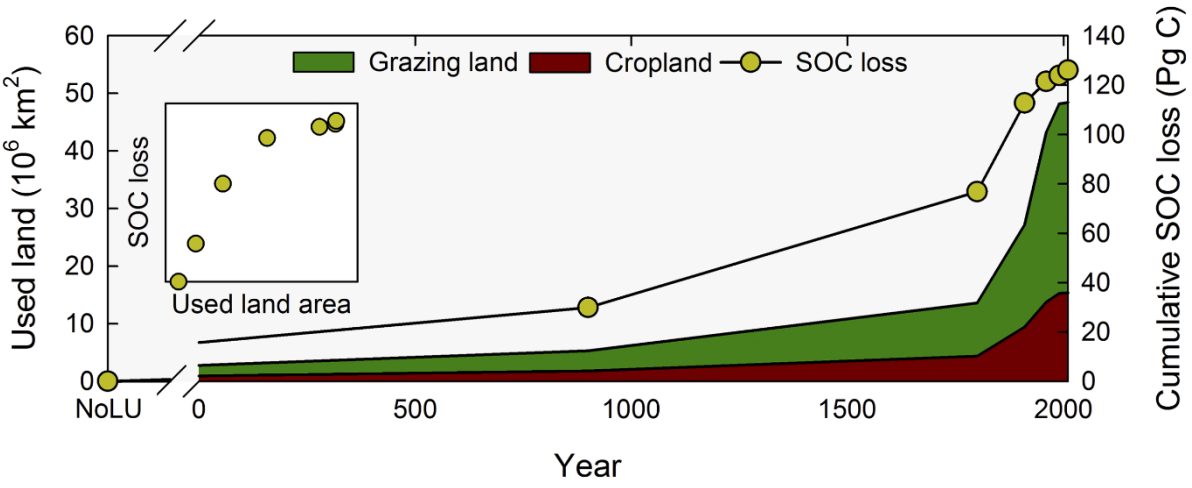


Figure 10: Historic evolution of conversion of undisturbed soils to grazing land and cropland and associated cumulative SOC loss (Sanderman et al., 2017)

1.6. Potential to reverse the SOC debt

The suggestion that this accumulated SOC debt can be reversed (Lal, 2004a) initially led to SOC and soil health receiving increased political and scientific attention (4per1000, 2018) but it has also attracted intense criticism (Paustian et al., 2016; Amundson and Biardeau, 2018). This idea is still being debated while various studies focus on different aspects of SOC sequestration potential involving biophysical, technical, economic or even societal limitations (Minasny et al., 2017; Amundson and Biardeau, 2018; Rumpel et al., 2018).

The optimistic scenario that best management practices can be achieved, allowing recovery of two thirds of the lost SOC corresponding to 88 Pg C (Lal, 2004a) is quite far from estimates provided by bottom-up approaches (taking into account that SOC equilibrium will be reached after 20–30 years), resulting in biophysical potential of SOC sequestration in the range of 8–28 Pg C (West et al., 2004; IPCC, 2006; Smith et al., 2008; Bossio et al., 2020). Others warn that even this scenario might not be achievable taking into account the current rise in global temperature and the increase in soil erosion rates (Smith et al., 2005; Amundson et al., 2015). Instead they suggest that the amount of carbon loss that can be reversed might be in the order of 10–30% at best or even below 10% (Sanderman et al., 2017).

A recent report specifically evaluating the potential of increasing SOC stocks and reaching the 4 per 1000 goal in France on a national scale based on extensive observational data and modelling showed that improving agricultural practices can offset ~41% of annual national agricultural emissions or 7% of total national GHG emissions (Pellerin et al., 2019). This study found a clear increasing effect of change in practices on SOC stocks, using observations from a network of long-term monitored sites with land use or land management changes or space for time substitutions at neighbouring plots (Fig. 11; left). The study assumed that the maximum potential of additional SOC storage was the difference between SOC stock of undisturbed soils (at equilibrium) and the current SOC stock of a cultivated soil. By definition, carbon additional storage would then depend on carbon inputs and the mean residence time of SOC (defined in Sect. 1.8.1.). Changing practices however can change both of these variables and will result in soils eventually reaching a new SOC equilibrium (after an infinite time period).

Predictions of the future of SOC under a given climatic and management scenario are uncertain due to lack of knowledge of site history, defining the initial conditions. Moreover, since most field trials are not longer than 20–30 years, it is difficult to validate model predictions outside of this range (Fig. 11; right side).

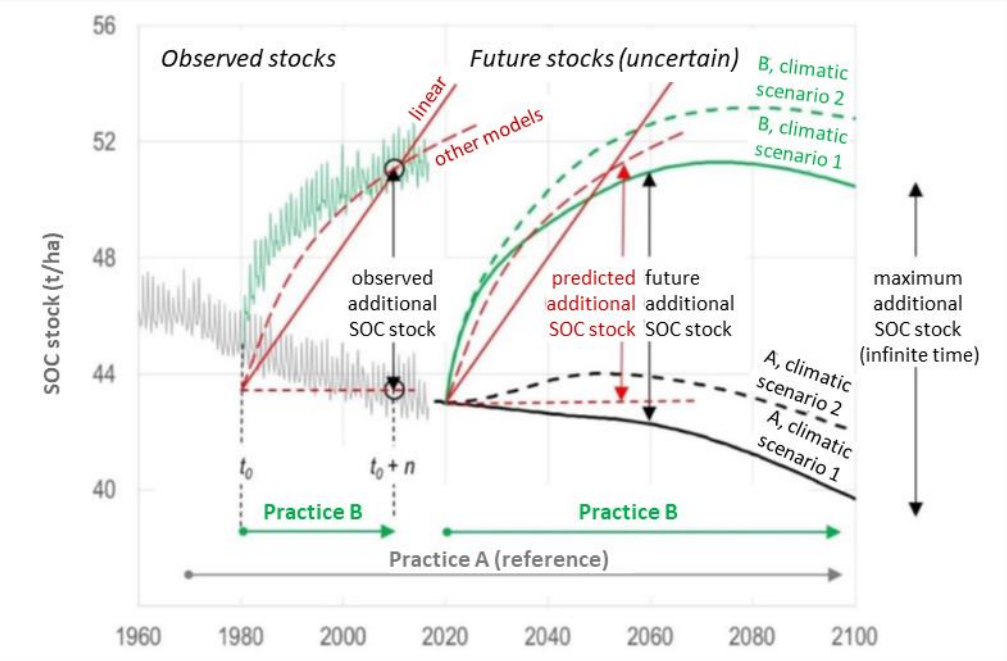


Figure 11: Potential of increasing SOC stocks by improving land management practices. Left side depicts observed changes in SOC stocks in field trials after change from a reference practice A to a new practice B. On the right side the graphs represent simulations of SOC evolutions under unchanged and new management practices for two climatic scenarios (translated from Pellerin et al., 2019).

Although the magnitude of the potential of increasing SOC stocks is uncertain, the benefits to soil functions, resistance to degradation and climate change mitigation are undeniable (FAO & ITPS, 2015; FAO, 2017). Climate smart soil management practices are not only an efficient way of removing CO_2 from the atmosphere but they also bring benefits to all soil functions including air quality, soil health, and biodiversity and are amongst the cheapest solutions currently available for promoting negative emissions (Griscom et al., 2017).

Still, developing appropriate strategies requires confidence in our ability to predict the evolution of SOC stocks through better assessment methods.

1.7. Monitoring SOC change: Measuring vs. modelling

Monitoring SOC stock changes in the field can be challenging. Not only do SOC stock changes occur slowly but the small magnitude of relative changes that are of interest for the exchange of CO_2 with the atmosphere compared to the large size of the SOC pool and the natural seasonal and spatial variation of SOC are difficult to detect (Smith, 2004). Thus carefully monitored long-term experiments (LTEs) maintained over multiple decades are of immense value for having accurate observations of the effect of management practices on SOC dynamics. The estimated required period for detecting a change in SOC stocks is between five and ten years (Smith et al., 2020), while the period needed for SOC to reach equilibrium with modified C inputs after a change in land use exceeds the lifetime of most LTEs (>50 years). In various parts of the world LTEs were often set up with the initial purpose of examining the effect of agricultural practices on crop yields. A few are even dating back to the 19th century (e.g., Rothamsted, UK), while in their vast majority they are found in the temperate zone. Although the geographical distribution of these sites is an important limitation (Smith et al., 2012), they provide a unique opportunity to test the accuracy of SOC models or to estimate SOC turnover rates and in extend calibrate models of SOC dynamics. In France, a prominent example of a collection of LTEs that was used in this study is the AIAL (Arvalis-INRA-Agro-Transfert-RT-LDAR) database, comprising all agricultural LTEs that started in France since 1970, including 455 treatments in a total of 53 sites (Clivot et al., 2019).

In the last decades, many countries have implemented soil monitoring networks (SMNs), with the purpose of obtaining reference soil data. Repeated soil sampling at constant locations after varying periods of time offers actual SOC stock change measurements over large spatial scales (Bellamy et al., 2005). However, land management and land use change is often not explicitly monitored in SMNs complicating the assessment of drivers of SOC change (Smith et al., 2007).

Such an example is the RMQS (Réseau de Mesures de la Qualité des Sols; Arrouays et al., 2002) network in France, with 2200 locations distributed in a 16×16km grid and a sampling frequency of 10–15 years. After multiple sampling campaigns SMNs can also be used to test SOC model performance. However, their use might be more interesting for statistical data-driven approaches combining the measurements on physical samples with spatial data (Luo and Viscarra-Rossel, 2020) or implementing space for time substitutions (Lugato et al., 2021) to infer environmental drivers of SOC change.

An important limitation of all field measurement data collection strategies is that their maintenance is costly. Collecting, analysing and storing physical samples requires significant funding and human resources, while even the most extensive networks will not cover all areas of the globe. Finally, although they are extremely valuable for establishing a baseline and collecting reference information, measurements alone cannot provide estimates for future changes in SOC (Smith et al., 2020). A detailed description of the various issues that still need to be resolved to have harmonized and comparable analytical and remote sensing data is presented in FAO (2017).

In contrast to physical measurements, modelling approaches are more flexible, cheaper and faster, but they also come with a greater uncertainty. Models of SOC dynamics are useful tools because they can provide estimates of SOC evolution on local or global scale. Nevertheless, their application has limits as well since they can only be applied in their range of calibration in terms of environmental, spatial and temporal dimension (Campbell and Paustian, 2015; Smith et al., 2020). Once a model has been correctly calibrated and sufficiently validated, it can be used to understand drivers of SOC change, or to estimate the impact of practices at unknown locations, or even be integrated into ecosystem models to link soil to the rest of the carbon cycle. Accurate operational models are indispensable for developing science-based strategies for better soil management (Eglin et al., 2010; Luo et al., 2016; Smith et al., 2020).

Combining analytical data and models is the only way to obtain robust simulations of SOC evolution, and improve our understanding of the dynamics controlling the evolution of this reservoir under rapidly changing conditions.

1.8. The Pandora box of SOC persistence

As mentioned in the previous section, neither measuring nor modelling SOC evolution can be considered as trivial. A characteristic of SOC relevant to its quality that is particularly difficult to assess and that contributes to the complexity of understanding SOC dynamics is its persistence. The spectrum in SOC persistence results from observed variations in its age, with various proportions having turnover times of only hours or days, while others persist over millennia (Trumbore, 1997, 2009).

1.8.1. So how old is SOC?

As for all other organic materials, radiocarbon dating can be used to estimate the mean age of soil organic carbon. This dating technique is based on the principle that during their lifetime plants will incorporate CO_2 from the atmosphere through photosynthesis, with the ^{14}C concentration of that time. Since ^{14}C is a radioactive isotope its decay can be quantified once that organism is dead. Knowing the concentration of ^{14}C in the atmosphere through time and so at the moment of death as well, it is possible to measure the remaining ^{14}C using mass spectrometry and calculate how long the organism has been dead for (Ellam, 2016).

Estimations of SOC age vary strongly across studies. Mean SOC ages of various size and chemical separation fractions of two French topsoils had SOC ages from less than 15 to 280 years (Balesdent et al., 1987). Using ^{14}C data on a soil formation chronosequence, Torn et al. (1997), modelled turnover times of SOC slower than 20 000 years. The average age of SOC of forest topsoils ranged between 200 and 1200 years (Trumbore, 2000). Topsoil SOC ages were estimated at 253 years in a grassland and at 340 years in an arable field topsoil (Jenkinson et al., 2008), while the global mean age of SOC in the top 1 meter was estimated at 3100 ± 1800 years (He et al., 2016), and at 4,830 ± 1,730 years (Shi et al., 2020).

Although it is evident that SOC is characterized by a wide spectrum of ages that can reach from very young to very old, using the ^{14}C technique on SOC has several limitations. A few are explained below but a more extensive discussion can be found in Trumbore (2009). First, estimates obtained with this method are representative of the mean age of a fraction or a bulk soil which are always composed of compounds with varying ages. This method is not able to predict the proportion of SOC that will be labile and the amount that will persist (Trumbore, 2000). Second, the mean age of SOC is not necessarily equal to its turnover[5] or its residence time[6]. Because soil is an open system and because biotransformation of SOM components is constantly taking place, recycling of carbon means that old organic matter in terms of ^{14}C can be present in a labile form and decompose fast (Gleixner et al., 2002; von Lützow et al., 2007; Sierra et al., 2014). Even though SOC with an older age is considered to have a slower turnover, as soil is an open system whose composition is changing with time, this is not an exact approximation (Six and Jastrow, 2002).

Despite the uncertainties associated with this method, ^{14}C ages clearly show that organic matter that should theoretically be thermodynamically labile can persist in the soil for thousands of years.

1.8.2. What explains SOC persistence?

The question can be viewed from two different perspectives: Either focusing on the conceptualization of the stabilization mechanisms that could be at play or on the environmental drivers that can be inferred empirically.

Since in the following section we discuss stabilization mechanisms it is important to differentiate between the two terms: persistence and stability/stabilization. While persistence is the resulting effect of interactions taking place in soil and a characteristic that can be inferred from SOC ages and field observations from long-term experiments, stability or stabilization on

[5] In systems at steady state (inputs=outputs), the turnover time is equal to the average age of the elements leaving the system, defined as the ratio of the total SOC pool size to the carbon outputs or inputs (Luo et al., 2019).

[6] Mean residence time is an umbrella term used in literature to describe either the turnover time or the mean age of SOM (Sierra et al., 2017).

the other hand is a much less clearly defined term. As we will see below it depends on many aspects and it is difficult to understand and even more so to quantify.

1.8.2.1. SOM persistence explained through conceptual mechanisms

Even though the debate on the main mechanisms and processes influencing SOC persistence is an old one, it is still an unresolved matter over which new contrasting opinions are constantly emerging. Until recently, three categories of SOM stabilisation mechanisms were considered: (1) increasing OM recalcitrance through selective mineralisation of more biodegradable litter inputs (2) physical protection, e.g. through the formation of aggregates, and (3) physico-chemical protection through the formation of organo-mineral associations (Sollins et al., 1996; Six et al., 2002; von Lützow et al., 2006; Campbell and Paustian, 2015)

1.8.2.1.1. Chemical recalcitrance of SOM

Early approaches aiming to study SOM fractions using chemical acid-base separation methods led to the development of the theory that complex humic acid macromolecules existed that were more stable and to the evolution of the humification model (e.g., Hayes, 1986). However, advancing SOM characterisation techniques have questioned the existence of these molecules in the soil environment and showed that the extracted humic substances were in reality a result of coagulation from smaller particles due to the experimental conditions (Schulten and Leinweber, 2000; Weil and Brady, 2016). The theory that inherent chemical recalcitrance of SOM is the main controlling mechanism responsible for its stabilization was discarded especially after the study by Schmidt et al. (2011). The authors present mean ages of different components of SOM showing that turnover of complex (e.g., lignin) and simple (e.g., glucose) molecules is not significantly different (Fig. 12). One exception is pyrogenic carbon showing consistently older ages than bulk SOM. Further studies support the idea that the soil decomposer community has the ability to mineralize SOC present in all natural compounds (e.g., Dungait et al., 2012) and suggest that other physico-chemical factors such as the ability of OM to form associations with minerals will define its persistence.

1.8.2.1.2. Physical protection of SOM in aggregates

The second classic mechanism proposed for the stabilization of SOM is the existence of a physical barrier between the decomposer community and SOM in cases where the latter is occluded in soil aggregates (e.g., Oades and Waters, 1991; Six et al., 2000). According to this theory aggregation can limit the enzymatic accessibility and obstruct degradation reactions. Especially the role of microagggregates (<50µm; Virto et al., 2008) is highlighted. One aspect limiting the ability of this mechanism to account for overall SOM persistence is the non-stability of the aggregates themselves, which are regularly destroyed and reformed in the soil (Baldock and Skjemstad, 2000; Virto et al., 2010). This would suggest that SOM protected in aggregates would partially be susceptible to decomposition. Moreover, a difficulty lies in our ability to study aggregates and differentiate between the role of aggregation and physico-chemical interaction of OM and minerals present in the clay fraction.

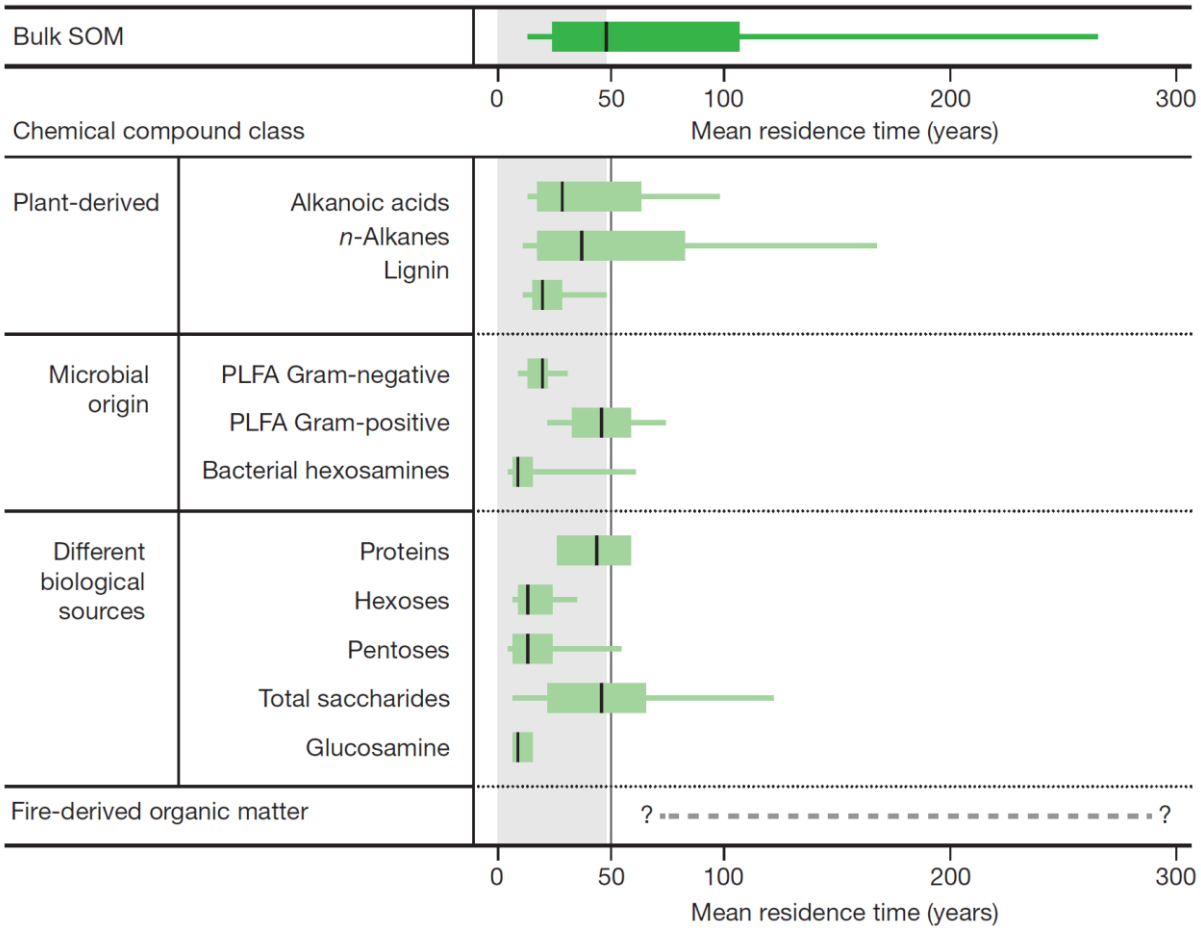


Figure 12 : Mean turnover times of various organic compounds found in SOM. The control of molecular structure on long-term decomposition of SOM is not significant, except in the case of pyrogenic carbon (Schmidt et al., 2011).

1.8.2.1.3. Physico-chemical protection of SOM through interaction with mineral surfaces

The third mechanism of SOM stabilization is its adsorption on mineral surfaces. The formation of a variety of bonds (e.g., covalent, complexation, hydrogen bonds, …) between SOM and minerals present in the clay soil fraction (<2µm) can limit its accessibility to microorganisms from an energetic point of view. This theory suggests that the binding energy will correspond to an additional amount of energy that a microorganism will have to spend to consume that particular molecule of SOM (e.g., Chenu and Stotzky, 2002; Kaiser and Guggenberger, 2003; Kleber et al., 2015; Hemingway et al., 2019; Mikutta et al., 2019). The existence of a relationship between the size of the clay fraction and SOM mean age was confirmed in multiple studies (e.g., Torn et al., 1997; Paul et al., 2001). Moreover, the mineralogy of the clay fraction is considered important, with metal oxy-hydroxides, reactive phyllosilicates and poorly crystalized alumosilicates amongst the most efficient for SOM protection (Mikutta et al., 2007; Sanderman et al., 2014; Rasmussen et al., 2018).

In addition, characteristics of the organic matter (e.g., functional groups, molecular size) will define its ability to participate in adsorption or other association reactions. Smaller molecules, often residues of microbial metabolic reactions, that are more oxidized and more reactive can be increasingly associated to minerals (Kleber et al., 2015; Pellerin et al., 2020).

The observation that SOM consists of organic residues of different sizes, at different decomposition stages and in different molecular groups, combined with the synergy of all of the above described mechanisms resulted in a complex view of overall SOM controls (Fig. 13; Lehmann and Kleber, 2015). The mechanical fragmentation of residual plant material by macrofauna such as earthworms and termites creates smaller SOM particles, but also the selective consumption of OM compounds by these organisms releases new metabolic products. These can either be stabilized or further decomposed by microbes. Further biotransformation reactions result through uptake of OM by microorganisms present in soils (fungi and bacteria) and their metabolic activity. Microbial biomass can degrade but also synthesize new organic matter components. Finally, SOM is constantly undergoing biotransformation reactions, transfers and stabilization and destabilization reactions.

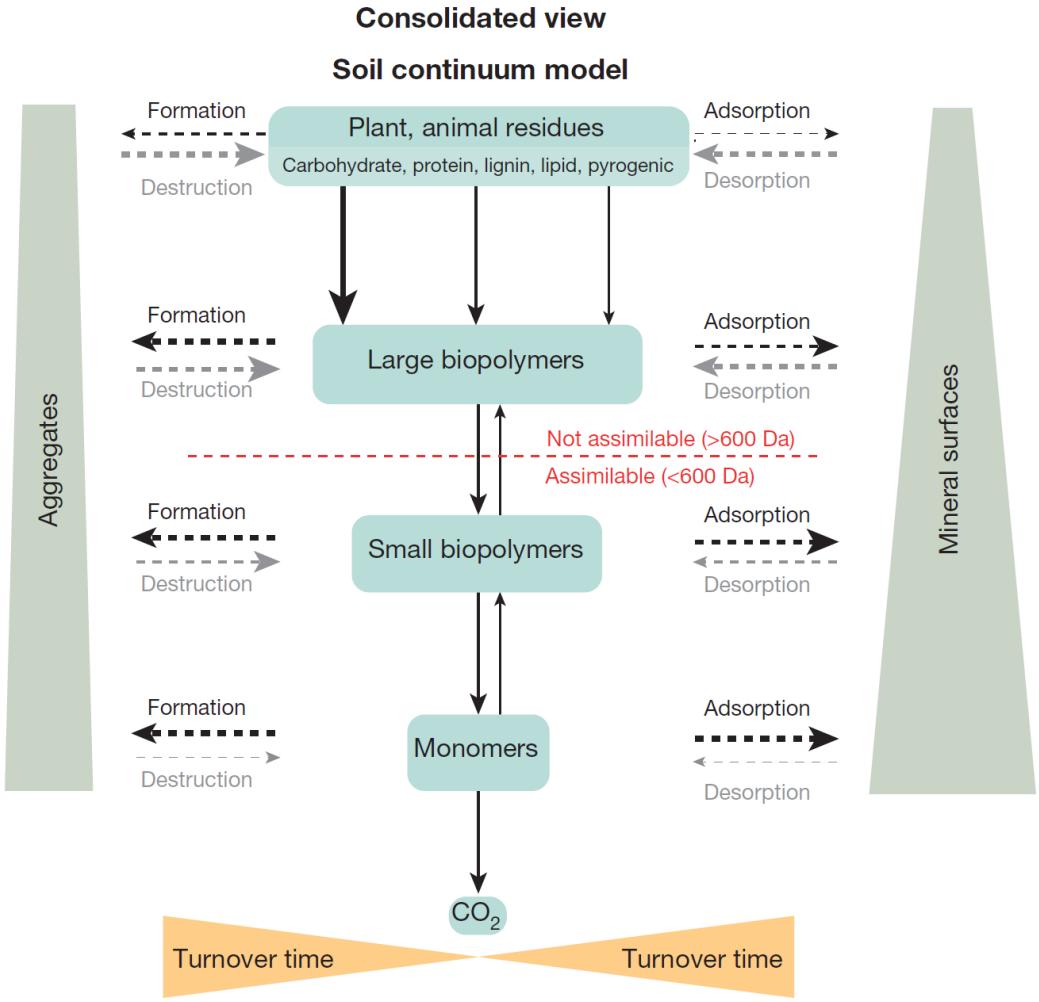


Figure 13: Consolidated view of the SOM continuum model explaining SOM turnover as a result of constant biotic and abiotic interactions (Lehman and Kleber, 2015).

1.8.2.1.4. Emerging views of SOM stabilization

Evolution of imaging techniques at the molecular level offering information on the distribution of SOM has inspired the emergence of new ideas regarding the conceptualization of processes driving SOM stability. The common point of these views is that they put the activity of soil organisms in focus. An emerging theory considers SOM resistance to degradation as a result of functional complexity (Lehmann et al., 2020). For this idea, the 3-D structure of soil matrix, the habitat heterogeneity, and the microbial diversity will control the mineralization of SOM. The observation that organic matter distribution at the nano-scale occurs in patches and the idea that microbes have a limited motility, has been proposed as a different definition of accessibility as the main control of SOM mineralization (Ekschmitt et al., 2008; Nunan et al., 2020).

Other studies go even further to suggest that activity of soil organisms might even control its structure, proposing that the chemical and physical changes it generates (e.g., exudation of gluing agents, rearrangement of soil material through root activity) will define the formation of aggregates (e.g., Rillig and Mummey, 2006).

Trying to predict the behaviour of the heterogeneous community of microorganisms that reside in soils, a complex environment by definition, is not an easy task. Therefore, for the incorporation of SOM mineralization controls in modelling frameworks a different approach is to focus on observed relationships between pedoclimatic characteristics and SOC stocks distribution.

1.8.2.2. Drivers of SOM turnover

Factors that have been shown to influence SOM turnover as well as the distribution of SOC stocks include temperature, soil water content, pH, texture and mineralogy, as well as the type of SOM input and soil depth (Balesdent et al., 2018; Luo et al., 2019; Luo and Viscarra-Rossel, 2020). These factors are in essence related to the above described mechanisms, as they have a strong influence on their efficiency. The multitude of effects of the different pedoclimatic factors on processes taking place in soils was summarized by Pellerin et al. (2019). Here we briefly mention a few examples. Temperature is a major factor defining SOC turnover mainly through favouring or hindering microbial activity. However, temperature will also play a role on the formation of adsorption and on the efficiency of diffusion processes. Another important factor is soil pH, also controlling microbial activity and the formation of organo-mineral associations since it can influence the surface charges of minerals. The bulk chemistry of SOM entering the soil will not only influence its bioavailability but it will also have a control over its interaction with mineral surfaces.

In the next section we will see how the evolution of theories regarding SOM stabilization has affected the evolution of models of SOC dynamics and how different models take into account drivers of SOM turnover.

1.9. A quick timeline of SOC models

The inherent complexity of the soil system and the variety of SOM controlling processes have led to a wide spectrum of hypotheses and conceptualizations for governing mechanisms of SOM turnover. Additionally, the need to represent the system at different temporal and spatial scales resulted in an important number and variety of models available today.

Here we draw a quick timeline that is in no way exhaustive. Its purpose is to provide a general overview with a few examples of the existing types of SOM models and their origin.

In many reviews (Mcgill, 1996; Feller and Bernoux, 2008; Bleam, 2017) one of the first mentions of a mathematical SOM modelling attempt is the simple soil organic nitrogen turnover model proposed by Jenny (1941), developed to evaluate the impact of agricultural practices on soil fertility. Observations from long-term agronomic experiments in central North America (Salter and Green, 1933) had shown that soil N dynamics can be described as an exponential decay function with different rates in different soils, or under different cropping systems. However, projecting the evolution of soil N at the scale of a century, with the assumption that N content will not reach zero but rather an equilibrium value, required the use of an additional term defining N fixation, thus resulting in the following equation:

$$\frac{dN}{dt} = -kN + A$$

where dN = change in soil organic N, dt = simulated period, k = first order decay rate constant (t^{-1}), and A = annual addition rate of soil organic N (in the form of organic matter; in mass t^{-1}).

The equation can be solved to calculate soil N content after a given period as:

$$N = N_E - (N_E - N_0)e^{-kt}$$

where N_0 = the initial amount of soil organic N, N_E = the amount of organic N in the soil at equilibrium (whereas at equilibrium: dN/dt = 0 thus N = N_E = A/k).

The model by Jenny was successfully used to describe soil organic N dynamics in different conditions (Woodruff, 1949; Jenkinson, 1990). It was also used in many studies with the objective of fitting long-term observations of N evolution to estimate model parameters such as the mineralization rate (k) and mean turnover time (1/k) (mentioned in Jenkinson, 1990).

Meanwhile, the scientific community and farmers gradually shifted their attention to total SOM and SOC dynamics because of their importance for soil quality. Unlike soil organic N whose evolution is mainly controlled by the input of fresh organic matter to soil, SOM dynamics are more complex. The most prominent difference is probably the heterogeneous nature of SOM containing components with highly variable turnover times. As the various components decompose at different rates, their relative proportions are bound to change over time. To accurately simulate SOM or SOC evolution this spectrum of stabilities has to be taken into account. Mathematically, this translates to a change in the decay rate (k) with time.

The different strategies for representing this change in k in SOM simulations, as enlisted in the review by McGill (1996), i.e., either adding more components or making k time-dependent, led to the different types of models available today. Here we distinguish two main categories: multi-compartment models and non-compartmental models, while taking a closer look at the evolution of multi-compartment models through time. A complete classification would require mentioning sub categories and hybrid models and is not conducted here.

1.9.1. A two-compartment model

An important example resulting from the idea that more components are necessary for correctly simulating SOC evolution is the model of Hénin and Dupuis (1945). They were the first to split organic matter into two compartments: organic residues and soil carbon (humified C). As organic residues (m) are returned to the soil each year, a fraction is rapidly mineralized ($m \cdot (1-k_1)$) and escapes into the atmosphere as CO_2, whereas the remaining organic material ($m \cdot k_1$) is “humified” and enters a more slowly cycling pool (humified C) that is also subject to mineralization ($k_2 \cdot C$; Fig. 14).

This model can be described by the following differential equation and its integrated form:

$$\frac{dC}{dt} = -k_1 \cdot m - k_2 \cdot C$$

$$C = \frac{k_1 \cdot m}{k_2} + \left(C_0 - \frac{k_1 \cdot m}{k_2}\right) e^{-k_2 t}$$

where C = SOC content, k_1 = humification rate, m = fresh organic matter (in the form of organic residues), C_0 = SOC content at the onset of simulated period, and k_2 = mineralization coefficient (whereas at equilibrium: $C = C_E = k_1 \cdot m/k_2$). The humification rate depends on the type of organic matter (lower for fresh straw material and higher for manure) and the mineralization rate depends on climatic conditions and soil characteristics.

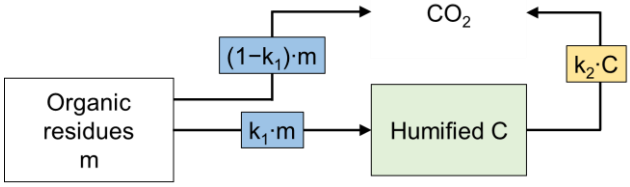


Figure 14: Conceptual representation of the simple two-compartment Hénin-Dupuis model. The model considers two organic matter compartments, organic residues (m) and humified organic carbon (C). Fluxes are expressed by arrows. Incorporation of fresh organic matter into soil is controlled by the humification coefficient k1 and loss of humified carbon through mineralization is controlled by the mineralisation coefficient k2 (modified after Mary and Guerif, 1994).

The Hénin-Dupuis model was used in numerous studies over the following decades to evaluate agricultural practices (Jenkinson and Rayner, 1977; Jenkinson and Johnston, 1977; Boiffin et al., 1986; Jenkinson et al., 1987). This simple adaptation led to important improvements of SOC projections compared to the model by Jenny. Some decades later however, several studies argued that the two-compartment model structure was inadequate and that SOM should be divided into at least two fractions, each with a different turnover rate (Janssen, 1984; Andrén and Kätterer, 1997).

1.9.2. Multi-compartmental models with conceptual kinetically defined pools

Since the 1980s, an approach valuing mathematical simplicity and based on the example set by Hénin and Dupuis led to the development of some of the simplest and most widely used models of SOC dynamics. In these models, the continuum of SOC decomposition is represented by a series of kinetically defined SOC pools characterized by different turnover times.

Splitting of SOC into pools, allows multi-compartmental models to use first-order kinetic laws, i.e., the flux of material from a pool and the quantity of material it contains are linked by a simple linear relationship. Amongst the most important advantages of these models, intrinsically related to their simple structure, are their transferability, calculative efficacy and adequate parametrization. This category includes world-renowned models such as Century (Parton et al., 1987) and Roth-C (Jenkinson and Rayner, 1977), and well parameterized mostly nationally used models such as the Swedish ICBM (Andrén and Kätterer, 1997), the French AMG model (Andriulo et al., 1999) or the Danish C-TOOL model (Taghizadeh-Toosi et al., 2014).

Century is a prominent example of a conceptual SOC model, developed in the United States to model SOC dynamics in the Great Plains (Parton et al., 1987). Initially used to evaluate the impact of crop production and management practices on SOC, it has the potential of linking plant growth and SOC evolution. Its SOC module however is quite simple with only two actively cycling SOC pools, the active and the slow pool, with turnover times of 1.5 and 25 years respectively. The model also has a passive pool with a very slow turnover (1000 years) that can be considered as inert for simulations at the decadal time scale.

The second most widely used model of SOC dynamics (according to the review of Manzoni and Porporato, 2009) is RothC. Since its development in the 1990s at the Rothamsted Experimental Station in England (Jenkinson and Rayner, 1977), the RothC model has been continuously adapting to model SOC dynamics in grassland and in forest soils and it was calibrated and applied on different soil types and climates. Similar to Century, RothC divides SOC into two actively cycling and one stable pool. Pools are defined according to their composition, namely microbial biomass (BIO) with a turnover time of 1.5 years, humus (HUM) with a turnover time of 50 years and the purely stable inert organic matter (IOM) pool

representing black carbon[7]. The turnover process in the RothC model is controlled by soil type, temperature, soil moisture and plant cover. Inputs into the soil occur by a fraction of fresh organic matter, the resistant plant material (RPM), while a more labile fraction of decomposable plant material (DPM) is directly lost to mineralization (Fig. 15).

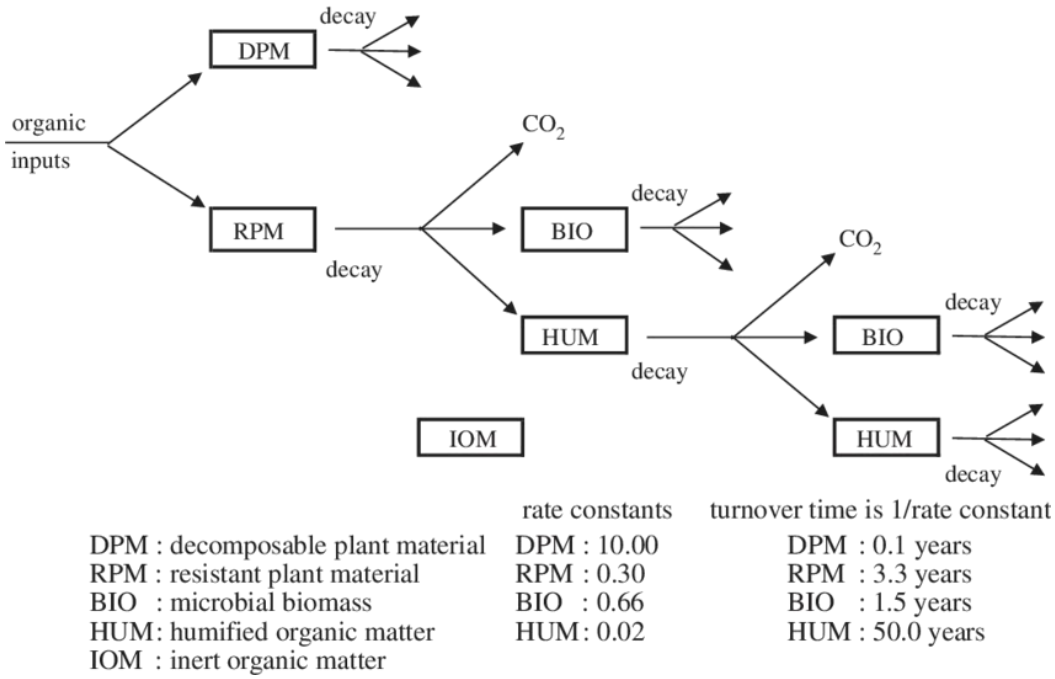


Figure 15: Conceptual diagram of the RothC model, including pool structure, fluxes, and turnover times (Coleman and Jenkinson, 1996).

Simple multi-compartmental models have been used in many climate zones and settings and they have served as operational predictive tools representing SOM in Earth System Models (ESMs). This wide geographical expansion and range in terms of applications was only possible because of the simple structure of these models, allowing sufficient parameterization, and mathematical and computational simplicity. Although the usefulness of these models to predict SOC evolution especially over the long-term has been proven through comparison with empirical data, they are currently being strongly criticized. Mainly, because their pools are only conceptual and kinetically defined and are thus difficult to measure and because they do not explicitly consider microbial processes.

[7] Black or pyrogenic carbon is a residue of incomplete combustion found in soils affected by fire

1.9.3. Mechanistic process-based models

Recent advances in analytical methods enabling precise imaging and chemical characterization of organic matter in undisturbed soil samples, have led to new understanding of processes involved in SOM stabilization (discussed in the previous section; (Schmidt et al., 2011; Lehmann and Kleber, 2015; Kögel-Knabner, 2017; Lehmann et al., 2020). The central role of microorganisms for the formation and degradation of SOM has inspired a new generation of models aiming to explicitly represent microbial interactions (Cotrufo et al., 2013; Wieder et al., 2015; Abramoff et al., 2018; Robertson et al., 2019). This has re-sparked interest in a debate that took place in the 1970s as the first microbial process-based models (Smith, 1979; Mcgill et al., 1981) were being developed in parallel to the conceptually simpler ones. One objective of many new generation models is to transition towards more "measurable" pools representing functionally meaningful SOM fractions. However, this often leads to more complex model structures, and a higher number of unknown model parameters to constrain (e.g., Millennial model; Fig. 16). Although new generation models might provide a more elegant representation of the processes controlling SOM dynamics, they also have the major drawback that they are very difficult to calibrate, validate against field data and scale-up (Sulman et al., 2018). This hinders particularly their potential to be used as operational tools to predict SOC evolution at the large scale.

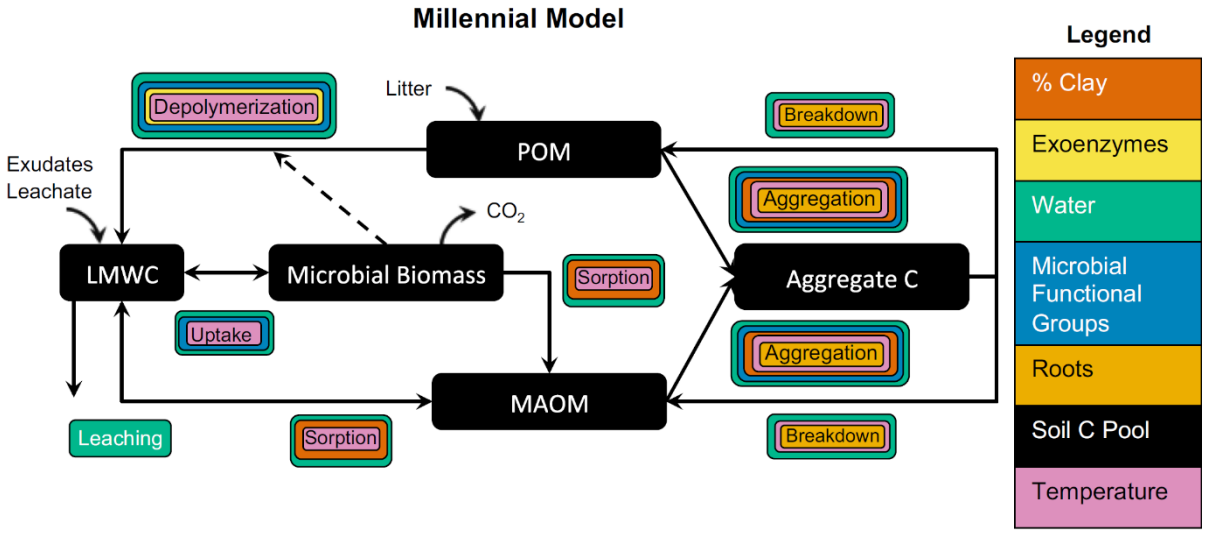


Figure 16: Conceptual diagram of the Millennial SOC model, including compartments (black), structure and fluxes (arrows and coloured boxes). The colour legend represents controlling biotic and abiotic factors considered in the model (Figure changed after Abramoff et al., 2018).

1.9.4. Non-compartmental models

In parallel to the evolution of multi-compartmental models, a second category of models neither splitting SOM into compartments nor applying first order kinetics were being developed as well since the 1980s. Originally proposed by (Carpenter, 1981), these models focus on the idea that SOM decomposition is a continuum (e.g., Q-model; Ågren and Bossata, 1996; Bruun et al., 2010). They represent SOM with all its heterogeneous compounds as one pool whose quality however changes over time. The issue of defining conceptual pool sizes and individual turnover rates is reformulated instead to defining patterns of mineralization dynamics. Since SOC decay in these models is not constant over time and is influenced by many factors these models are more complex to constrain (Jenkinson, 1990).

So which is the best model?

It has been clear to modellers that different types of models exist for different purposes (Mary and Guérif, 1994; Mcgill, 1996). It is important to recognize that models are developed to test specific hypotheses relevant at a specific spatial and temporal scale (Campbell and Paustian, 2015). Each model can only be evaluated and applied in the range in which its parameters were calibrated and its structure was tested. The current trend of increasing structural complexity and non-linearity in SOC model development (Manzoni and Porporato, 2009) is also accompanied by an increasingly difficult parametrization and transferability. As put by Blankinship et al. (2018) a robust SOM modelling approach must integrate three aspects: emerging understanding (theory), analytical measurements (data), and mathematical representation (numerical modelling). Without the necessary data to inform and evaluate a model it is impossible to know how well a model functions and use it to predict changes in the modelled system. In conclusion, there is no silver bullet solution but for a model to be operational a balance between available quantitative data and numerical model structure is key. The choice of the most appropriate model should be taking into consideration all of the points mentioned above.

In this work we chose to work with AMG (Andriulo et al., 1999), a simple multi-compartmental model of SOC dynamics widely used in France. The AMG model is described in detail in the next section, but briefly comparing its model structure to the ones of RothC and Century, we can see that it is similar, except the fact that it ignores the existence of the most labile carbon pool (residence time less than 1 year). All three models consider one pool to be stable at the pluri-decadal scale (<100 years), although it varies in size (~ 10% of SOC in RothC, 40–50% in Century, and 40–65% in the AMG model).

1.10. The French AMG model — Development, parametrization and application

Since its development in 1999 by Andriulo, Mary and Guérif (Andriulo et al., 1999), the AMG model has seen a wide use in France and worldwide for evaluating the impact of agricultural practices on SOC stocks in croplands (Saffih-Hdadi and Mary, 2008; Bouthier et al., 2014; Autret et al., 2016; Martin et al., 2019; Levavasseur et al., 2020; Nowak and Marliac, 2020). Throughout the decades the model has been constantly adapting to include important drivers according to current understanding and new concepts emerging in soil science (Saffih-Hdadi and Mary, 2008; Clivot et al., 2019; Mary et al., 2020).

In this section a description of the current version of the model is provided, followed by a short review of important works regarding its development and parametrization and finally some prominent examples of its application.

1.10.1. Model description

The AMG model operates at an annual time step. It is characterized by a simple structure consisting of three carbon pools: fresh organic matter, and two SOC fractions, an active and a stable pool (Fig. 17). The model allows transfer of carbon from the fresh organic matter pool either to the atmosphere through microbial mineralization or into the active pool. Organic carbon from the active pool is also subject to mineralization, forming a second direct flux of CO_2 from the soil into the atmosphere. SOM decomposition follows first order kinetics with a rate defined by the coefficient of mineralization k (year^{-1}), controlled by climatic conditions and soil characteristics. The h coefficient controls the yield of crop residues transformation into active carbon and depends on the type of fresh organic matter. No carbon exchange with the stable SOC pool is possible since it is considered inert and remains unchanged over the simulation period (typically several decades). Mathematically the AMG model can described by two simple equations (Clivot et al., 2019).

$$QC = QC_S + QC_A$$

$$\frac{dQC_A}{dt} = \sum_i m_i h_i - k \cdot QC_A$$

where QC is the total SOC stock ($t \cdot ha^{-1}$), QC_S is the stable SOC stock ($t \cdot ha^{-1}$) defined as a fraction of initial SOC stock QC_0 constant for a specific treatment, QC_A is the active SOC stock ($t \cdot ha^{-1}$), t is the time in years, m_i is the annual C input from organic residue i ($t \cdot ha^{-1} \cdot yr^{-1}$), h represents the fraction of C inputs which is incorporated in SOM after 1 year, and k is the mineralization rate constant associated with the active C pool (yr^{-1}).

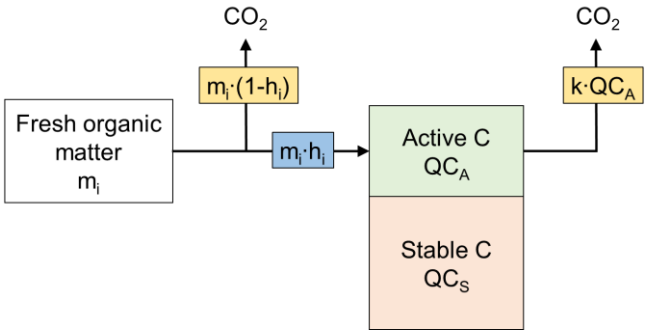


Figure 17: Schematic representation of the pool structure and fluxes of the AMG model. A fraction (1-h) of fresh organic matter (m) is yearly mineralized and released in the atmosphere, whereas a fraction (h) is incorporated into the active SOC pool (C_A). The coefficient of mineralization (k) controls carbon discharge from C_A into the atmosphere, while no exchange is possible with the stable SOC pool (C_S) (modified from Duparque et al., 2013 and Levavasseur et al., 2020).

Moreover, the model can simulate simultaneously the evolution of C_3 and C_4 stocks (expressed below as QC_3 and QC_4) according to the equation proposed by Balesdent et al. (1987):

$$QC_3 = \frac{\delta^{13}C_S - \delta^{13}C_4}{\delta^{13}C_3 - \delta^{13}C_4} \cdot QC$$
$$QC_4 = \frac{\delta^{13}C_S - \delta^{13}C_3}{\delta^{13}C_4 - \delta^{13}C_3} \cdot QC$$

where $\delta^{13}C_S$ is the measured ^{13}C isotopic ratio in the soil, and $\delta^{13}C_3$ and $\delta^{13}C_4$ are the isotopic compositions of C_3 and C_4 plants in the crop rotation.

The mineralization coefficient k of the active SOC pool is calculated according to a potential mineralization rate k_0 and the cumulative effect of pedological and climatic characteristics including mean annual temperature (T; °C), soil moisture expressed as the difference between cumulative annual water inputs (through precipitation and irrigation) and potential evapotranspiration (H; mm), clay content (A; g kg^{-1}), calcium carbonate content ($CaCO_3$; g kg^{-1}), soil pH and C/N ratio:

$$k = k_0 \cdot f(T) \cdot f(H) \cdot f(A) \cdot f(CaCO_3) \cdot f(pH) \cdot f(C/N)$$

Graphical representations of the non-linear functions used in the above equation are shown below (Fig. 18; Clivot et al., 2017). Their fitting parameters and threshold values can be found in Clivot et al. (2019).

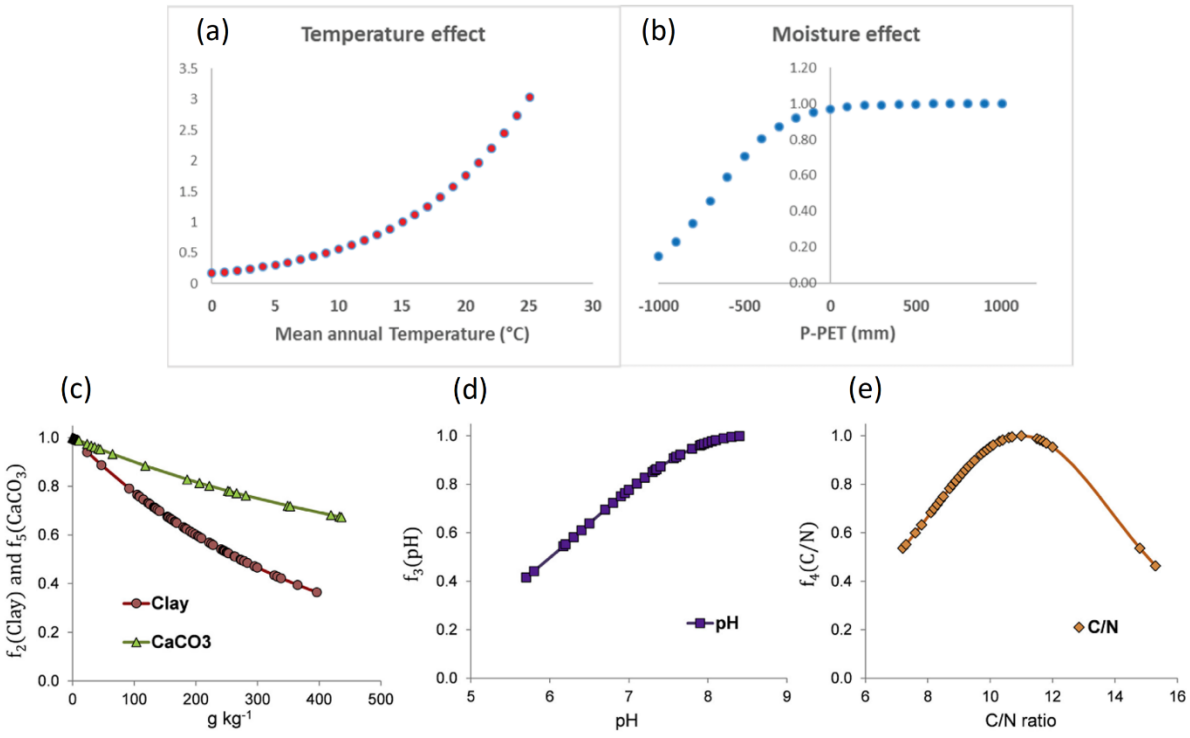


Figure 18: Set of six non-linear functions showing the dependence of the mineralization rate k of the active carbon pool C_A on environmental parameters. (a) Mean annual temperature, (b) soil moisture, (c) soil clay and $CaCO_3$ content, (d) soil pH, and (e) soil C/N ratio (Plots (a) and (b) from Mary and Ferchaud personal communication; Plots c-e from Clivot et al., 2017)

The model does not include a module for the calculation of carbon inputs but a detailed method was developed and is provided in Clivot et al (2019). The authors propose using allometric coefficients from Bolinder et al. (2007) for estimating the distribution of carbon in the different parts of a plant, relating harvest index measured in the field with corresponding above and below carbon inputs for each crop (Fig. 19). The accumulation of carbon to a specific depth is taken into account in the method by Clivot et al. (2019) by an asymptotic equation proposed by Gale and Grigal (1987) and according to differences in root distribution of various crops (Fan et al., 2016).

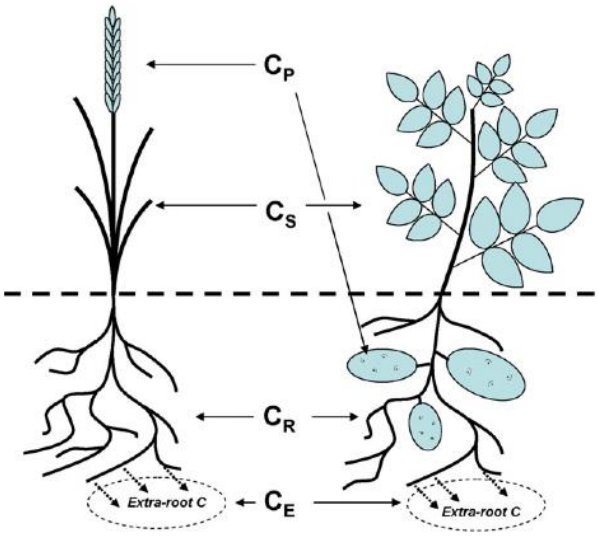


Figure 19: Schematic diagram of the distribution of carbon in the different parts of a plant. CP is the carbon in the harvested product, CS is the carbon in the aboveground residue, CR is the carbon contained in the roots and CE is the carbon extracted by roots (from Bolinder et al., 2007).

Finally, in the method by Clivot et al (2019) humification coefficients for aerial parts of plants returned to the soil as organic residues can be determined using their specific C/N ratio (Machet et al., 2017) and functions defined in Justes et al. (2009). Humification coefficient for belowground inputs is considered uniform and is calculated according to data published in Balesdent and Balabane (1996) and Kristiansen et al. (2005). A former study conducted with AMG inverse modelling of field observations and incubation experiments provides humification coefficients of various exogenous organic matter amendments (Bouthier et al., 2014).

1.10.2. Milestones in the development, parametrization and application of the AMG model

Important works in the development and parametrization of the AMG model begun with the study of Mary and Guérif (1994) attempting to model SOC dynamics in different treatments of the long-term experimentation sites of Rothamsted (UK) and Grignon (FR), using the Hénin and Dupuis (1945) two-compartment model. The authors demonstrated that the model was unable to reproduce SOC evolution at contrasting treatments (in terms of organic matter inputs) of the same site using a constant mineralization coefficient. In other words, from a point of view regarding its dynamic evolution, SOM had to be heterogeneous. The simplest approach to represent this heterogeneity in a model was to divide SOM into an active and a biogeochemically stable pool, or very slowly cycling pool (Mary and Guérif, 1994). Based on observations from the Rothamsted long-term field experiment, they estimated that about one third of SOC should be in the active pool and they showed that the three-compartment model significantly improved simulations of long-term SOC stocks evolution at the studied sites.

This adaptation of the Hénin-Dupuis model became the AMG model, named after Andriulo, Mary and Guérif (1999) who tested and calibrated the model to track the evolution of SOC stocks under crop rotations in Argentina. The monitored trials included changes from C_3 (wheat or soybean) to C_4 (maize) plants, offering the opportunity to follow the evolution of young and old carbon separately by measuring the natural ^{13}C abundance change over time. The authors showed that the size of the stable SOC fraction ranged between 60–68% of the initial SOC stock, that the humification coefficient depended on type of organic matter input and that the mineralization rate was sensitive to management practices (tillage).

The specific 65% estimate for the stable SOC proportion still used today was obtained for the first time in a study by Mary and Wylleman (2001). The authors followed a similar methodology taking advantage of ^{13}C data series to fit the model on observations from a long-term field experiment at Boigneville in France.

With the establishment of an estimate for the pool partitioning Saffih-Hdadi and Mary (2008) focused on the dependence of the mineralization rate on environmental factors, namely temperature and clay content. They borrowed a function describing the control of temperature on the mineralization rate proposed by Balesdent and Recous (1997), obtained from incubation experiments, and adapted to field conditions by Mary et al. (1999). For a given temperature range the temperature effect increased exponentially up to a threshold value after which a slower increase occurred. The effect of clay content on the mineralization rate was described by an exponential relationship according to observations from incubation experiments (Brisson et al., 2003). Although these functions would later be fine-tuned by Clivot et al. (2017, 2019), the authors showed that the model reproduced well the effects of straw residue export on SOC stocks in various pedoclimatic conditions (nine sites in Sweden, Denmark, France and Thailand). However, they argued that optimum values of stable SOC proportion could be lower (down to 40%) in some cases with varying in C inputs.

The current version of AMG (Clivot et al., 2019), was developed by adding four additional environmental functions affecting the mineralization rate of the active carbon pool. The relationship between pedological characteristics (i.e., $CaCO_3$ content, soil pH, and C/N ratio) and in-situ nitrogen mineralization was estimated by Clivot et al. (2017) at 65 field experiments under bare fallow (France). The obtained functions were incorporated in the SOC model that was evaluated against data from 60 long-term field experiments in 20 sites in France (Clivot et al., 2019). As described above, this study included an important amount of work for defining carbon inputs based on experimental data, thus allowing the optimization of the potential mineralization rate k_0, resulting in a value of 0.29 yr^{-1} for the network of studied sites.

The final above described version of the AMG model has been repeatedly used in the last years in France and in Europe and its performance to reproduce observed changes in SOC stocks was shown to be satisfactory. Comparative studies against other multi-compartmental models (such as RothC and Century; Martin et al., 2019; Bruni et al., 2021; Farina et al., 2021) confirm that the huge amount of parametrization work and the simple and efficient model structure of AMG make it a top candidate for use in cropland.

The model was used to simulate the evolution of SOC stocks at seven sites in Europe (Denmark, England, France and Sweden) with repeated exogenous organic matter application (Levavasseur et al., 2020) and performed well. Recently, the priming effect was incorporated in the model (Mary et al., 2020) and the evolution of SOC stocks in a 47-year experiment was improved in the topsoil layers of different tillage treatments. In a French national project designed to evaluate current inventory methods for croplands (CSOPRA) the AMG model was used alongside RothC, Century and ORCHIDEE, to simulate SOC stock evolution at long-term experiments (AIAL database) and to predict SOC stocks at sites of a soil monitoring network (RMQS). Overall, the study showed that AMG performed best, providing the least biased predictions and the lowest errors for the selected sample set.

Finally, in a work in preparation (Bruni et al., 2021), in an effort to evaluate the necessary carbon inputs to achieve the yearly 4‰ increase in SOC stocks goal set by the homonymous initiative, the authors compare predictions provided by six models (Century, RothC, ICBM, AMG, MIMICS and Millennial) at 17 sites in Europe. Amongst the used models, AMG was shown to simulate most accurately the changes in SOC stocks over time when all models are applied with their default parametrization.

The simplicity and predictive value of AMG led to the development of a decision support tool (SIMEOS-AMG; Duparque et al., 2013) freely available and adapted for use by farmers, agricultural advisors, researchers, as well as for teaching purposes. SIMEOS-AMG uses as input soil and climatic data, information on land-cover and management practices, and provides predictions of the long term effect of current and alternative agricultural practices on SOC stocks. Moreover, the AMG model has been included in the structure of the STICS crop model, used to simulate crop growth, soil water, and C and N balance at a daily time step (Whisler et al., 1986; Brisson et al., 2003, 2010).

1.11. Assessing SOC persistence and estimating conceptual SOC pool sizes

Multi-compartmental models reviewed in the previous section lack accuracy due to a current knowledge gap that is the correct partitioning of the various SOC pools they consist of at the onset of the simulation. Model SOC pools characterized by a given turnover time are conceptually defined and thus not directly measurable. The challenge of "measuring the modellable" (Elliott et al., 1996) fuelled numerous attempts aiming to develop methods to separate functionally homogeneous (non-composite) pools or infer SOC pool sizes and their turnover times resulting in a myriad of methods applied today.

1.11.1. Processes of SOC fractionation

Although the term "fractionation" usually refers to the variety of chemical and physical separation methods, that have been developed to isolate SOC fractions according to their chemical and physical characteristics, the same term can be used to describe the process of biological fractionation occurring in the soil.

1.11.1.1. In-situ approaches

Various kinds of long-term field experiments present different opportunities for studying the evolution of SOC *in situ* under real pedological, climatic and biological conditions. Such methods provide information characteristic of the undisturbed soil environment in real time. These are some of the greatest advantages of long-term field experiments as they allow for the evolution of different SOC fractions over time to be observed and the importance of different drivers to be estimated. A major limitation of these sites is their rare occurrence and their geographical limitation.

1.11.1.1.1. Long-term bare fallow sites

A unique example for *in situ* biogeochemical fractionation of SOC are so called long-term bare fallow experiments (LTBF; Rühlmann, 1999). These are field experiments where soil plots are kept free of vegetation over multiple decades in order to eliminate organic carbon inputs. Through this extreme intervention (of constantly removing all plants), and through regular sampling and monitoring it is possible to observe and measure the decay of the organic carbon initially present at these sites (Fig. 20). Moreover, these experiments provide a unique opportunity to obtain soils enriched in persistent SOC, as the more labile SOC will decay first leading the relative proportion of persistent SOC to increase. Fitting an exponential decay model to the observed SOC evolution trend allows for estimation of parameters such as the decay rate and the persistent SOC content (plateau; Barré et al., 2010). This is the closest we can get to quantifying SOC behaviour under the influence of real temperature and precipitation variations, and in an undisturbed mineralogy and soil environment. This valuable information is only available at rare locations where soil scientists and agronomists have initiated experiments decades ago and continuously managed and maintained them ever since.

One of the strongest criticisms associated with the LTBF trials is the disturbance of the SOC system through the removal of such an important ecosystem aspect, vegetation. In terms of SOC dynamics these experiments are lacking a representation of the priming effect (i.e., the emerging theory stating that fresh SOM inputs enhance mineralization of existing SOC by boosting microbial activity).

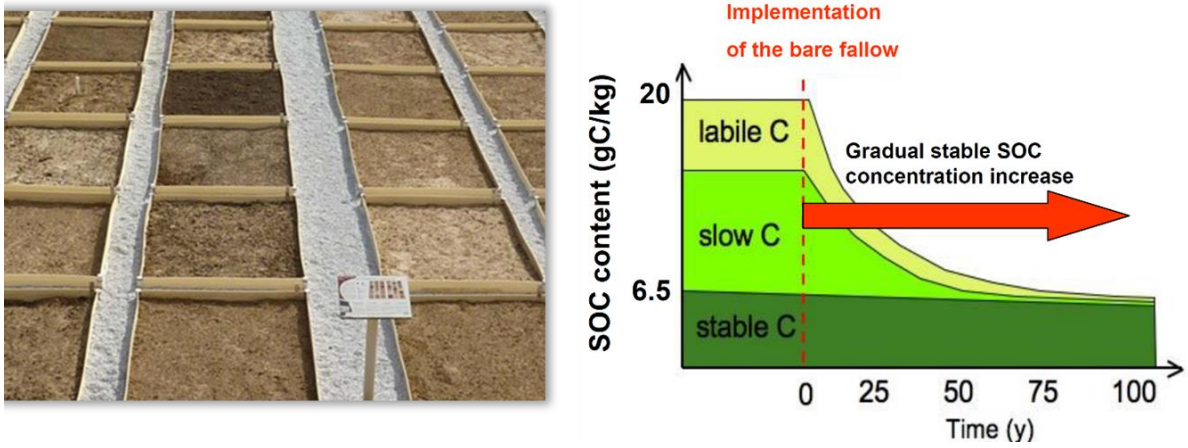


Figure 20: To the left: Picture from the Versailles long-term bare fallow experiment (©INRAE). To the right: Schematic representation of the evolution of the different SOC pools at an LTBF site (Chenu, 2020).

In the next paragraph we discuss a different kind of agronomic experiment and we argue that SOC persistence estimated based on LTBF sites has a chance of being relevant in more realistic conditions as well.

1.11.1.1.2. C_3–C_4 chronosequences and natural ^{13}C abundance

This is a method based on the principle of organic isotopic fractionation. Different photosynthesis mechanisms have evolved in plants impacting the distribution of C isotopes in the organic compounds they produce. The so called "C_3" metabolic pathway is characteristic of plants such as rice, wheat, soya, and potatoes and it results in a $^{(8)}\delta^{13}C$ signature between −24 and −34‰, averaging at −27 ‰, while "C_4" plants such as maize, sugar cane, millet and sorghum have a $\delta^{13}C$ between −6‰ and −19‰ with an average value of −12‰ (Balesdent et al., 1987; Ellam, 2016). The concept behind this method is that since the signature of the plants will be preserved upon organic matter incorporation into the soil or during its decomposition (Fig. 21), ^{13}C natural abundance can be used as a tracer of SOM evolution. In cases with a known time of transition from one vegetation type to another, and where timeseries of soil samples are available, it is possible to track the change in ^{13}C signature with time and estimate rates of fresh SOM incorporation and turnover of native SOM. This method has been used as a tracer for dynamics of total SOC (Rasse et al., 2006; Mary et al., 2020) as well as for following the behaviour of specific SOC fractions (Balesdent, 1987; Skjemstad et al., 1990; Fernández-Ugalde et al., 2016).

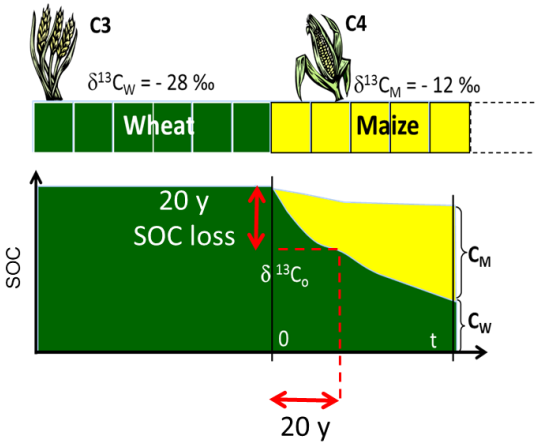


Figure 21: Schematic representation of the evolution of SOC at a C_3–C_4 chronosequence according to the change in $\delta^{13}C$ signature after a vegetation change (Barré, 2020; adapted from Balesdent et al., 1991).

[8]Definition of $\delta^{13}C = \left[\left({}^{13}C/{}_{12}C_{sample} - {}^{13}C/{}_{12}C_{standard}\right) - 1\right] \times 1000$ (Ellam, 2016)

A unique opportunity for comparing SOC decay estimated with the 13C natural abundance method and with the LTBF method is provided at the experimental site of Versailles, which includes both a C3–C4 trial ("Les Closeaux") and LTBF plots. Evidence shows that the observed decay of C3 follows a similar pattern as the decay of bulk SOC (C. Chenu, personal communication; mentioned in Barré, 2020), suggesting that the absence of SOC inputs does not have a significant impact on SOC turnover.

1.11.1.1.3. Long-term monitored agronomic experiments

Long-term monitoring of agricultural practices combined with precise knowledge over the long-term land-use history of a site allows for optimization of specific model parameters. When measurements of all the necessary data (e.g., pedoclimatic, C input, land cover) are available and precise the hypothesis can be made that fitting a model to the observed SOC evolution would result in obtaining the mathematically optimal value of a given parameter, such as mineralization rate or SOC pool partitioning (e.g., Clivot et al., 2019). This method is fast and mathematically elegant but its precision depends entirely on the accuracy of the collected measurements and on the suitability of the chosen model.

1.11.1.2. Laboratory methods: chemical, physical, and physico-chemical fractionation

Most of these procedures (reviewed by von Lützow et al. (2007) and summarized in a pedagogical manner by Don and Poeplau; https://www.somfractionation.org/) were developed with the goal of separating, studying and quantifying functionally homogeneous conceptual pools used in SOC models. The common basic principle behind the various developed approaches is the concept of the three traditionally considered mechanisms responsible for SOC stability mentioned previously (Sect. 1.8.2.1.).

For example, physical fractionation is based on the idea that SOC physical properties are crucial to its persistence and that it is possible to separate SOC fractions along specific physical boundaries e.g., size, density, cohesive aggregate strength. The main assumption behind physical fractionation is that the various interactions between minerals and organic matter are crucial to its stabilization (Six et al., 2002; Eusterhues et al., 2003; Kaiser and Guggenberger, 2003).

Often physical fractionation procedures apply various steps towards physical separation of particles, such as disaggregation, size and density separation, dispersion and sedimentation with the goal of defining the fraction of soil that is associated to minerals in various ways. One example of physical separation is the often used particle size fractionation, aiming to separate the fine clay fraction, or even sub-fractions of clay sized material (Christensen, 1992). Density fractionation is also often used to separate the light SOM fraction which is considered to be accessible and not associated to minerals from a heavier fraction composed of organo-mineral complexes. Another type of physical fractionation is based on the magnetic sensitivity of minerals present in the clay fraction, aiming namely at separating organic matter associated to oxides, known to form strong associations (Torn et al., 1997; Kleber et al., 2004)

Chemical fractionation methods, on the other hand, aim to separate components with different inherent chemical recalcitrance, or to eliminate the soil mineral fraction entirely, in order to isolate and study SOM separately. These include extraction, hydrolysis and oxidation methods.

Fractionation methods combining physical and chemical procedures aim to account for the combination of the effects of particle sizes, chemical composition, soil matrix and organo-mineral associations and differentiate between protected and unprotected SOC.

1.11.1.2.1. Important milestones and historical references

Efforts to separate SOC fractions date back to 1923 as mentioned in Christensen (1992). In more recent years, many studies have concentrated on defining the distribution of SOC amongst stability pools of one of the most commonly used SOC dynamics model, the RothC (Jenkinson and Rayner, 1977). As a reminder, RothC divides soil organic matter into five compartments: one stable inert pool (IOM) and four active pools: decomposable, and resistant plant material, humus, and microbial biomass carbon (DPM, RPM, HUM and BIO), distinguished by different turnover times (see Sect. 1.9.2 for a more detailed model description; Fig. 22).

An attempt to separate measurable SOC fractions correlated to conceptual pools of RothC is presented in Balesdent (1996). Using both physical and chemical procedures, the author assorted SOC into four fractions of decreasing particle sizes and increasing ^{13}C ages. The presented method allowed to satisfactorily relate the >50µm fraction to the RPM pool.

However, the observed link was interpreted as an approximation and not an exact quantification since particles of higher stability are partially present in the larger-size fraction and vice-versa.

Another fractionation procedure was proposed by Skjemstad et al. (2004), based on a combination of pre-existing protocols (Fig. 22). Using soils from two Australian LTE sites, the authors physically separated SOC into two size fractions, larger ([9]POC), and smaller than 53µm. They determined the proportion of the smaller-sized fraction that corresponded to charcoal by photo-oxidation and magnetic resonance spectroscopy ([10]ROC), and defined the humic pool ([11]HOC) by difference (i.e., HOC=TOC-(POC+ROC); Fig. 22). In this study, the three isolated fractions were related to the RPM, HUM and IOM pool of RothC, and they were successfully used to initialize dynamic SOC simulations at the two sites.

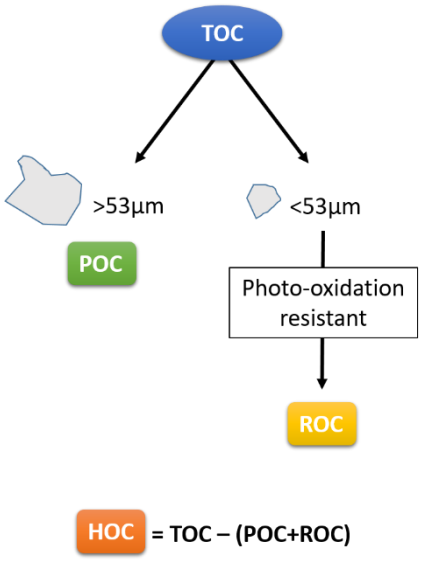


Figure 22: Schematic diagram of the fractionation method proposed by Skjemstad et al. (2004). Rectangles represent the various fractions (POC: particulate organic carbon, ROC: resistant organic carbon and HOC: humic organic carbon) while the arrows represent procedures (i.e., sieving, oxidation and calculation of HOC through subtraction).

The correlation of model and observed data was improved significantly, but only after optimizing the RPM decomposition rate to match pedoclimatic conditions. Moreover, although the use of this fractionation approach seems to be appropriate for Australian soils, its relevance to global or European soils is questionable, considering that the particularly important influence of wildfires in Australian soils might not hold true in different environments where mechanisms of SOC stabilization are more variable.

[9] POC: Particulate organic carbon
[10] ROC: Resistant organic carbon
[11] HOC: Humic organic carbon

In a study conducted on agricultural topsoils from Northern Germany, Weihermüller et al. (2013) were able to show that the >53µm fraction separated according to the protocol proposed by Skjemstad et al. (2004) was positively correlated ($R^2 > 0.7$) to the size of RPM pool of RothC under assumption of SOC equilibrium. The approach developed by Skjemstad et al. (2004) was slightly modified in a study by Herbst et al. (2018) and applied on soil samples from agricultural experimentation sites in Germany and the UK. The authors investigated the correlation of the POC and HOC fractions to the RothC pools RPM and HUM, respectively. Their results showed that under assumption of SOC equilibrium, there was a strong positive correlation ($R^2 > 0.8$) between the HOC fraction and the HUM pool. For the same conditions, a weaker correlation ($R^2 = 0.5$) was observed between the POC fraction and the RPM pool. The authors draw the ambiguous conclusion that although fractionation could be used to improve spin-up model runs, it can also lead to biases.

In an effort to adapt the idea of physico-chemical SOC fractionation to European soils, Zimmermann et al. (2007a) attempted to design an analogous protocol (Fig. 23 & 24). In this study the authors use size-density fractionation as well as chemical oxidation (NaOCl) to divide SOC into five fractions: “DOC” (Dissolved organic carbon; <0.45µm), “rSOC” (resistant SOC; 0.45-63µm and resistant to chemical oxidation), “s+c” (silt and clay; calculated as the fraction 0.45-63µm minus the rSOC), “S+A” (sand and stable aggregates; >63µm and heavier than 1.8 g cm^{-3}) and “POM” (particulate organic matter; >63µm and lighter than 1.8 g cm^{-3}; Fig. 23). The fraction of “rSOC” corresponds directly to the IOM pool of RothC, while the “POM” and “DOC” fractions are summed up and then split into DPM and RPM according to a coefficient calculated based on a SOC equilibrium scenario. Similarly, the SOC remaining in the “s+c” and “S+A” pools is distributed into the BIO and HUM fractions (Fig. 24).

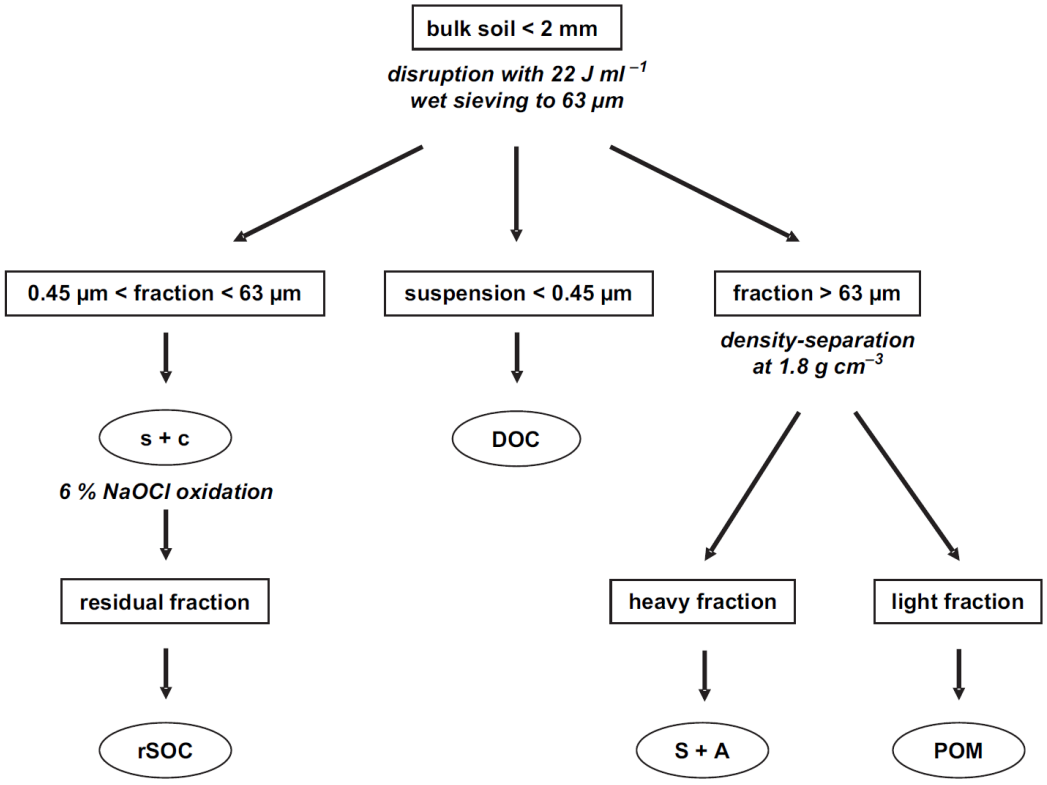


Figure 23: Schematic diagram of fractionation method proposed by Zimmermann et al. (2007). The five separated fractions are shown as ellipses. "s+c" stands for silt and clay, "rSOC" for resistant SOC, "DOC" for dissolved organic carbon, "S+A" for sand and stable aggregates and "POM" for particulate organic carbon (Zimmermann et al., 2007a).

The authors applied this approach on 123 agricultural and managed grassland soils in Switzerland and they showed that the size of the SOC fractions defined with this method was correlated with the size of the model pools in equilibrium.

Although this approach is labour-intense it has been applied in numerous studies in lack of a more appropriate method. In a study using agricultural soils in Ireland, Dondini et al. (2009) repeated this procedure but showed that for the sites used in their study, a good correlation was found only between two out of the three physically separated fractions and RothC pools (HUM and IOM). Leifeld et al. (2009a) attempted to estimate RothC pool sizes for a collection of alpine grassland soils along an altitude gradient. The authors showed that the correlation between the size of model pools and the estimated SOC fractions was weak, and they conclude that the hypothesis of a relationship between fractions obtained with this fractionation protocol and conceptual model pools does not hold true for high altitude grassland soils.

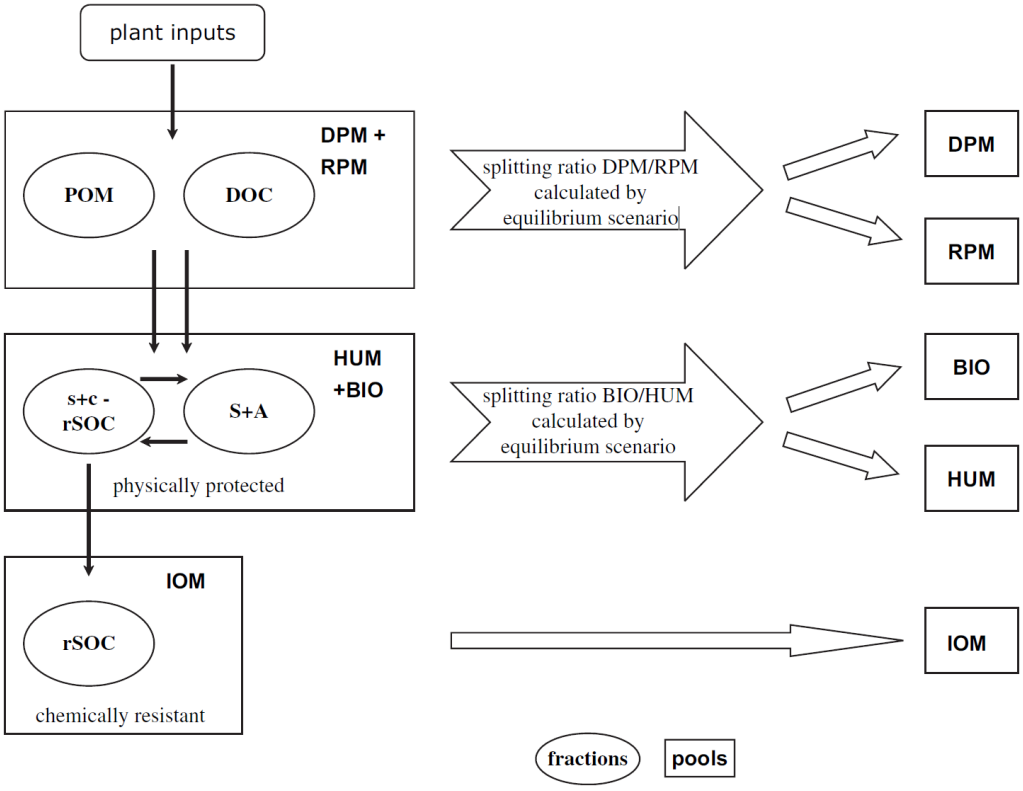


Figure 24: Conceptual diagram showing the procedure for splitting SOC fractions into pools (Zimmermann et al., 2007a).

In an attempt to better understand the limitations of the method proposed by Zimmermann et al. (2007a) caused by its complexity, Poeplau et al. (2013)conducted a comparative study aiming to investigate its reproducibility. The authors followed an identical protocol to analyse the same soils at six independent laboratories. Their work showed significant disagreements for the estimation of all the stability pools (coefficient of variation 14–138%), caused by slight differences in sample treatment. Yet this method continues to be applied, often with slight adaptations, resulting in a multitude of different experimental protocols, which hinders direct comparison of studies following different procedures. Characteristically, Poeplau et al. (2018) conducted a study to evaluate 20 different currently used fractionation methods. Their work provides useful information as to which method might be more appropriate, depending on available samples and the research question, but it emphasizes an inconsistency issue as well.

A more universal and drastically simplified fractionation approach was recently proposed by (Lavallee et al., 2019), separating SOM into just two size-fractions, particulate organic matter (POM) and mineral-associated organic matter (MAOM). These two fractions are considered to differ in regards to their formation, functioning and more importantly their persistence. Moreover, they are easy to separate, as the procedure includes only two steps: soil dispersion to break apart occluded particles, and sieving at 53µm. Although this method has clear advantages such as its simple use compared to complex fractionation methods, discrepancies might still occur due to the variable effect of dispersion on different soil types (Just et al., 2021). Additionally, evidence suggests that SOC within these pools has variable turnover times. Even though the "unprotected" POM fraction is thought to cycle much faster (Gregorich et al., 2006), studies have shown that average age of POM can be similar to bulk SOC (Schmidt et al., 2011) and in some cases it can even exceed 100 years (Trumbore, 2000). It is evident that in all cases, SOC pools contain a mixture of labile and more persistent materials, which makes the question of actually testing and evaluating the performance of this and every fractionation method independently a crucial one.

The lack of complete and independent validation as SOC pool initialization methods is a common drawback to all of the fractionation methods mentioned above. A sufficient validation would require that a fractionation method designed to separate measurable SOC pools is actually used to initialize SOC pool sizes and that the obtained projections of SOC dynamic evolution match actual observations. Currently SOC fractionation methods are evaluated based on their correlation to modelled compartments based on steady-state assumption. Yet, many studies argue that this assumption is unrealistic especially for arable land, even after long periods of known land-use history. In addition, such incomplete evaluation approaches entirely lack evidence regarding the added value brought to dynamic SOC simulations by fractionation-based initializations.

One of the few studies attempting to use a fractionation method to actually initialize SOC simulations and evaluate the improvement brought by this approach, was conducted by Leifeld et al. (2009b) using SOC time series from a long-term agronomical trial in Switzerland. The authors show that the improvement in the accuracy of simulations brought by the Zimmermann et al. (2007a) initialization was non-significant.

Another attempt at applying the same approach is presented in Nemo et al. (2016) who evaluated the fractionation procedure using four agronomical sites in Europe. Their conclusion was similar, as the improvement brought to the precision of model projections was non-significant compared to a spin-up simulation.

The sensitivity of SOC dynamic models to the correct initialization of the sizes of their kinetic pools emphasizes the need for appropriate and validated methods (Luo et al., 2016). Specifically, determining the size of the stable pool with precision is crucial for the accuracy of SOC projections (Clivot et al., 2019; Smith and Falloon, 2000; Taghizadeh-Toosi et al., 2020) The large uncertainties in SOC evolution predictions are an important issue not only in soil science and agronomy but also in ecosystem modelling (Todd-Brown et al., 2013). Poor representation of the SOC module in Earth system models can lead up to an overestimation of global SOC storage potential of up to 40% (He et al., 2016).

Alternative, robust and validated initialization methods are imperative considering the complex non-standardized and time-consuming nature and the questionable performance of current approaches.

1.12. A recent advancement in estimation of SOC persistence: the PARTY$_{SOC}$ model

1.12.1. Thermal instead of physico-chemical fractionation

In the past decade, ramped thermal analysis has shown great potential to obtain proxies for the biogeochemical stability of soil organic carbon (Plante et al., 2013; Gregorich et al., 2015; Soucémarianadin et al., 2018). Especially when accompanied by information on SOC bulk chemistry, such as depletion or enrichment in hydrogen (Barré et al., 2016) or specific moieties (Sanderman and Grandy, 2020), thermal stability showed a strong correlation with classic tools used for estimating the age of carbon atoms (^{14}C; Plante et al., 2013) and *in situ* estimations of biogeochemical SOC stability (LTBF sites; Barré et al., 2016).

1.12.2. Link between thermal and biogeochemical stability *in situ*

A pioneering study linking thermal stability to biogeochemical SOC persistence was conducted by Barré et al. (2016). Compiling a unique data set of archived samples from LTBFs in north-western Europe, the authors analysed sequences of persistent SOC enrichment (with LTBF duration) using a variety of methods. They identified energetic and chemical characteristics of persistent SOC, namely its high thermal stability, low energy content and depletion in hydrogen (Fig. 25). Amongst the four analytical methods applied: Rock-Eval® thermal analysis, TG-DSC, NEXAFS (Near Edge X-Ray Absorption Fine Structure) spectroscopy and Mid-IR spectroscopy, indicators obtained with Rock-Eval® showed the most systematic correlation to bare fallow duration and thus SOC persistence.

1.12.3. The first quantitative thermal-analysis-based machine-learning model predicting SOC persistence

As explained earlier in this manuscript, LTBF experiments offer the opportunity to infer the concentration of persistent SOC at a given site (see Sect. 1.11.1.1.1.; Barré et al., 2010). By applying a Bayesian curve fitting method to the observed SOC decay over the duration of the LTBF (longest trial ~80 years), Cécillon et al. (2018) estimated a site-specific concentration of

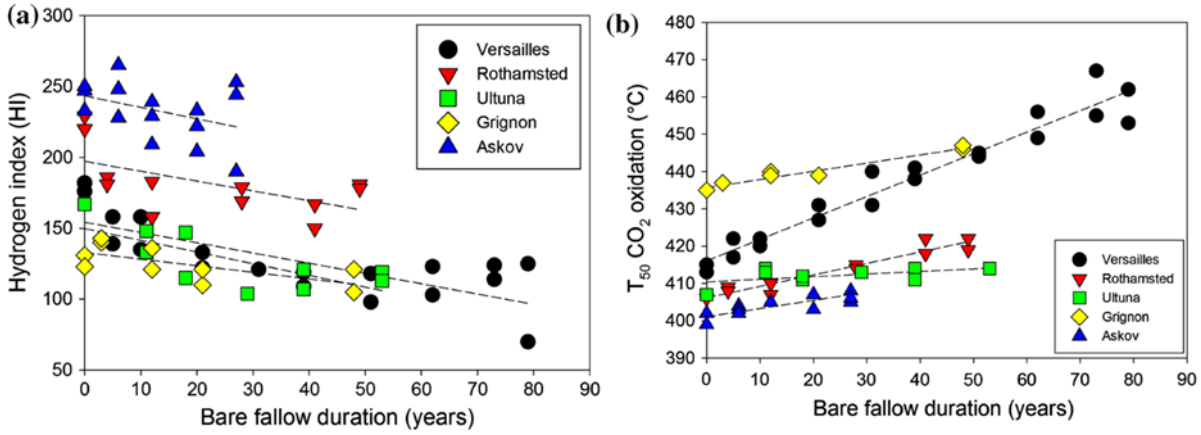


Figure 25: (a) Hydrogen index (in mg HC g TOC^{-1}) and (b) Temperature at which 50 % of CO_2 was released during oxidation as a function of bare fallow duration (Barré et al., 2016).

centennially persistent SOC at four European sites (Versailles, France; Grignon, France; Rothamsted, United Kingdom; and Ultuna, Sweden). The authors assumed that the persistent SOC concentration would remain unchanged over the scale of a century and that it would be uninfluenced by land-management practices. This assumption allowed (i) calculating the proportion of centennially persistent SOC at each sampling time during the experiment at LTBF plots and (ii) at neighbouring plots under different land-uses (manure amendment, straw or composted straw amendment, grassland, and cropland).

Thus a database was generated consisting of time series of SOC evolution at various locations under various land-uses with known changes in persistent SOC proportion with time. Moreover, all archived samples (n=118) were analysed with Rock-Eval® and 30 parameters per sample were obtained, characterizing their organic matter thermal stability and bulk chemistry.

Finally, based on the correlations between Rock-Eval® parameters and biogeochemical stability, Cécillon et al. (2018) trained a non-parametric machine-learning algorithm using a random-forest approach to accept Rock-Eval® parameters as predictor variables and generate a value of centennially persistent SOC proportion as a target variable.

This model was internally validated in the same study (“out-of-bag” validation; R^2=0.91, RMSEP=0.07) showing that the algorithm could efficiently predict the centennially persistent SOC proportion of samples from the same sites but which were not included in the learning set of the model.

This version was recently adapted (Cécillon et al., 2021; ANNEX 1) to form the most up-to-date version of the model currently available and used in this study.

1.12.4. Rock-Eval® thermal analysis machine-learning models PARTY$_{SOC}$v2.0 and PARTY$_{SOC}$v2.0$_{EU}$

The complete work by Cécillon et al. (2021) presenting two new versions of the model and explaining in detail their differences compared to the older version (see ANNEX 1 Supplementary Table S1) can be found in ANNEX 1 of this manuscript.

Here we briefly discuss some of the most important updates. First, the network of sites used for model calibration was expanded in the geographical sense, including samples from Germany (Bad Lauchstädt) and Colombia (La Cabaña). Second, a new type of long term trial was included in the model's training data set, a C_3–C_4 vegetation change chronosequence (La Cabaña, Colombia). Third, the calculation of persistent SOC concentration was adapted to account for errors or uncertainties in the Bayesian fitting method as follows: in LTBFs where modelled estimations of persistent SOC concentration exceeded the lowest SOC concentration observed *in situ*, the latter value was used instead. Fourth, the contribution of each reference site to the training set in terms of number of samples was equalized (n=15 per site). Fifth, the number of Rock-Eval® parameters used as predictors was reduced to 18, out of which some were directly related to SOC content, and which were all correlated to the proportion of centennially stable SOC (Spearman's rho > 0.5).

The conservative estimation of the prediction error of the European version of the model (PARTY$_{SOC}$v2.0$_{EU}$) was obtained with a leave-one-site-out validation and was in the order of RMSE=0.15 (RRMSE=0.27) for the six independent validation sites. These results showed promise for the use of the model on independent sites from similar pedoclimates.

In this study we used the most up-to-date European (PARTY$_{SOC}$v2.0$_{EU}$) version of the PARTY$_{SOC}$ model. In this manuscript the name PARTY$_{SOC}$ is often used without any indication to the specific version. Since only the European (PARTY$_{SOC}$v2.0$_{EU}$) version of the model was used throughout this work, the name PARTY$_{SOC}$ refers to the European version, unless specified otherwise.

1.13. The principles and methodology evolution of the Rock-Eval® technique

The apparatus was originally invented to assess the type, maturity degree and petroleum potential of source rocks (Espitalié et al., 1977). Rock-Eval® thermal analysis served for almost five decades as a standard technique in the petroleum industry(Espitalie et al., 1985b, a, 1986; Lafargue et al., 1998; Behar et al., 2001). Since the 1990's after several upgrades on its hardware, the technique saw new alternative applications, such as characterization of recent sediments, soil organic matter analysis and even analysis of pure biochemical compounds (Lafargue et al., 1998; Di-Giovanni et al., 2000; Disnar et al., 2003; Hetényi et al., 2005; Carrie et al., 2012a; Baudin et al., 2015; Gregorich et al., 2015; Barré et al., 2016; Sebag et al., 2016; Soucémarianadin et al., 2018).

1.13.1. The Rock-Eval® 6 device

The Rock-Eval® 6 TURBO device (Vinci Technologies, France) used in this study, is equipped with two micro ovens, as well as a fully automated sampler, a flame ionisation detector (FID) and two infra-red (IR) cells. The apparatus set-up allows direct monitoring of HC, CO and CO_2 effluents during programmed heating of the sample (Fig. 26). More precisely, two successive steps can be applied, pyrolysis and oxidation, generating three and two thermograms respectively. For the analysis, 20–100 mg of sample (depending on its carbon content) are placed in a crucible made of an incolloy, designed for high corrosion resistance as well as strength at high temperatures. A fine grid forming the bottom and the top of the crucible allows the circulation of carrier gas. The crucible enters the micro oven automatically, where its top and bottom are in direct contact with the two ends of the thermocouple wires. Due to this setup the true temperature occurring in the micro ovens can be measured in real time with a precision of 0.5°C.

In the case of sequential pyrolysis and oxidation, the sample first undergoes pyrolysis under an inert (N_2) atmosphere. The released gas passes through a heated splitter (400 °C) where it is divided into two equal parts. One half is led to the FID, where the released hydrocarbons are quantified. Hydrocarbon effluents can be measured accurately with an FID sensitivity of 100 pA to 1 µA. Simultaneously, the second half of the pyrolysis effluents is guided to the IR cells to register the CO and CO_2 generation.

During the second analysis step, the residual carbon is oxidized under laboratory atmosphere, while CO and CO_2 effluents are directly monitored by the IR detectors. The sensitivity of the IR cells is in the order of 12 ppm for CO and 25 ppm for CO_2 (Vinci Technologies).

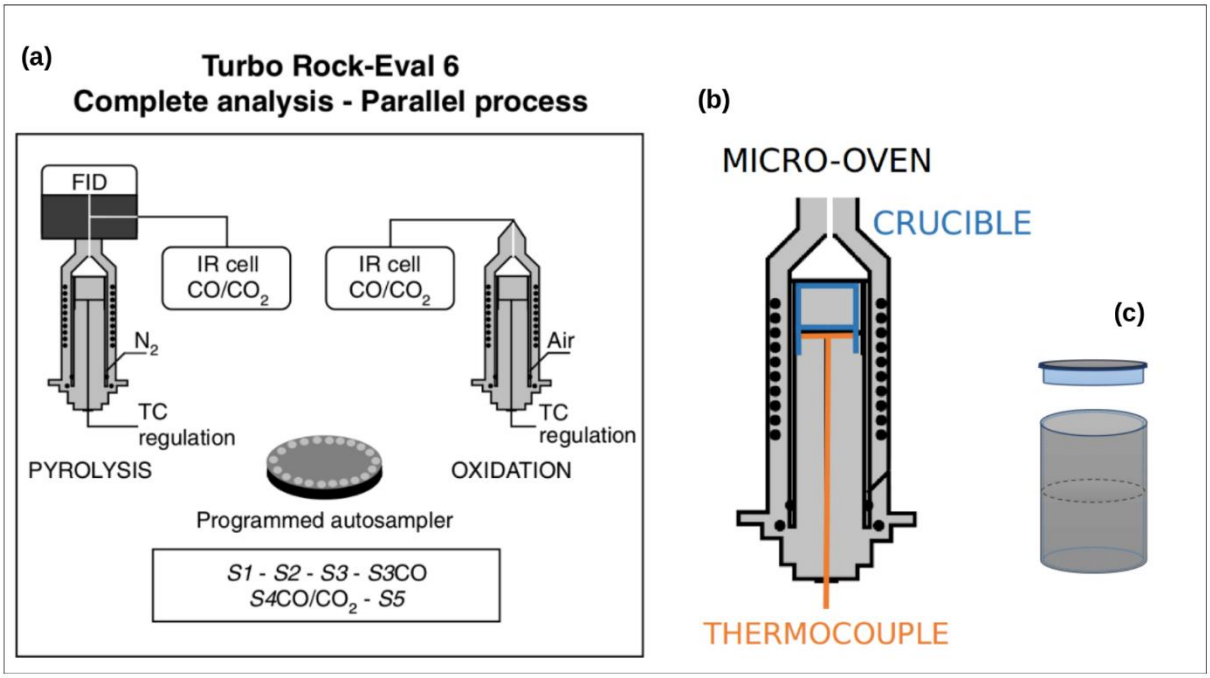


Figure 26: (a) The set-up of the Rock-Eval®6 apparatus including the two micro ovens, the used detectors and monitored gases. In the lower middle box are some of the acquisition parameters (the names of the obtained thermogram peak areas; Behar et al., 2001). (b) Closer look at a micro-oven and the positioning of the thermocouple in direct contact with the crucible (changed after Behar et al., 2001). (c) Schematic drawing of a crucible.

1.13.1.1. Detection of gases during Rock-Eval® analysis

1.13.1.1.1. Flame Ionisation Detection

FID is an instrument that is often used to measure the concentration of organic species in a gas flow. The sample is burned in a hydrogen flame, which leads to the release of ions and electrons. As ions are generated inside the detector, they are transported between two electrodes where a difference of potential of a few hundred volts is occurring. The presence of the produced ions can generate an electric current in the order of Pico amps (10^{-12} A). This current can be recorded, converted to a voltage and then filtered and amplified as specified. The intensity of the resulting current is directly proportional to the amount of ions present in the detector (The Flame Ionization Detector, 2005).

1.13.1.1.2. Infrared Spectroscopy

This method is based on absorption spectroscopy. Through the interaction of radiation with matter, chemicals can be identified and studied. A beam of a specific spectrum (IR) is passed through the sample. Characteristically, matter absorbs the radiation that matches its vibrational frequency, also known as resonance frequency. By examining the transmitted light, the amount of energy and the specific frequency of absorption is revealed. Vibrational and thus absorbing frequency is dependent on the structure of a molecule. It can be influenced by the masses and the positions of the atoms forming a molecule as well as the associated vibronic coupling (interaction between electronic and nuclear vibrational motion). As a result, the infrared spectrum of a substance is characteristic of its chemical composition (Derrick et al., 1999).

1.13.1.2. Output data

The output data is displayed in five thermograms. Each thermogram presents time information on the x-axis (that can be converted to temperature if the heating routine is known) measured in seconds, and the intensity of the signal of a given detector (showing the evolution of a gas) on the y-axis (in mV for the FID or ppm for the IR signal). Peak areas are defined for each thermogram to interpret the origin of carbon (Behar et al., 2001). The FID signal recorded during pyrolysis (HC_PYR thermogram) consists of two peaks: S1, showing the release of free hydrocarbons, and S2, representing pyrolyzed hydrocarbons with respect to the oil generation. The IR signal monitoring CO_2 released during pyrolysis (CO2_PYR thermogram) can be divided into two peaks: S3, formed through CO_2 organic origin, and S3', formed by CO_2 of inorganic origin. The CO signal recorded by IR detection during pyrolysis (CO_PYR thermogram) also has two peak areas, S3CO and S3'CO, similarly representing effluents of organic and inorganic origin. The IR signal of CO_2 detected during oxidation (CO2_OX thermogram) includes two peak areas as well, S4CO2 and S5, representing an organic and inorganic source respectively. A single peak area is considered for the CO gas detected by IR during oxidation (CO_OX thermogram), called S4CO, and is attributed to organic carbon.

The peak area limits are pre-defined for the HC_PYR, and CO2_PYR thermograms and they can be located automatically or manually for the CO_PYR, CO2_OX and CO_OX thermograms.

The cut-off in each case is defined as the minimum value of gas production observed in a specific range, namely before the occurrence of the signal corresponding to inorganic carbon. A detailed overview of the limits and the interpretation of each area is illustrated in Fig. 27. The most important inorganic carbon signal due to the presence of carbonates has a characteristic peak recorded in the CO2_OX thermogram after 611°C (s. Fig. 27).

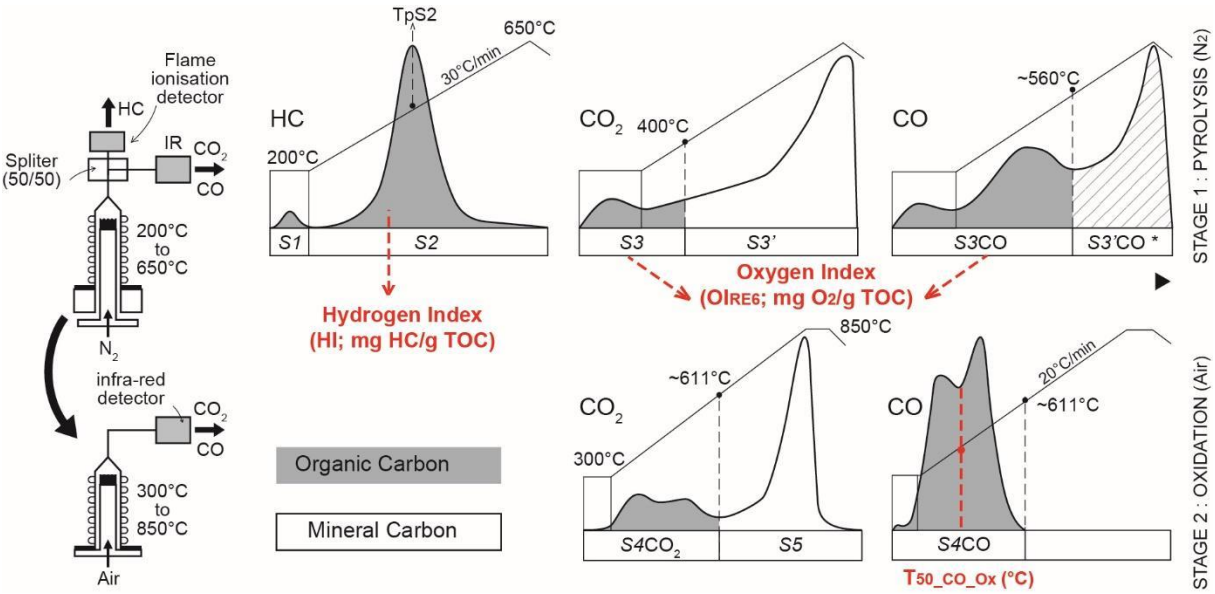


Figure 27: Representation of five thermograms obtained by Rock-Eval® analysis. Grey areas correspond to detection of carbon of organic origin while white areas correspond to inorganic carbon (©Baudin, after Behar et al., 2001).

1.13.1.3. Acquisition parameters

Based on the generated thermograms, several parameters can be calculated based on integrated information on the carbon content, the size of different carbon fractions and the bulk chemistry and thermal stability of the analysed organic matter. In order to keep this section short and avoid repeating technical information too many times in the manuscript, we direct the reader to Chapter 2, Supplementary Table 1 for more details. The most important and relevant information is repeated in the methods section of each chapter.

2. The work conducted during this thesis in three chapters

Structure of writing

After this general introduction giving the context of this work and reviewing the current state of the art relevant to this topic follows a presentation of the work conducted in the frame of this thesis in the last three years. The work is presented in three chapters, each addressing a different research question.

Organization of topics

Each chapter is introduced by an overview reminding the reader of the objectives, hypotheses, highlights, conclusions and perspectives of each part. The body of each chapter consists of a journal paper produced during the thesis (chapter 1 and 2) or one in preparation (chapter 3). Each chapter can stand independently as it contains a summary of the work, its context, a review of relevant literature, a description of the methods and materials used, a results section, a short discussion section and a section with conclusions and perspectives. Still, the chapters are presented in an order chosen to create a story showing how these questions are relevant to each other. They are bound together by a general extended discussion at the end of the manuscript whose goal is to answer points mentioned earlier on and pave the way for the general conclusions and perspectives of this work.

Research questions

My thesis work can be summarized in three objectives, which I tried to investigate and describe in the three following chapters.

The research question of the first chapter is if we can apply the $PARTY_{SOC}$ model to sites outside of its calibration set and if so, evaluate the potential of this approach for improving the accuracy of SOC simulations.

The next two chapters focus on learning more about the limits and possibilities of the approach using laboratory experiments. Namely, in the second chapter we ask if it is possible to compare Rock-Eval® parameters of samples originating from different depths for purposes of harmonisation of existing databases.

The question of the third and final chapter is if it is possible to progress towards a more mechanistic understanding of the link between thermal and biogeochemical stability. Specifically, can we study the role of adsorption on mineral surfaces − a classic SOC protection mechanism – using Rock-Eval® and if so what is its influence on thermal stability?

CHAPTER 1

Using estimations of SOC persistence predicted with machine-learning and Rock-Eval® thermal analysis to improve the accuracy of simulations of SOC dynamics.

In this chapter we address the question of the poor accuracy of SOC projections by models, caused by the lack of precise information on initial SOC kinetic pool partitioning.

Our objectives were: first, to test the latest version of the thermal analysis-based machine-learning model $PARTY_{SOC}$ (providing estimations of persistent SOC) on independent samples and second, to use the obtained information to initialize the pool partitioning of AMG (Andiulo et al., 1999; currently the most accurate model for simulating SOC stock evolution in French cropland) to estimate the potential improvement on SOC simulations.

We hypothesized that the $PARTY_{SOC}$ model can be applied efficiently on new soil samples from similar pedoclimates as its calibration set (North-western Europe). Second, since the chosen model of SOC dynamics (AMG) and the $PARTY_{SOC}$ model have the same SOC pool architecture, we expected the pool partitioning obtained with these two approaches to be directly comparable. Third, we hypothesised that the use of the pool partitioning predicted by $PARTY_{SOC}$ would improve the accuracy of SOC stock projections, compared to the models default initialization.

This work was conducted on archive and recent soil samples from 9 French long-term agronomic experiments, including 32 treatments. We showed that the $PARTY_{SOC}$ model predicts a SOC pool partitioning that accounts for legacy effects of land cover and soil management practices and is optimal for AMG. Coupling the two models, i.e., initializing the pool partitioning of AMG using $PARTY_{SOC}$ allowed us to reproduce the observed SOC changes on a pluri-decadal scale at the 32 treatments. The simulations were especially improved in sites with complex land cover and soil management histories, where SOC stocks were far from equilibrium. As this is the case for most European croplands, we estimate the net potential for improvement brought by this method to be substantial.

Additionally, this initialization method is quick, has a low cost and is easy to implement. As such it presents a great potential for obtaining accurate predictions of SOC stocks evolution over the coming decades at large scale.

Another important message we underline in this work is the value of simple multi-compartmental models of SOC dynamics as a tool for accurately predicting SOC stock changes. We demonstrate that when properly initialized a simple model provides unbiased and precise simulations (on the pluri-decadal scale). Contrary to a new generation of mechanistic models that are mathematically more complex and require additional parametrization work and validation, simple models on the other hand, once their kinetic pools are correctly initialized, can provide reliable SOC projections at the plot, ecosystem and global scale.

Nevertheless, before this initialization approach can be applied globally, multiple steps are necessary. First, more sites that allow for precise estimation of *in situ* SOC persistence (such as carefully monitored long-term agronomical experiments, long-term bare fallow sites, and C_3-C_4 chronosequences) will have to be included in the calibration set of an updated version of the $PARTY_{SOC}$ model to cover a larger range of pedoclimates. The updated version of the model will have to be validated in relation to similar long-term experiments from the new cover areas. Moreover, for the final part of coupling $PARTY_{SOC}$ to a global SOC dynamics modelling scheme, several questions remain. Even though it is a simple model, and therefore easily transferable, the use of AMG at new sites will require a lot of data (especially for model evaluation purposes): climatic records, history of land cover and precise land management practices. Alternatively, the structure of $PARTY_{SOC}$ could be revised, to match the pool structure of more widely used models of SOC dynamics such as RothC (Coleman and Jenkinson, 1996) or CENTURY (Parton et al. 1987), in which case new indicators able to split SOC into a larger number of pools or pools with different mean turnover times will have to be determined. These points are discussed in more detail in the general discussion of this manuscript.

The work presented in this chapter was conducted in close collaboration with Bruno Mary, Fabien Ferchaud, Hugues Clivot and Fabien Levavasseur. The result of this work was published in the EGU journal Biogeosciences. The paper under the title “A robust initialization method for accurate soil organic carbon simulations” can be found on the following pages.

A robust initialization method for accurate soil organic carbon simulations

Eva Kanari[1,2], Lauric Cécillon[1,3], François Baudin[2], Hugues Clivot[4], Fabien Ferchaud[5], Sabine Houot[6], Florent Levavasseur[6], Bruno Mary[5], Laure Soucémarianadin[7], Claire Chenu[6], Pierre Barré[1]

[1]Laboratoire de Géologie, École normale supérieure, CNRS, Université PSL, IPSL, Paris, France
[2]ISTeP, UMR 7193 Sorbonne Université, CNRS, Paris, France
[3]Normandie Univ, UNIROUEN, INRAE, ECODIV, Rouen, France
[4]Université de Reims Champagne Ardenne, INRAE, FARE, UMR A 614, 51097 Reims, France
[5]BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université Picardie Jules Verne, F-02000, Barenton-Bugny, France
[6]UMR ECOSYS, INRAE, AgroParisTech, Université Paris-Saclay, Thiverval-Grignon, France
[7]ACTA - les instituts techniques agricoles, Paris, France

Abstract

Changes in soil organic carbon (SOC) stocks are a major source of uncertainty for the evolution of atmospheric CO_2 concentration during the 21st century. They are usually simulated by models dividing SOC into conceptual pools with contrasted turnover times. The lack of reliable methods to initialize these models, by correctly distributing soil carbon amongst their kinetic pools, strongly limits the accuracy of their simulations. Here, we demonstrate that PARTY$_{SOC}$, a machine-learning model based on Rock-Eval® thermal analysis optimally partitions the active and stable SOC pools of AMG, a simple and well validated SOC dynamics model, accounting for effects of soil management history. Furthermore, we found that initializing the SOC pool sizes of AMG using machine-learning strongly improves its accuracy when reproducing the observed SOC dynamics in nine independent French long-term agricultural experiments. Our results indicate that multi-compartmental models of SOC dynamics combined with a robust initialization can simulate observed SOC stock changes with excellent precision. We recommend exploring their potential before a new generation of models of greater complexity becomes operational. The approach proposed here can be easily implemented on soil monitoring networks, paving the way towards precise predictions of SOC stock changes over the next decades.

1. Introduction

Soil organic carbon (SOC) plays an important role in sustaining soil functions and associated soil ecosystem services worldwide (IPCC, 2019). It is the largest terrestrial organic carbon reservoir, with the upper two meters of soil storing 2400 Pg C, three times more carbon than the atmosphere (Jobbagy and Jackson, 2000). A mere 4 per 1000 annual decrease in SOC stocks (*ca.* 10 Pg C yr^{-1}) may double the global annual anthropogenic CO_2 emissions, while an equivalent increase may compensate them (Balesdent and Arrouays, 1999). This is the concept behind the 4 per 1000 initiative (Rumpel et al., 2018) that aims at increasing SOC stocks to fight global warming while ensuring food security, two Sustainable Development Goals of the United Nations (UN General Assembly, 2015). This initiative and other political headway have placed the question of managing SOC stocks and assessing the global SOC sequestration potential at the top of political and scientific agendas (Vermeulen et al., 2019; FAO, 2019; Amelung et al., 2020). Despite this particular attention, the prediction of SOC stock changes remains very uncertain, which makes soils a major source of uncertainty for the evolution of atmospheric CO_2 concentration (Todd-Brown et al., 2014; He et al., 2016; Shi et al., 2018).

Models of SOC dynamics can predict future SOC stock evolution by simulating carbon transfer into the soil mostly through plant organic matter inputs, and microbial SOC mineralization resulting in a CO_2 flux from the soil to the atmosphere. They can have structures of various complexities reflecting our mechanistic understanding of SOC dynamics (Luo et al., 2016; Shi et al., 2018). However, most models dedicated to prediction, including those used in Earth System Models, have a simple structure dividing SOC into conceptual pools with contrasted turnover times (Manzoni and Porporato, 2009; Todd-Brown et al., 2014; He et al., 2016). These multi-compartmental models of SOC dynamics are the best option we currently have to foster science-based SOC preservation and sequestration actions, given the strong uncertainty of more complex models (Shi et al., 2018; Crowther et al., 2019; Cécillon, 2021a; Dangal et al., 2021; Lee et al., 2021). Predictions of SOC stocks evolution provided by such simple models are very sensitive to the initial distribution of SOC amongst the different kinetic pools (Smith and Falloon, 2000; Luo et al., 2016; Clivot et al., 2019). This makes the question of SOC kinetic pool partitioning a priority for improving the accuracy of multi-compartmental SOC dynamics models (Luo et al., 2016; Taghizadeh-Toosi et al., 2020).

The most commonly used method to initialize the size of SOC kinetic pools is to run spin-up simulations until a steady-state equilibrium for SOC is reached, eventually matching the initial SOC stock measurement (Wutzler and Reichstein, 2007; Taghizadeh-Toosi et al., 2020). However, this method has two well-known limitations. First, climatic, SOC input, and land-use or land-cover data extending over long time periods required by this approach are highly uncertain. Second, assuming steady-state equilibrium for SOC at the onset of model simulations is often unrealistic. This is due to the history of the simulated sites that often includes disturbances (*e.g.* fire), as well as previous changes in climate, land-use and soil management that prevent SOC pools with slow turnover times from being at equilibrium (Wutzler and Reichstein, 2007; Poeplau et al., 2011; Oberholzer et al., 2014; Herbst et al., 2018; Clivot et al., 2019). Alternative initialization procedures are needed to address these issues (Bruun and Jensen, 2002; Wutzler and Reichstein, 2007; Taghizadeh-Toosi et al., 2020).

In some models of SOC dynamics, like the AMG model (Clivot et al., 2019), a default initial SOC pool size distribution is prescribed according to basic information on land-use history (*i.e.* long-term cropland *vs.* long-term grassland; Clivot et al., 2019). This approach does not take into account the effect of recent changes in land-use or historical soil management practices on SOC pool distribution. To better reflect the effect of the frequent state of non-equilibrium of SOC on its partitioning into conceptual kinetic pools, another approach has been proposed, relating results from SOC fractionation methods with SOC kinetic poolsizes (*e.g.* Zimmermann et al. (2007a) or Skjemstad et al. (2004) for the RothC model; Dangal et al. (2021) for the DAYCENT model). However, this approach also suffers from important drawbacks. First, SOC fractionation procedures are tedious and cannot be implemented on large-scale studies, though this problem may be solved by using soil infrared spectroscopy or environmental variables and machine- learning (Zimmermann et al., 2007b; Barthès et al., 2008; Baldock et al., 2013; Cotrufo et al., 2019; Viscarra Rossel et al., 2019; Dangal et al., 2021; Lee et al., 2021; Lugato et al., 2021; Sanderman et al., 2021). Second, their reproducibility is questionable (Poeplau et al., 2013, 2018), and third, their use for initializing model SOC pool sizes has never been properly validated. A proper validation would require showing that (1) the size of measured SOC fractions matches the one of model kinetic pools, and that (2) simulations of SOC dynamics are more accurate using this initialization strategy, compared to default simulations (on independent validation sites while other model parameters remain constant).

Reasonably good correspondence between measured or soil-spectroscopy-estimated SOC fractions and modelled SOC conceptual pools has been reported in a number of studies, though with some notable discrepancies (Zimmermann et al., 2007a; Leifeld et al., 2009a; Herbst et al., 2018; Dangal et al., 2021). Conversely, the studies that attempted to initialize model SOC pool sizes using a SOC fractionation scheme generally reported no improvement in the accuracy of simulations of SOC dynamics compared to a default or a spin-up initialization approach (Leifeld et al., 2009b; Nemo et al., 2016; Cagnarini et al., 2019). Only two studies showed that an initialization based on a SOC fractionation scheme yielded more accurate simulations of observed SOC dynamics, but at the cost of modifying the decomposition rate of SOC kinetic pools (Skjemstad et al., 2004; Luo et al., 2014).

An alternative approach using Rock-Eval® thermal analysis has recently been proposed — under the name PARTY$_{SOC}$ model — to estimate SOC kinetic pool sizes (Cécillon et al., 2018, 2021). PARTY$_{SOC}$ is a machine-learning model trained on Rock-Eval® data of soil samples from long-term experiments (LTEs) where the size of the centennially stable SOC fraction can be estimated (*e.g.* sites including a bare fallow treatment). PARTY$_{SOC}$ incorporates recent key elements of the new understanding of SOC dynamics (Dignac et al., 2017), showing that the centennially stable SOC fraction has specific chemical and energetical characteristics that are measurable quickly (*ca.* 1 h per sample) and at a reasonable cost (less than USD 60) using Rock-Eval®; it is thermally stable (*i.e.* high activation energy) and it is depleted in hydrogen (Gregorich et al., 2015; Barré et al., 2016; Hemingway et al., 2019; Poeplau et al., 2019; Chassé et al., 2021).

In this study, we tested if the PARTY$_{SOC}$ machine-learning model, built on a totally independent data set from North-western Europe, could be used to initialize the distribution of SOC pools of the simple AMG model (Clivot et al., 2019) and improve the accuracy of its simulations. The default version of AMG is currently the most accurate model for reproducing the observed SOC stock dynamics in diverse French agricultural LTEs at the pluri-decadal scale (Martin et al., 2019). The efficient use of this model at sites covering an important pedological and climatic variability (including oceanic, continental, and tropical climate) provides further support to its robustness (Levavasseur et al., 2020; Farina et al., 2021; Saffih-Hdadi and Mary, 2008).

In this model, SOC is simply divided into two pools, the "stable SOC (C_S)" that is considered as inert at the time scale of the simulation and the "active SOC (C_A)" that has a mean turnover time of a few decades. A recent study (Clivot et al., 2019) determined that the optimal initial proportion of stable SOC (C_S/C_0) can deviate from the model's default value (0.65 in croplands), so that a more precise initialization of the C_S/C_0 proportion would significantly improve AMG simulations of SOC dynamics. Here, we hypothesized that the SOC pool partitioning as determined by the $PARTY_{SOC}$ machine-learning model (Cécillon et al., 2021, 2018) would be close to the mathematically optimal one for the AMG model, therefore, improving the accuracy of its SOC dynamics simulations compared to default initialization. We tested our hypothesis on 32 treatments from nine independent French agricultural LTEs (experiment duration from 12 to 41 years with a median of 21 years) in which the AMG optimal SOC pool partitioning could be determined by *ex-post* optimization and for which topsoil samples collected at the onset of the experiment were available (Table 1). These LTEs were croplands established in different pedoclimates that have experienced contrasted soil management practices and land-use histories. All available initial topsoil samples were analysed with Rock-Eval® and the results were used in the European version of the $PARTY_{SOC}$ model, $PARTY_{SOC}v2.0_{EU}$ (Cécillon et al., 2021), to compute the centennially stable SOC proportion of each topsoil sample.

2. Materials and methods

2.1. Experimental sites

This work was conducted on nine French agricultural LTEs (Supplementary Material Fig. 1). Seven LTEs including 29 treatments were selected from the dataset presented in Clivot et al. (2019), from sites with availability of initial topsoil samples. Two additional LTEs (Colmar and Feucherolles) including a total of three treatments were obtained from the dataset published in Levavasseur et al. (2020), selecting control treatments without organic amendments and with available initial topsoil samples. Basic site and topsoil characteristics are reported in Table 1 and Supplementary Material Table 1. Information necessary to run AMG simulations on a total of 32 treatments (initial soil physico-chemical properties, detailed information on management practices and observed climatic data during all experiments) were obtained from Clivot et al. (2019) for the 29 treatments of the seven sites and from Levavasseur et al. (2020) for the three treatments of the sites of Colmar and Feucherolles.

Table 1 : Main information on the nine French agricultural long-term experiments used in this study. All sites had been croplands for at least 20 years before the onset of all experiments. Additional site information including climatic variability amongst sites and long-term history of land cover is provided in Supplementary Material Table 1.

	Auzeville	**Boigneville**	**Colmar**	**Doazit**	**Feucherolles**	**Grignon-Folleville**	**Kerbernez**	**Mant**	**Tartas**
Soil type (WRB 2014)	Luvisol	Haplic Luvisol	Calcaric Cambisol	Luvisol	Gleyic Luvisol	Luvisol	Cambisol	Dystric Luvisol	Luvic Arenosol
LTE onset	1968	1970	2000	1967	1998	1958	1978	1975	1972
Simulated period	1975–2010	1970–2011	2000–2018	1977–1989	1998–2019	1989–2008	1978–2005	1975–1992	1976–1997
Number of treatments	4	12	1	2	2	2	5	2	2
Sampling dates (number of samples)	1975 (4), 2010 (8)	1970 (29), 1998 (10), 2017 (32)	2000 (4), 2018 (6)	1977 (4), 1989 (4)	1998 (8), 2013 (8), 2018 (8)	1989 (8), 2008 (8)	1978 (6), 1991 (6), 2005 (12)	1975 (4), 1992 (4)	1976 (4), 1997 (4)
Crop rotation	Annual crop rotation	Annual crop rotation	Annual crop rotation	Maize mono culture	Annual crop rotation	Annual crop rotation	Silage maize monoculture (KERB_C incl. raygrass)	Maize mono culture	Maize mono culture
Considered depth (cm)	30	29	28	25	28	30	25	28	28
Initial SOC stock (tC ha^{-1})	34.68	42.40	45.20	26.35	43.80	55.85	81.98	38.75	45.25
Reference	(Colomb et al., 2007)	(Dimassi et al., 2014)	(Obriot, 2016)	(Lubet et al., 1993)	(Noirot-Cosson et al., 2016)	(Barré et al., 2008)	(Vertès et al., 2007)	(Messiga et al., 2010)	(Morel et al., 2014)

2.2. Archive soil samples from experimental sites

Our final soil sample set included 181 topsoil samples. At each site the soil was sampled to include the whole ploughing depth (Table 1). At all sites, except Boigneville where the soil was sampled in five sublayers, the ploughing layer was sampled as one homogeneous layer. Of the final samples, 71 were from starting dates of the nine LTEs, 24 from LTEs intermediate dates and 86 from LTEs final dates. All samples were air-dried or dried at 40 °C, sieved to < 2 mm and finely ground to < 250 µm using a ball mill (Retsch, Germany).

2.3. Rock-Eval® analysis of archive soil samples

All soil samples were analysed using a Rock-Eval 6® Turbo apparatus (Vinci Technologies). The samples were first pyrolyzed in an inert N_2 atmosphere, then oxidized under ambient air (O_2). The heating routine applied during pyrolysis was as described in Disnar et al. (2003), starting with a three-minute isotherm at 200 °C followed by a heating ramp of 30 °C min^{-1} up to 650 °C. For the oxidation step, a one-minute isotherm was kept at 300 °C and was directly followed by a heating ramp of 20 °C min^{-1} until 850 °C was reached, followed by a five-minute isotherm at 850 °C (Baudin et al. 2015; adapted from Behar et al., 2001).

Based on five generated Rock-Eval® thermograms, 18 parameters were calculated for each sample, and were then used as predictors by the random forests model. These include Total Organic Carbon (TOC; in g C kg $soil^{-1}$) — the amount of organic C released during the analysis as a proportion of sample weight; Pyrolyzed organic Carbon (PC; in g C kg $soil^{-1}$) — the sum of C released as HC, CO and CO_2 during pyrolysis step; the ratio of PC to TOC (PC/TOC); S2 peak area (g C kg $soil^{-1}$) — the hydrocarbon gas released within the range of the pyrolysis temperature ramp; the ratio of S2 to PC (S2/PC); PseudoS1 peak area (g C kg $soil^{-1}$) — the sum of C released as HC, CO and CO_2 during the first 200 seconds of pyrolysis (after Khedim et al., (2021); Hydrogen Index (HI; in mg HC g TOC^{-1}) — the amount of hydrocarbons released as a ratio of TOC; the ratio of HI to Oxygen Index (HI/OI_{RE6}) — where OI_{RE6} is calculated as the amount of oxygen released as CO and CO_2 gases normalized to TOC. Finally, various temperature parameters ($T70_{HC\_PYR}$, $T90_{HC\_PYR}$, $T30_{CO2\_PYR}$, $T50_{CO2\_PYR}$, $T70_{CO2\_PYR}$, $T90_{CO2\_PYR}$, $T70_{CO\_OX}$, $T50_{CO2\_OX}$, $T70_{CO2\_OX}$, $T90_{CO2\_OX}$; in °C) are included in the predictors set. They describe evolution steps, namely at which temperature a specific amount (e.g. 30, 50, 70 or 90%) of a given gas was released according to each thermogram (Cécillon et al., 2018).

It is important to note that no pre-treatment of $CaCO_3$-containing samples was necessary before Rock-Eval® analysis. The slow pyrolysis and oxidation steps of the Rock-Eval® method allow distinguishing carbon of organic and mineral form, since the latter is released above a given temperature. For the calculation of all of the above parameters, only the part of each thermogram corresponding to organic carbon was taken into account. For this purpose, upper temperature integration limits for Rock-Eval® temperature parameters were set at 560 °C for the CO and CO_2 pyrolysis thermograms, and at 611 °C for the CO_2 oxidation thermograms (Cécillon et al., 2018; Supplementary Material Fig. 2). R scripts used for computing Rock-Eval® parameters are available on the Zenodo platform (Cécillon, 2021b).

2.4. The PARTY$_{SOC}$ machine-learning model

The most up-to-date European version of this model, calibrated on soils from North-western Europe, used in this study, is described in detail in Cécillon et al. (2021). This model uses the 18 above-mentioned Rock-Eval® thermal analysis parameters as predictors and estimates the centennially stable SOC proportion in a topsoil sample. The model consists of a trained non-parametric machine learning algorithm, using the random forests approach to estimate centennially stable SOC proportions in unknown topsoils from centred and scaled Rock-Eval® parameters. In this study the obtained centennially stable SOC proportion of each topsoil sample, was converted to centennially stable SOC content by multiplying the predicted proportion by the total SOC content. The PARTY$_{SOC}$v2.0$_{EU}$ model, available on Zenodo (Cécillon, 2021b), was used without any adaptation.

2.5. The AMG model of soil organic carbon dynamics

The AMG model (Andriulo et al., 1999) was developed based on the two-compartment SOC model proposed by Hénin and Dupuis (1945). It is characterized by a simple structure consisting of three carbon pools: fresh organic matter, and two SOC fractions, an active and a stable pool (Supplementary Material Fig. 2). The model allows transfer of carbon from the fresh organic matter pool either to the atmosphere through microbial mineralization or into the active pool. Organic carbon from the active pool is also subject to mineralization, forming a second direct flux of CO_2 from the soil into the atmosphere. SOM decomposition follows first order kinetics with a rate defined by the coefficient of mineralization k (year^{-1}), controlled by climatic

conditions and soil characteristics. The *h* coefficient controls the yield of crop residues transformation into active carbon and depends on the type of fresh organic matter. No carbon exchange with the stable SOC pool is possible since it is considered inert and remains unchanged over the simulation period (here from 12 to 41 years; see Table 1). Considering the stable SOC pool as mathematically inert at this time scale is in line with consistent observations of a significant pluri-decadal persistent SOC fraction in long-term bare fallows and C_3-C_4 vegetation change chronosequences (Barré et al., 2010; Balesdent et al., 2018).

The AMG model can be mathematically described by two simple equations (Clivot et al., 2019):

$$QC = QC_S + QC_A \,, \tag{1}$$

$$\frac{dQC_A}{dt} = \sum_i m_i \, h_i - \, k \cdot QC_A \,, \tag{2}$$

where QC is the total SOC stock ($t \cdot ha^{-1}$), QC_S is the stable SOC stock ($t \cdot ha^{-1}$) defined as a fraction of initial SOC stock QC_0 (s. Sect. 2.6) constant for a specific treatment, QC_A is the active SOC stock ($t \cdot ha^{-1}$), t is the time in years, m_i is the annual C input from organic residue i ($t \cdot ha^{-1} \cdot yr^{-1}$), h is a coefficient representing the fraction of C inputs which is incorporated in SOM after 1 year related to the type of organic residue, and k is the mineralization rate constant associated with the active C pool (yr^{-1}).

The AMG parameters (*h* and *k*) have been determined by experimental results (Clivot et al., 2019). This approach differs from most multi-compartmental SOC dynamics models for which decay rates of slower pools were calibrated indirectly, assuming an equilibrium state for SOC (Wutzler and Reichstein, 2007). The simple structure of the AMG model and the experimental determination of its decomposition rates make it less susceptible to the problem of equifinality compared to other multi-compartmental models of SOC dynamics (Luo et al., 2016; Clivot et al., 2019). Furthermore, AMG has been validated with $\delta^{13}C$ tracer data of long term alternative sequences of C_4 and C_3 crops (Mary et al., 2020).

The version of AMG used in this study was AMGv2, described in detail in (Clivot et al., 2019). Input data necessary to run simulations of SOC stocks with AMG include crop type, annual crop yields and information regarding management of crop residues. These are used to compute annual aboveground and belowground C inputs from plants, here according to the method proposed by (Bolinder et al., 2007) and adapted by (Clivot et al., 2019).

The coefficient of mineralization k (year^{-1}) is calculated according to soil characteristics (clay and carbonate contents, pH and C:N ratio) and climatic conditions (mean annual temperature, precipitation and potential evapotranspiration; Clivot et al., 2019).

2.6. Soil organic carbon pool partitioning in the AMG model

2.6.1. Default CS/C0 initialization

Two default values can be used for initialization of SOC pool distribution in AMG, depending on land-use history before the onset of simulations. The initial proportion of C_S/C_0 equals 0.65 for sites with a long-term arable land-use history. Former long-term grassland sites are expected to have lower C_S/C_0 and the value of 0.40 was used in previous studies (Saffih-Hdadi and Mary, 2008; Clivot et al., 2019). Since all sites used in this study had been under arable land for at least 20 years before the onset of the experiment, a default value of 0.65 was used.

2.6.2. PARTY$_{SOC}$-based initialization of C_S/C_0

The PARTY$_{SOC}$-based initialization of C_S/C_0 was derived from data obtained with Rock-Eval® analysis of initial topsoil samples from each LTE. Here, C_S/C_0 was estimated using the following simple 4-step procedure: first, topsoil samples from the LTE's onset were analysed with Rock-Eval® and the 18 thermal parameters described in Sect. 2.3 were calculated for each sample. Second, the thermal parameters were used as input for the PARTY$_{SOC}$ machine-learning model described in Sect. 2.4 which was run for this sample set resulting in a sample-specific prediction of centennially stable SOC proportion. Third, the obtained values were averaged per LTE. Fourth, the site mean of the centennially stable SOC proportion was used (as C_S/C_0) to initialize simulations of SOC stocks for the various treatments of every site (the site standard deviation is reported on Fig. 28 and in Supplementary Material Table 2). Supported by the evident common land-use history shared by the multiple treatments of each site before the onset of simulations and as the SOC stocks and centennially stable SOC contents were very homogeneous amongst each site, we also performed simulations of 17 treatments for which soil samples from the onset of the LTE were not available. In these cases, we considered that the C_S/C_0 of the treatment was equal to the mean value of the respective site (Supplementary Material Table 1 and 2).

2.6.3. *Ex-post* optimization of C_S/C_0

Following a least squares optimization approach, the best fit of the AMG model on observed SOC stocks time series was obtained and the optimal initial SOC pool partitioning (C_S/C_0) was estimated accordingly for each site (Clivot et al., 2019). In sites with C_3-C_4 vegetation change chronosequences where $\delta^{13}C$ long-term monitoring data were available, the model was adapted to simultaneously match the observed evolution of C, C_3 and C_4 stocks (Clivot et al., 2019) for a given treatment.

2.7. Calculation of the centennially stable SOC content

The content of the centennially stable SOC pool of each LTE at initial, intermediate and final dates was estimated through multiplication of the $PARTY_{SOC}$ estimates of the proportion of the centennially stable SOC at a given date by the corresponding total SOC content previously determined using elemental analysis (Clivot et al., 2019; Levavasseur et al., 2020). For example, for the onset of an LTE where t=0: $C_S = C_S/C_0 \cdot C_0$, where C_S is the stable SOC content (g C kg soil^{-1}), and C_0 is the total SOC content (g C kg soil^{-1}) at time t=0.

2.8. Statistics

The fit between $PARTY_{SOC}$ predictions of centennially stable SOC proportion and AMG *ex-post* optimized stable SOC proportion was assessed by a linear regression model. The same approach was applied for the evaluation of the agreement between centennially stable SOC content and AMG *ex-post* optimized stable SOC content of initial samples. The evaluation of the performance of the AMG model, for the different SOC pool partitioning initialization methods, was also based on simple linear regressions between simulated and observed SOC stock values. Statistical terms used to express the strength and the statistical significance of the relationships were the coefficient of determination (R^2), and the associated probability value (*p-value*). Prediction bias and model error were expressed as mean difference (*BIAS*), and relative mean square error (*RMSE*). The relative root mean square error (*RRMSE*) and the normalized root mean square error (*NRMSE*) were used to compare the error of different data sets (with a different range of predictions; Smith et al., 1996; Wallach, 2006; Otto et al., 2018).

$$R^2 = \left(\frac{\sum_{i=1}^{n}((O_i - \bar{O}) \cdot (S_i - \bar{S}))}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2} \cdot \sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}} \right)^2, \quad (3)$$

$$BIAS = \frac{1}{n} \sum_{i=1}^{n}(S_i - O_i), \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n}(S_i - O_i)^2}, \quad (5)$$

$$RRMSE = \frac{RMSE}{\bar{O}}, \quad (6)$$

$$NRMSE = \frac{RMSE}{O_{max} - O_{min}}, \quad (7)$$

where: O and S are the observed and simulated values, n is the number of observations, $\bar{O}$ and $\bar{S}$ are the means of observations and simulations, respectively, and O_{max} and O_{min} are the maximum and the minimum value observed.

The observed and simulated total SOC stock change *dQC* was calculated as follows for each treatment:

$$dQC_{obs} = QC_{obs,t_2} - QC_{obs,t_1}, \quad (8)$$

$$dQC_{sim} = QC_{sim,t_2} - QC_{obs,t_1}, \quad (9)$$

where QC_{obs} is the observed SOC stock at time *t*, QC_{sim} is the SOC stock at time t simulated with AMG, t_1 indicates the start and t_2 the end of simulation period.

All data processing and statistical analyses were performed within the "R" programming environment (version 3.4.2; R Core Team, 2017). For plotting, packages ggpmisc (Aphalo, 2016), reshape2 (Wickham, 2007) and ggplot2 (Wickham, 2016) were used.

3. Results

3.1. Accurate soil organic carbon pool partitioning

Centennially stable SOC proportion values were predicted by the $PARTY_{SOC}$ machine learning model (Cécillon et al., 2021) using Rock-Eval® data measured on initial topsoil samples. The mean value for each independent site was plotted against the stable SOC proportion as determined by AMG *ex-post* optimization (Fig. 28). The initial centennially stable SOC proportion values predicted with $PARTY_{SOC}$ ranged from 0.44 to 0.74, with a mean value of 0.59, whereas AMG optimal *ex-post* estimations of stable SOC proportion covered almost the same range, from 0.45 to 0.74, with a mean value of 0.61. The two approaches were strongly correlated ($R^2 = 0.63$, significant at the $p < 0.05$ level), with a linear regression slope close to 1 ($a = 0.9$) and intercept close to 0 ($b = 0.04$), showing an unbiased relationship between $PARTY_{SOC}$ estimates of the centennially stable SOC proportion and the AMG *ex-post* optimized stable SOC proportion at the onset of the nine LTEs. Although a slight discrepancy was observed for higher stable SOC proportion values, the results validate our hypothesis showing that centennially stable SOC proportion determined by Rock-Eval® thermal analysis and the $PARTY_{SOC}$ machine-learning model built on fully independent data provides a good estimate of the optimal stable SOC proportion of the AMG model for unrelated French agricultural soils. When expressed as content (g C kg $soil^{-1}$), the fit between the $PARTY_{SOC}$ predictions of the centennially stable SOC determined on initial topsoil samples and the *ex-post* optimized stable SOC content values was excellent ($R^2 = 0.95$; Supplementary Material Fig. 4; optimal stable SOC content ranged from 4.37 to 12.75 g C kg $soil^{-1}$ across the nine sites). Furthermore, the method appears to be reliable since additional Rock-Eval® measurements on topsoil samples from intermediate and final dates of the LTEs showed that the $PARTY_{SOC}$ predictions of the centennially stable SOC content remained remarkably constant during the experimental period at most sites (Supplementary Material Fig. 5).

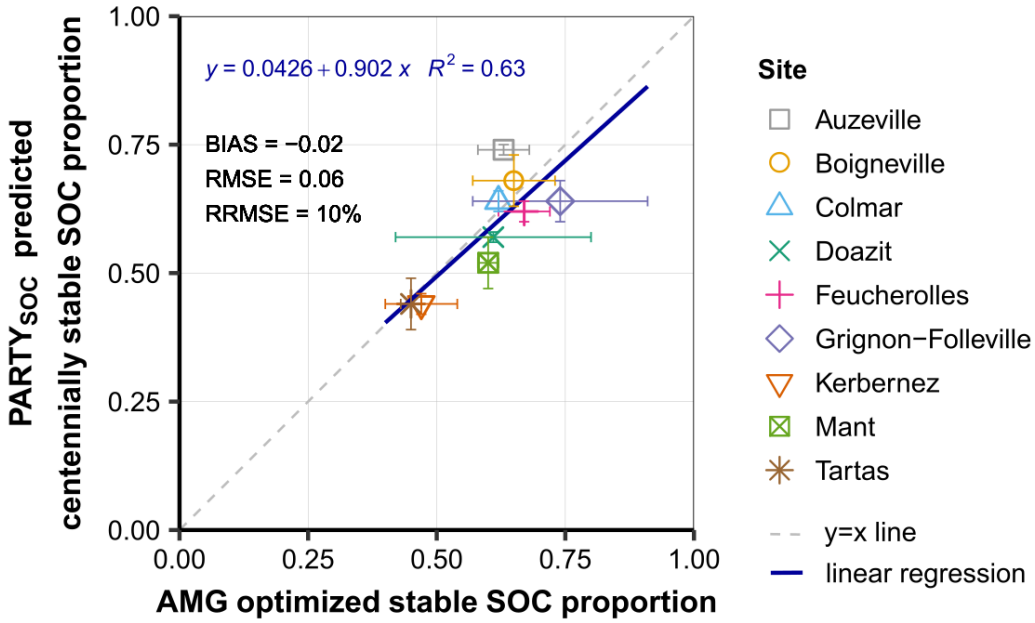


Figure 28 : Performance of the $PARTY_{SOC}$ model to predict the centennially stable SOC proportion compared to the AMG *ex-post* optimized stable SOC proportion. Points represent site-mean values based on initial topsoil samples from nine independent French long-term experiments. Statistics refer to the linear regression between x and y values (blue solid line). Horizontal error bars show the uncertainty associated with the AMG optimal stable SOC proportion, calculated as the standard deviation of treatment-wise AMG optimizations. Vertical error bars represent the prediction error of the centennially stable SOC proportion values, calculated as the standard deviation of the $PARTY_{SOC}$ model predictions on initial topsoil samples.

3.2. More accurate soil organic carbon simulations

In a second step, we investigated if a $PARTY_{SOC}$-based initialization of the SOC pool partitioning could improve the accuracy of SOC stock simulations of the AMG model. To do so, we compared SOC stock simulations obtained with three different initializations. We first ran AMG using the default initialization method for the SOC pool partitioning (C_S/C_0 = 0.65 since all LTEs were under cropland for at least two decades before their onset; Table 1). Then, we ran AMG simulations using the $PARTY_{SOC}$-based initialization method by defining C_S/C_0 as the site-mean centennially stable SOC proportion determined by the $PARTY_{SOC}$ model. Finally, we ran AMG using the *ex-post* optimization method to initialize the SOC pool partitioning for each site. For all three initialization procedures, the simulated SOC stock change between the initial and last sampling date for each treatment of each site was plotted against the measured SOC stock change (Fig. 29a–c). Observed SOC stock change ranged from +6 to −24 Mg C ha^{-1} for the 32 treatments. In spite of a rather good mean agreement (RMSE = 5.95 Mg C ha^{-1}), the AMG model initialized with the default procedure provided predictions of SOC stock change rather far from what was observed in two out of nine LTEs (Fig. 29a). Using the $PARTY_{SOC}$-based initialization method improved AMG simulations compared to the default method, bringing them much closer to the observed SOC stock changes (RMSE = 3.60 Mg C ha^{-1}; Fig. 29a, b). $PARTY_{SOC}$-based initialization of AMG resulted to unbiased simulations (BIAS = 0.06 Mg C ha^{-1}) and a strong decrease in the mean error of prediction. Unsurprisingly, AMG initialized using *ex-post* optimized C_S/C_0 proportions predicted SOC stock changes very close to the observed ones (RMSE = 2.12 Mg C ha^{-1}; Fig. 29c). AMG simulations from *ex-post* optimized and $PARTY_{SOC}$-based initializations were remarkably comparable (Fig. 29b, c). The SOC stock simulations produced with AMG for each independent treatment are presented in Supplementary Material Fig. 6.

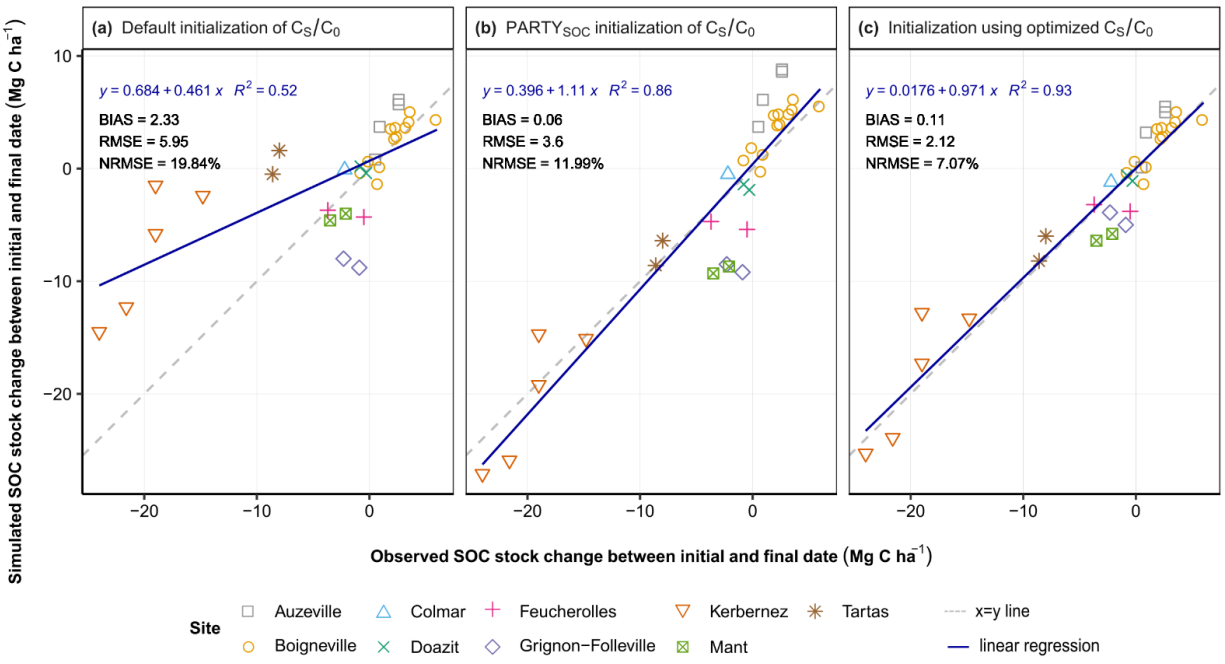


Figure 29: Observed *vs.* simulated change in SOC stocks between the initial and final date of 32 treatments from nine French long-term experiments. The three panels show the performance of the AMG model for three different initialization approaches. Initial SOC kinetic pool sizes were defined using a, the default value for cropland (C_S/C_0 = 0.65), b, the centennially stable SOC proportion predicted by the $PARTY_{SOC}$ model and c, the AMG *ex-post* optimized C_S/C_0 proportion. Statistics refer to the linear regression between x and y values (blue solid line). Points represent the values for the 32 treatments for which AMG simulations were run.

It is noteworthy that the $PARTY_{SOC}$-based initialization improved the fit between observed and simulated SOC stock change, compared to AMG default initialization, especially for treatments that experienced the greatest SOC stock loss (Fig. 29a, b). In treatments that experienced no SOC stock change or a slight increase in SOC stock, the $PARTY_{SOC}$-based initialization did not improve the simulations but resulted in highly reliable predictions, similarly to AMG default or optimized initialization methods (Fig. 29a–c). This is likely explained by the history of land cover and soil management practices of the different sites. Sites presenting treatments with no change or a slight increase in SOC stocks were predominantly sites with a long cropland history (*e.g.* site of Boigneville; Supplementary Material Table 1), for which the default AMG C_S/C_0 value of 0.65 is nearly optimal. Conversely, the two sites, Kerbernez and Tartas, where the *ex-post* optimized C_S/C_0 values were far below the default value (Fig. 28) have a more complex history of land use and soil management practices. The site of Kerbernez is a former grassland (during the first half of the 20^{th} century; Supplementary Material Table 1) that was converted into a cropland two decades before the implementation of its arable LTE, in 1958. The site of Tartas was cultivated for a longer time before the LTE onset, however it was turned to grassland for a period in the 19^{th} century (Supplementary Material Table 1) and received applications of poultry manure for several years before the LTE began. In these two sites, characterized by an optimal AMG C_S/C_0 much lower than the default value, the $PARTY_{SOC}$ machine-learning model predicted values very close to the optimal C_S/C_0 values (Fig. 28).

4. Discussion

Our study demonstrates that the PARTY$_{SOC}$ method based on Rock-Eval® thermal analysis (Cécillon et al., 2018, 2021) can estimate the initial SOC pool partitioning of the AMG model of SOC dynamics while improving its accuracy in a series of diverse and independent French LTEs. Contrary to previous studies (Skjemstad et al., 2004; Luo et al., 2014), no modifications of the decomposition rate of SOC kinetic pools were necessary to improve model predictions. The PARTY$_{SOC}$ initialization method never severely affected the model simulations while it strongly improved them at sites where SOC stocks were far from an equilibrium state due to historical changes in soil management or land use. Areas with past changes in land use and soil management represent a large yet poorly known part of arable land in France and Europe (Fuchs et al., 2015; Erb et al., 2017) where SOC stocks and slow-cycling SOC pools are far from equilibrium (Wutzler and Reichstein, 2007; Herbst et al., 2018; Clivot et al., 2019; Taghizadeh-Toosi et al., 2020). Therefore, by accounting for these legacy effects of site history on SOC pool partitioning, the PARTY$_{SOC}$-based initialization of the AMG model should result in more accurate simulations of SOC dynamics at a national or continental scale.

Our findings combined with results reported in recent ensemble modelling studies (Martin et al., 2019; Farina et al., 2021), suggest that despite its simple structure and when properly initialized (*e.g.* using the PARTY$_{SOC}$ model) the AMG model is unsurpassed for predicting observed SOC stock changes in French agricultural LTEs, and is amongst the best available modelling frameworks of SOC dynamics in European arable land (Martin et al., 2019; Farina et al., 2021). Our results demonstrate that there is still potential to increase the accuracy of simple multi-compartmental models of SOC dynamics, bringing their simulations very close to the observed values of SOC stock changes. Developing other Rock-Eval®-based initialization methods specifically designed to match the carbon pool design of other multi-compartmental SOC dynamics models such as RothC (Coleman et al., 1997) is a promising research area. More generally, we recommend that the potential of multi-compartmental SOC dynamics models be fully explored and exploited by soil biogeochemists before a new generation of models of increased complexity becomes operational. While new models including the diversity of microbial communities and related processes are emerging (Crowther et al., 2019; Lehmann et al., 2020), the uncertain structure and parametrization of more complex models is hindering their application as robust predictive tools (Shi et al., 2018).

At the same time, simple conceptual models of SOC dynamics like AMG combined with novel initialization methods and data-based approaches such as PARTY$_{SOC}$ show promising improvements (Cécillon, 2021a; Dangal et al., 2021; Lee et al., 2021). The low prediction error of the AMG model when its SOC pool distribution is initialized with the PARTY$_{SOC}$ method even challenges the ability of more complex modelling approaches to achieve better performance, given the uncertainty on observed values of SOC stock changes (Schrumpf et al., 2011).

The continental or worldwide implementation of the AMG model with the PARTY$_{SOC}$-based initialization of SOC pools distribution will require additional work. First, the PARTY$_{SOC}$ machine-learning model (Cécillon et al., 2018, 2021) will have to be validated on a wider range of pedoclimates. This method initially built on LTEs coming from North-western Europe (Cécillon et al., 2018), has now been successfully extended to new soil types and a new climate (tropical; Cécillon et al., 2021). The good agreement between optimal AMG stable SOC proportion values and PARTY$_{SOC}$ predictions reported here suggests that most agricultural LTEs with accurate AMG simulations could be used as reference sites for the PARTY$_{SOC}$ model, lifting an important technical limitation to its geographical expansion (Cécillon et al., 2021). Second, the improved accuracy of model simulations using a PARTY$_{SOC}$-based initialization will also have to be demonstrated for a wider pedoclimatic range (*i.e.* worldwide LTEs; such as those referenced by the International Soil Carbon Network; Nave et al., 2015). Third, Rock-Eval® data from the new application areas will be required. Rock-Eval® is a high-throughput technique that is well adapted to the analysis of large soil sample sets provided by large-scale soil monitoring programs. We recommend implementing Rock-Eval® measurements in national and continental soil monitoring networks.

5. Conclusions

Combining Rock-Eval® thermal analysis with the PARTY$_{SOC}$ machine-learning model should be considered as an emerging key approach with demonstrated ability to improve the accuracy of simulations of SOC dynamics, complementary to other SOC cycling proxies (Bailey et al., 2018; Wiesmeier et al., 2019). The progressive large-scale delivery of these complementary data related to SOC dynamics will strengthen model predictions of SOC stock changes at the national to global scale, necessary for implementing efficient climate change mitigation policies (FAO, 2020).

CHAPTER 2

On the harmonisation of Rock-Eval® data

The exponential increase of studies focusing on SOM, kick-started by the realization of its importance for climate change mitigation, brought along an enormous amount of data recorded on soil samples collected using different sampling strategies (notably different sampling depth intervals). Harmonization of the available data is more imperative than ever for the correct incorporation of information into large scale studies including SOC modelling and mapping of SOC stocks and SOC sequestration potential.

This chapter addresses the question of data comparability amongst soil samples obtained at different depths. Using samples from 10 plots, located at 8 forest sites in France, we conducted a soil mixing experiment, aiming at finding an appropriate method for calculating Rock-Eval® parameters of a soil profile (0–50 cm) by combining Rock-Eval® results recorded on its sublayers (0–30 and 30–50 cm).

We hypothesized that in temperate soils most Rock-Eval® parameters will be independent of the sampling strategy (i.e., that Rock-Eval® parameters characteristic of SOM of a given soil layer could be obtained from Rock-Eval® parameters measured in its sublayers), as we expect the mineral matrix effect to be low in these soils. However, it was also anticipated that mixing of contrasting materials might cause changes on the Rock-Eval® signal of composite samples.

The Rock-Eval® parameters measured on composite samples were generally in good agreement with the calculated ones. However, for parameters derived from the hydrocarbon signal (S2) the relationship between measured and calculated values was unsatisfactory for some sites. These were sites with a very clay-rich deep soil layer and a surface layer with a coarser texture, where mixing caused an addition of clays leading to increased hydrocarbon retention by the mineral matrix during pyrolysis. Therefore, in the context of the ongoing work towards harmonization of databases, we concluded that in most temperate soils, Rock-Eval® characterization of a soil layer, including its centennially persistent carbon proportion can be inferred from sublayers characteristics.

Regarding clays and their interference with the hydrocarbon signal, we provide an empirically obtained threshold value of 20% difference in clay content between sub-layers above which mixing of soil samples might cause changes in the hydrocarbon thermogram shape and related parameters.

In addition, in this chapter we discuss the importance of an often neglected error introduced in many studies by the application of linear mixing equations on ratios. This error is of general interest and is not limited to soil studies. Moreover, since many Rock-Eval® parameters are complex terms, involving multiple acquisition parameters and correction factors, attention must be paid when calculating a mixed value. A simple approach is appropriate for most classic Rock-Eval® parameters with small adaptations depending on their defining equation.

Finally, a last guideline was derived from this work, regarding a specific category of parameters; temperature parameters. These are by definition particularly sensitive to changes on the thermogram shape and should be calculated after signal reconstruction.

Some interesting perspectives for future research could be to actually quantify the effect of different soil mineral matrices on the Rock-Eval® signal, to better understand the interactions between SOM and minerals. A more detailed insight could be gained by measuring specific surface area where this is possible either of the whole soil or of different fractions. As in this study the mineral effect was negligible in most cases, it was not possible to study the effect of changes in mineralogy rather than or in addition to texture. Using temperate and tropical soils or artificial soil samples with different composition but similar texture could be helpful for gaining a more detailed understanding of the importance of different soil characteristics and testing the limits of the Rock-Eval® approach. More perspectives and ideas for improvement are presented in the general discussion.

A draft presenting this work entitled “Predicting Rock-Eval® thermal analysis parameters of a soil layer based on samples from its sublayers; an experimental study on forest soils” was published in Organic Geochemistry and can be found on the pages that follow.

Predicting Rock-Eval® thermal analysis parameters of a soil layer based on samples from its sublayers; an experimental study on forest soils.

Eva Kanari[a,b*], Pierre Barré[b], François Baudin[a], Alain Berthelot[c], Nicolas Bouton[d], Frédéric Gosselin[e], Laure Soucémarianadin[f], Florence Savignac[a], Lauric Cécillon[b,g]

[a]Institut des Sciences de la Terre de Paris, Sorbonne Université, CNRS, 75005 Paris, France
[b]Laboratoire de Géologie, CNRS, École normale supérieure, PSL University, IPSL, Paris, France
[c]FCBA, Délégation territoriale Nord-Est, 21170 Charrey-sur-Saône, France
[d]VINCI Technologies, 92022 Nanterre, France
[e]INRAE, UR Ecosystèmes Forestiers, 45290 Nogent-sur-Vernisson, France
[f]ACTA – les instituts techniques agricoles, 75595 Paris, France
[g]Univ. Normandie, UNIROUEN, INRAE, ECODIV, Rouen, France

Abstract

Soil sampling depths strongly vary across soil studies. Stocks of elements (such as C, N) or organic matter in a soil layer can be simply calculated from stocks measured in its sublayers. This calculation is less obvious for other soil characteristics, such as soil organic carbon (SOC) persistence, complicating the comparison of results from different studies. Here, we tested whether Rock-Eval® parameters of a soil layer, characterizing soil organic matter and its biogeochemical stability, can be determined using Rock-Eval® data measured on its sublayers. Soil samples collected in 10 plots located in eight French forest sites, taken up at two different depths (0–30 cm, 30–50 cm), and their mixtures were analysed with Rock-Eval®. Expected values for the Rock-Eval® parameters of the soil mixtures were calculated either: (1) as the weighted mean of Rock-Eval® parameters measured on the two sublayers, or (2) based on a signal reconstructed as the weighted mean of Rock-Eval® thermograms recorded on the two sublayers. Our results showed a good agreement between measured and expected Rock-Eval® parameter values. However, when the clay content strongly differed between the two soil sublayers, the amount of pyrolyzed hydrocarbons measured on the soil mixtures was slightly lower than expected. We conclude that it is reasonable to calculate Rock-Eval® parameters of a soil layer, from the Rock-Eval® signature of its sublayers. Our findings facilitate the harmonization of Rock-Eval® data from large scale soil studies using different sampling depths.

1. Introduction

Recognition of the central role of soil and soil organic matter (SOM) for food security and ecosystem functioning (FAO, 2005, 2017) has led to enhanced political and scientific attention on soil health and SOM dynamics (e.g., Global Soil Partnership, Montanarella, 2015; International Year of Soils, FAO 2015; 4per1000 initiative; 4p1000, 2018; Rumpel et al., 2018). As a result, over the last five decades the number of studies focusing on SOM has grown exponentially (Smith et al., 2017). The first large-scale soil monitoring projects with focus on soil organic carbon (SOC) were initiated (e.g., RMQS in France, Arrouays et al., 2003; LUCAS in Europe, European Commision Joint Research Centre, 2021; Orgiazzi et al., 2018; WoSIS, a worldwide soil profile database, Batjes et al., 2020; ISRIC, 2021) and an enormous amount of soil samples and data were collected worldwide, without any harmonization regarding the depth of the sampled soil layers. According to the research question, each soil study may follow a different strategy for the selected sampling depths, with possible complications regarding the comparability of available information. For instance, in some studies the 0–30 cm layer is divided in several sublayers (e.g., 0–10 cm, 10–20 cm and 20–30 cm) or sampled according to pedogenetic horizons, whereas in some other cases the 0–30 cm layer is sampled as one homogeneous layer. Independent of the sampling method, the SOC stock of a soil profile is calculated as the sum of the SOC stocks obtained in the layers comprising that profile (FAO, 2018). Even though determining SOC stock of a given layer is straightforward, it may be much less obvious for other parameters such as SOM characterization results.

One promising analytical technique in SOM research is Rock-Eval® thermal analysis. A time-efficient and inexpensive method, it can be used on large sample sets to quantify soil organic and inorganic carbon and characterize SOC bulk chemistry and thermal stability (Saenger et al., 2013; Gregorich et al., 2015; Sebag et al., 2016; Soucémarianadin et al., 2018). Moreover, a strong empirical link exists between parameters obtained with this method and in situ observed SOM biogeochemical stability (Barré et al., 2016; Poeplau et al., 2019). Encouraged by this observation, Rock-Eval® thermal analysis was used to build a non-parametric random forests algorithm, $PARTY_{SOC}$, providing quantitative information on SOC biogeochemical stability, by predicting the proportion of centennially stable SOC (Cécillon et al., 2018, 2021). As the application range of the Rock-Eval® method is increasingly expanding, samples from different projects and different sampling strategies are being analysed.

Therefore, it is of great importance to determine whether Rock-Eval® parameters characterizing SOM of a given soil layer can be determined from Rock-Eval® parameters measured on its sublayers and, similarly, if centennially stable SOC assessed using $PARTY_{SOC}$ in a given soil layer can be determined from $PARTY_{SOC}$ estimates based on measurements on its sublayers. If true, this would simplify the comparison of data between studies and the building of datasets on harmonized sampling depths, for instance the 0–30 cm layer, a standard for soil studies (IPCC, 2006; FAO, 2018).

Rock-Eval® analysis consists of thermal treatment of a soil sample following a predefined temperature ramp. Simultaneous and continuous detection of effluents generates five thermograms in total that describe the evolution of carbon containing gases (HC, CO and CO_2) during the analysis. A large number of Rock-Eval® parameters can be calculated from the five thermograms (Behar et al., 2001; Saenger et al., 2013; Sebag et al., 2016; Cécillon et al., 2018, 2021; Khedim et al., 2021). Parameters obtained with this method are characteristic of the SOM and its interaction with the soil mineral matrix, since soil samples are analysed with no previous isolation of SOM or removal of carbonates. We hypothesised that, in a large variety of temperate forest soils Rock-Eval® parameters characteristic of SOM of a given soil layer could be obtained from Rock-Eval® parameters measured on its sublayers. This hypothesis is far from trivial as differences in the composition of the mineral matrix can influence the Rock-Eval® signal (Espitalié et al., 1980, 1984). This could considerably complicate the assessment of Rock-Eval® parameters of a given soil layer from Rock-Eval® parameters measured on its sublayers and comparison of results amongst studies with different sampling depth protocols.

To test this hypothesis, we conducted Rock-Eval® analysis of surface (0–30 cm) and deeper (30–50 cm) samples from 10 plots located in eight forest sites in France as well as various binary mixtures of the two soil layers. Some parameters (i.e., ratios) required a minor adaptation of the calculation method used to determine Rock- Eval® parameters of the mixtures based on Rock-Eval® results obtained for the two sublayers, while for other parameters (i.e., temperature parameters) a new calculation method was introduced. Finally, this study provides some basic guidelines that facilitate the direct comparison of Rock-Eval® data of samples from different depths.

2. Materials and methods

2.1. Sampling sites and soil characteristics

Samples used in this study belong to 10 contrasting plots distributed amongst seven forest sites of the French "GNB" network (Gosselin et al., 2015; 9 plots) and one site of the "FCBA" network (Berthelot, 2018); 1 plot; part of the IN-SYLVA France national research facility (INRAE, 2020); Fig. 30, site map; Cécillon et al., 2017). At each plot, two composite soil layers (layer A (0–30 cm) and layer B (30–50 cm)) were sampled, resulting in two "natural" samples per plot (n = 20 samples for the 10 plots). Soil samples were dried, sieved to < 2 mm and ground to < 250 µm before further processing. Basic characteristics of the 20 "natural" samples are provided in Table 2. Soil bulk chemistry was highly variable with organic carbon content ranging from 0.38 to 11.28 wt.%, pH measured in water was between 4.1 and 8.2 and carbonate content was negligible in all samples except one that contained 35 wt.% $CaCO_3$. The sample selection covered a wide range of soil textures with sand content varying between 4 and 85 wt.% and clay content between 2 and 55 wt.% (Table 2).

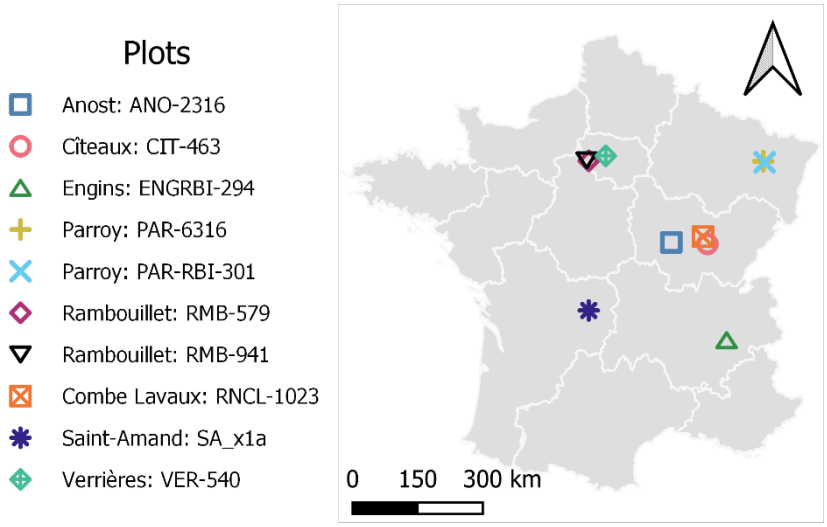


Figure 30: Map of the 10 plots used in this study, located in eight forest sites in France.

2.2. Soil mixtures

Using the two "natural" soil samples available from each plot, we composed five binary soil mixtures per plot by combining 90:10, 75:25, 50:50, 25:75 and 90:10 in mass proportion of soil from layer A and layer B (Equation 1; n = 50 binary mixtures for the 10 plots). The exact used soil mass was measured using a precision scale whereas the weighting error allowed was in average 0.39% (calculated as the mean weighted error). In addition, two pure end-member samples per plot were set aside, composed of 100% layer A and 100% layer B soil respectively, resulting in a total number of 70 samples.

$$Mixture_{AB} = Layer_A \cdot f_i + Layer_B \cdot (1 - f_i), \quad \textbf{(1)}$$

where $Layer_A$: soil surface layer (0–30 cm), $Layer_B$: soil sub-surface layer (30–50 cm), f_i: mixing ratio = 1, 0.9, 0.75, 0.5, 0.25, 0.1, 0 ± 0.39%

2.3. Rock-Eval® thermal analysis

All 70 soil samples including the 20 "natural" samples and the 50 composed mixtures were analysed using a Rock-Eval® 6 Turbo apparatus (Vinci Technologies). A small amount of soil (approx. 60 mg), was required for the analysis that took place in two consecutive steps, during which carbon-containing effluents were directly detected. First, the sample underwent pyrolysis in an inert (N_2) atmosphere, and then oxidation in the presence of O_2 (ambient air). The heating routine applied during pyrolysis was as proposed by Disnar et al. (2003), including a three-minute isotherm at 200 °C, succeeded by a heating ramp of 30 °C·min^{-1} up to 650 °C. Oxidation started with a one-minute 300 °C isotherm followed by a heating ramp of 20 °C·min^{-1} up to 850 °C and a final five-minute isotherm at 850 °C (oxidation routine presented in Baudin et al. (2015) as "bulk rock/basic" method). Simultaneous detection of effluents during both analysis steps generated five thermograms per sample in total describing the evolution of hydrocarbons during pyrolysis (HC_PYR), and CO and CO_2 during both pyrolysis and oxidation steps (CO_PYR, CO2_PYR, CO_OX, CO2_OX). A schematic representation of the analysis steps and the generated thermograms is presented in Supplementary Fig. 1.

Table 2: Information on the geographical location, soil type, soil texture and soil chemistry of the study plots. The reference coordinate system used here was WGS84 and texture analysis was done without previous decarbonisation of the samples. C org EA is the organic carbon content measured by Elemental Analysis.

Plot ID	Sampling depth (cm)	Site	Latitude	Longitude	Soil type (WBR 2014)	Clay (wt.%)	Fine silt (wt.%)	Coarse silt (wt.%)	Total silt (wt.%)	Fine sand (wt.%)	Coarse sand (wt.%)	Total sand (wt.%)	pH (water)	$CaCO_3$ (wt.%)	C org EA (wt.%)
ANO-2316	0–30	Anost	47.10216	4.06001	Dystric Cambisol	20.8	24.3	10.3	34.6	7.6	37.0	44.6	4.1	0	11.28
	30–50					9.8	28.1	12.0	40.1	8.2	42.0	50.2	4.49	0	5.23
CIT-463	0–30	Cîteaux	47.07216	5.06185	Luvic Stagnosol	11.6	28.4	31.5	59.9	24.9	3.6	28.5	4.75	0	1.65
	30–50					10.6	27.9	31.8	59.7	26.2	3.5	29.7	4.71	0	0.53
RNCL-1023	0–30	Combe Lavaux	47.21097	4.94095	Calcaric Cambisol (Colluvic)	44.7	35.3	16.5	51.8	2.6	0.9	3.5	6.6	0	3.42
	30–50					33.9	26.8	14.4	41.2	11.4	13.5	24.9	8.19	35.2	0.38
ENG-RBI-294	0–30	Engins	45.23063	5.59497	Calcaric Cambisol	20.9	43.7	27.6	71.3	7.1	0.7	7.8	4.71	0	7.26
	30–50					23.4	40.4	27.8	68.2	7.7	0.7	8.4	5.45	0	2.21
PAR-6316	0–30	Parroy	48.65225	6.61094	Vertic Cambisol	31.2	39.9	13.5	53.4	8.0	7.4	15.4	5.34	0	3.16
	30–50					55.1	25.1	7.8	32.9	5.2	6.7	11.9	6.6	0	0.86
PAR-RBI-301	0–30	Parroy	48.63205	6.65036	Vertic Cambisol	15.6	45.4	26.7	72.1	5.3	6.9	12.2	4.64	0	2.60
	30–50					36.8	33.8	19.8	53.6	4.1	5.5	9.6	5.42	0	0.45
RMB-579	0–30	Rambouillet	48.66017	1.77172	Luvic Stagnosol	16.1	20.1	22.3	42.4	23.1	18.4	41.5	4.65	0	1.77
	30–50					27.2	17.5	20.9	38.4	19.1	15.3	34.4	4.98	0	0.66
RMB-941	0–30	Rambouillet	48.67083	1.71289	Entic Podzol (Arenic)	2.3	5.8	8.0	13.8	55.9	28.0	83.9	4.23	0	1.66
	30–50					3.3	4.7	7.3	12.0	54.2	30.6	84.8	4.73	0	0.45
SA_x1a	0–30	Saint-Amand	45.80000	1.77000	Dystric Cambisol	15.3	14.1	7.6	21.7	20.7	42.3	63.0	4.19	0	4.75
	30–50					11.8	11.1	6.2	17.3	21.7	49.2	70.9	4.54	0	1.18
VER-540	0–30	Verrières	48.75656	2.24006	Stagnic Luvisol	13.7	28.3	46.7	75.0	8.1	3.2	11.3	4.15	0	2.32
	30–50					18.2	26.3	44.9	71.2	8.0	2.6	10.6	4.27	0	0.60

2.4. Measured Rock-Eval® parameters

Although these parameters are calculated on the thermograms obtained from Rock-Eval® analysis of the "natural" samples and the composed mixtures, they will be referred to as "measured parameters" in this manuscript to avoid confusion with "calculated parameters" (defined in section 3 of this chapter). Classic Rock-Eval® parameters were acquired using the RockSix software (Vinci Technologies). These included: six automatically generated "peaks" defined as specific areas of the three pyrolysis thermograms (S1, S2, S3, S3', S3CO and S3'CO; Supplementary Fig. 1, Supplementary Table 1), the amount of pyrolyzed carbon (PC corresponding to the sum of organic C released as HC, CO and CO_2 during pyrolysis), total organic carbon (TOC corresponding to the amount of organic C released during the analysis), mineral carbon (MinC corresponding to the amount of C released through carbonate cracking), hydrogen index (HI corresponding to the ratio of released hydrocarbons to TOC), and oxygen index (OI_{RE6} corresponding to the ratio of the amount of oxygen of organic origin released to TOC). In addition, further parameters used as predictors by the $PARTY_{SOC}v2.0_{EU}$ model were calculated based on the obtained thermograms using R scripts available on Zenodo (https://zenodo.org/record/4446138#.YDe84Xlw2SQ; Cécillon et al., 2021; Cécillon, 2021b). These included: PseudoS1 (the sum of carbon released during the first 200 seconds of pyrolysis—200 °C isotherm—in form of HC, CO and CO_2), the ratio S2/PC (ratio of the amount of hydrocarbons released excluding the first 200 seconds of pyrolysis to pyrolyzed carbon), the ratio PC/TOC, the ratio HI/OI_{RE6}, and ten temperature parameters (e.g., T30, T50, T70, T90) that describe evolution steps, i.e., at which temperature 30, 50, 70 and 90% of a given gas was released. Finally, we calculated three additional parameters proposed in previous soil studies using the Rock-Eval® technique. These were: The I-index (assessing the preservation of thermally labile "immature" hydrocarbons; Sebag et al., 2016), the R-index (describing the proportion of refractory SOM released as hydrocarbons after 400 °C; Sebag et al., 2016), and the thermolabile hydrocarbon index (TLHC-index corresponding to the proportion of hydrocarbons evolved between 200 °C and 450 °C; Saenger et al., 2015). A detailed description of the definition, units and equations used to calculate all parameters can be found in Supplementary Table 1.

2.5. The Rock-Eval®-based PARTY$_{SOC}$ model

In this study, we used the Rock-Eval®-based random forest model PARTY$_{SOC}$v2.0$_{EU}$ (https://zenodo.org/record/4446138#.YDe84Xlw2SQ) proposed by Cécillon et al. (2021). This model was calibrated on data from 6 long-term agricultural experiments including a bare fallow treatment in Northwestern Europe and can predict the centennially persistent SOC proportion in a topsoil sample (0–30 cm). The model requires a set of 18 Rock-Eval® parameters (Supplementary Table 1) characteristic of a sample and provides a prediction of the stable SOC proportion for soils from similar pedoclimates. Here we run the model three times, first using measured Rock-Eval® parameters as predictors (obtained as described in section 2.4.) and then calculated ones obtained by two different methods (defined in section 3). The result of the first run was treated as one of the measured parameters during data processing. In the context of this work we merely aimed at evaluating the possibility of calculating the centennially stable SOC proportion of a soil profile based on its sublayers. As this statistical model does not include any forest or deep soil samples in its learning set nor has it been tested or validated on forest or deeper soil samples, we could not provide any interpretation regarding the absolute measured centennially stable SOC proportion values. Instead, we investigated the possible ways of using the model when combining several soil layers.

2.6. Statistics

To evaluate the performance of the two calculation methods in predicting measured Rock-Eval® parameters for soil mixtures we conducted a statistical assessment of the relationship between measured and calculated values for each parameter. “Natural” samples were excluded from all calculations. The coefficient of determination R^2 was estimated to assess the existence and strength of a correlation between the measured and calculated values. R^2 was accompanied by the *p*-value, indicating the probability of a statistically significant relationship, whereas the threshold of significance level was set at 0.05. The mean difference between calculated and measured values (BIAS) was used to test if the relationship was biased. The root mean square error (RMSE) was used as an expression of the magnitude of the error of prediction, as it is defined as a sum of the residuals or deviations between calculated and measured parameters. To compare not only the two calculation methods but also the uncertainty associated with each parameter independent of unit or range, we estimated the normalized root mean square error (NRMSE) defined as the RMSE divided by the range of observations.

$$R^2 = \left(\frac{\sum_{i=1}^{n} \left((M_i - \bar{M}) \cdot (C_i - \bar{C}) \right)}{\sqrt{\sum_{i=1}^{n} (M_i - \bar{M})^2} \cdot \sqrt{\sum_{i=1}^{n} (C_i - \bar{C})^2} } \right)^2$$

$$BIAS = \frac{1}{n} \sum_{i=1}^{n} (C_i - M_i)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (C_i - M_i)^2}$$

$$NRMSE = \frac{RMSE}{M_{max} - M_{min}}$$

where: M_i and C_i are the measured and calculated values, n is the number of observations, $\bar{M}$ and $\bar{C}$ are the means of measurements and calculations, respectively and M_{max} and M_{min} are the maximum and minimum of measured values.

3. Calculation

The central idea of our experiment is that by knowing the exact mixing ratios used to prepare the soil mixtures and having analysed “natural” samples composed only of Layer A or Layer B we can calculate expected values of Rock-Eval® parameters for the mixtures as a weighted mean of the recorded signals. Since many of the parameters are composite terms, involving multiple thermogram areas, mass balance corrections and ratios (Supplementary Table 1), we had to resort to their defining equations to find the most appropriate way of calculating expected parameters. According to the type of parameter, we calculated expected parameters for soil mixtures: (1) based on acquisition parameters (AP) and (2) for temperature parameters only, based on reconstructed thermograms (RT).

3.1. Rock-Eval® parameters calculated based on acquisition parameters (AP)

Expected values of Rock-Eval® parameters defined as single thermogram areas (e.g., S1, S2, S3, etc.; Supplementary Fig. 1), as a sum of multiple thermogram areas (e.g., PC, TOC; Supplementary Table 1) or as evolution steps (e.g., $T90_{HC\_PYR}$; Supplementary Table 1) were calculated as a weighted mean, following Equation 2. For each parameter and each soil mixture we applied a weighting directly on the corresponding Rock-Eval® parameters measured for the two pure "natural" samples (100% Layer A, 100% Layer B) according to the known exact mixing ratio used for the mixture preparation.

$$Calculated\ Rock\text{-}Eval®\ parameter_{Mixture_{AB}}(AP) = AP_{Layer_A} \cdot f_i + AP_{Layer_B} \cdot (1 - f_i) \quad \textbf{(2)}$$

where AP: Measured Rock-Eval® acquisition parameter

For parameters that represent ratios, a different procedure was followed. Directly applying a weighted mean equation on parameters that represent ratios (i.e., HI, OIRE6, S2/PC, PC/TOC, HI/OIRE6, R-index and TLHC-index; Supplementary Table 1) often leads to mathematical errors (Perdue and Koprivnjak, 2007). More precisely, following Equation 2 would mean neglecting to apply the weighted mean correction on the denominator of the ratio (see Equation 3 below for the basic application of equation 2 to OI_{RE6} — defined as the proportion of O_2 of organic origin relatively to TOC).

$$OI_{RE6\ Layer_A} \cdot f_i + OI_{RE6\ Layer_B} \cdot (1 - f_i) = \frac{O_{2\ Layer_A} \cdot f_i}{TOC_{Layer_A}} + \frac{O_{2\ Layer_B} \cdot (1 - f_i)}{TOC_{Layer_B}} \neq$$

$$Calculated\ OI_{RE6\ Mixture_{AB}}(AP) \quad \textbf{(3)}$$

In this case, this would lead to an overestimation of OI_{RE6}. The correct Equation for ratios Perdue & Koprivnjak (2007), is presented below for OI_{RE6} (Equation 4):

$$Calculated\ OI_{RE6\ Mixture_{AB}}(\mathrm{AP}) = \frac{O_{2\ Mixture\ AB}}{TOC_{Mixture\ AB}} = \frac{O_{2\ Layer_A} \cdot f_i + O_{2\ Layer_B} \cdot (1 - f_i)}{TOC_{Layer_A} \cdot f_i + TOC_{Layer_B} \cdot (1 - f_i)} \quad \textbf{(4)}$$

Having already calculated expected values of thermogram areas with the AP-based method, we made use of them to simplify the calculation approach. We calculated Rock-Eval® parameters representing ratios based on their defining equations (e.g., Equation 5 for OI_{RE6}; or Supplementary Table 1 for further parameters). We simply replaced basic terms in the defining equation (e.g., S3, S3CO, TOC) with previously calculated expected values of Rock-Eval® parameters for the soil mixtures (e.g., $S3_{Mixture\ AB}$ (AP), $S3CO_{Mixture\ AB}$ (AP), $TOC_{Mixture\ AB}$ (AP)).

$$Calculated\ OI_{RE6\ Mixture_{AB}}(AP) = \frac{\left[S3_{Mixture_{AB}}(AP) \times \frac{32}{44}\right] + \left[S3CO_{Mixture_{AB}}(AP) \times \frac{16}{28}\right]}{TOC_{Mixture_{AB}}(AP)} \quad \textbf{(5)}$$

Similarly, when calculating the I-index, applying a weighting to the logarithm itself would not have been mathematically correct, thus the weighting was applied on the areas of the HC_PYR thermogram that are used in the I-index formula (areas used to calculate A1, A2, A3; Sebag et al., 2016).

3.2. Rock-Eval® parameters calculated based on reconstructed thermograms (RT)

The above described way of calculating expected values for temperature parameters, as the weighted mean of temperature parameters measured on "natural" samples, is not mathematically correct. Expected values for temperature parameters of soil mixtures were therefore calculated following a reconstructed thermograms (RT) approach. Based on the mixing ratio that was used to compose each mixture we reconstructed the expected shape of the five thermograms of each soil mixture as the weighted mean of the shapes of the thermograms obtained for the two "natural" samples (Equation 6).

$$Thermogram_{Mixture_{AB}} = Thermogram_{Layer_A} \cdot f_i + Thermogram_{Layer_B} \cdot (1 - f_i) \quad \textbf{(6)}$$

The RT method was used to calculate expected values for temperature parameters only (Cécillon et al., 2018, 2021; Supplementary Table 1), as by definition the AP and RT approaches are equivalent when considering all other parameters, whose calculation is based on thermograms areas. Since the RT approach is more complicated to implement, we also kept the AP method for the calculation of temperature parameters, to test whether the results were very different from what can be obtained using a correct but more time-consuming approach.

Signal processing, integrations of thermogram areas and calculation of Rock-Eval® parameters were conducted using R version 3.4.2 (R Core Team, 2017), the trapz function of the pracma package (Borchers, 2018) and the hyperSpec package (Beleites and Sergo, 2018). The R script used to calculate the reconstructed thermograms is provided as supplementary material (Supplementary R script 1).

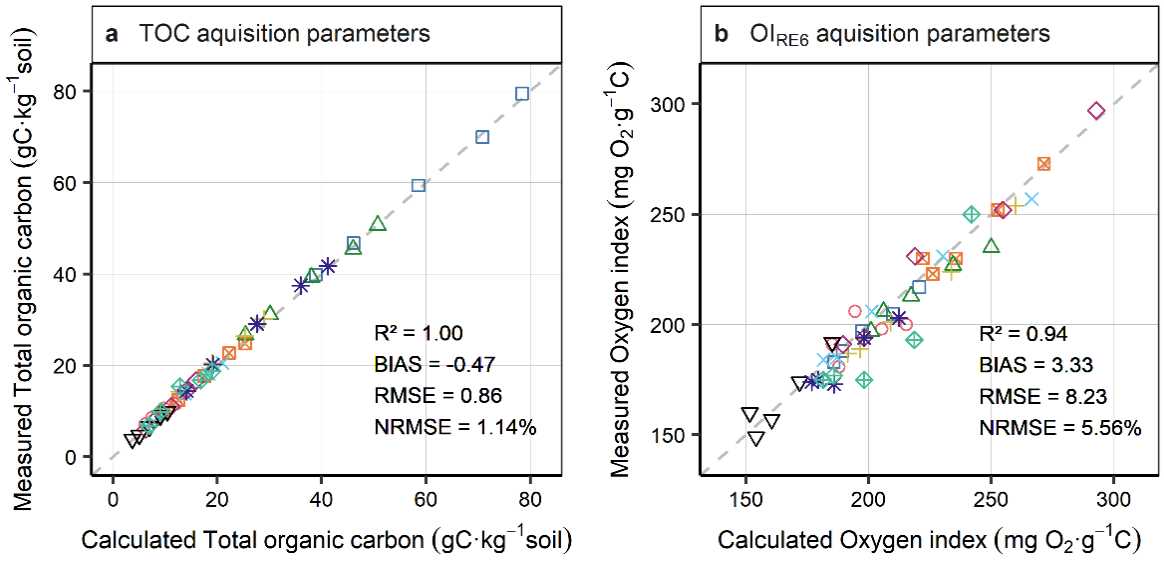
3.3. Three ways of calculating stable SOC proportion of soil mixtures using the PARTY$_{SOC}$v2.0$_{EU}$ model

Expected values of stable SOC proportion of soil mixtures were calculated in three different ways. First, the stable SOC proportion of soil mixtures was calculated as a weighted average of PARTY$_{SOC}$v2.0$_{EU}$ predictions obtained for layer A and layer B. Second, the stable SOC proportion of the soil mixtures was predicted by running the PARTY$_{SOC}$v2.0$_{EU}$ model, using calculated Rock-Eval® parameters as predictors, obtained with the AP method. Third, we run the PARTY$_{SOC}$v2.0$_{EU}$ model using the calculated predictors (AP method) but replacing temperature parameters with ones obtained with the RT method.

4. Results

4.1. A general agreement between calculated and measured Rock-Eval® parameters

Comparison of calculated and measured Rock-Eval® parameters showed that the AP method was able to accurately predict the values of Rock-Eval® parameters for the soil mixtures (overall median NRMSE obtained for the 29 parameters of 8.5%; Supplementary Table 2). For classic Rock-Eval® parameters such as TOC there was a perfect correlation between measurements and predictions (R^2_{AP} = 1), with a negligible bias ($BIAS_{AP}$ = −0.47 g C · kg^{-1} soil) and a very low percentage error ($NRMSE_{AP}$ = 1.1%; Fig. 31a). The calculated values of OI_{RE6} were also very close to the measured ones ($NRMSE_{AP}$ = 5.6%, Fig. 31b). An overestimation was observed for HI, for two out of ten plots ($BIAS_{AP}$ = 17 mg HC · g^{-1} C; Fig. 31c). This effect increased with distance of the mixture composition from the "natural" samples and was the strongest for the 50:50 soil mixtures. However, the overall observed relationship between calculated and measured values was still very strong (R^2_{AP} = 0.90). The correlation between measured $T90_{HC\_PYR}$ and calculated with the AP method was equally strong (R^2_{AP} = 0.91) and the relationship was unbiased with a low percentage error ($BIAS_{AP}$ = 1.66 °C, NRMSE = 8.8%; Fig. 31d).



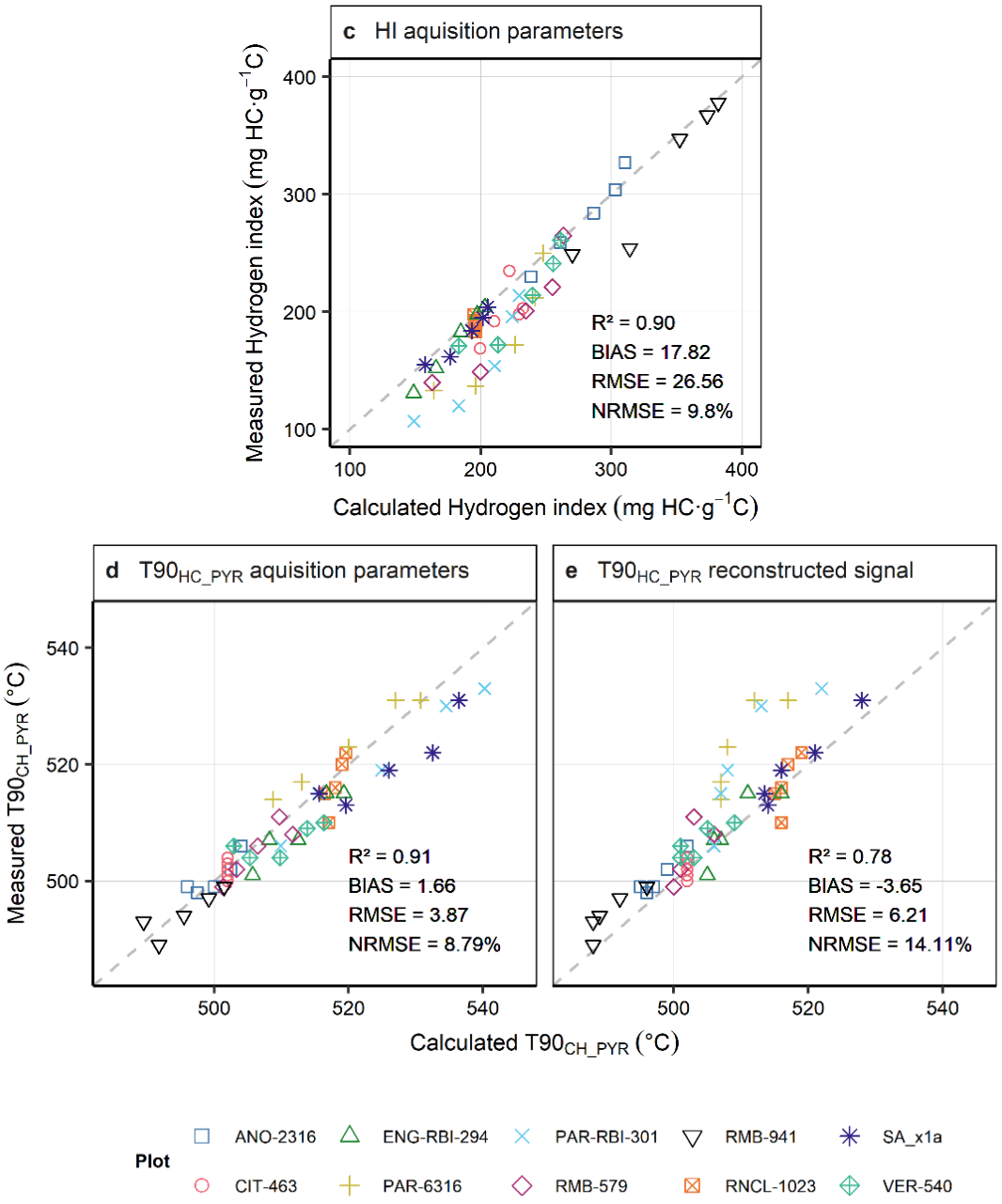


Figure 31: Comparison of measured and calculated values for four Rock-Eval® parameters, presented in three simple plots and one twin plot. The measured values of Rock-Eval® parameters for TOC, OI_{RE6}, HI and $T90_{HC\_PYR}$ are plotted on the y-axes. On plots a, b and c and on the left side of the twin plot the x-axis represents values calculated on the acquisition parameters-based method (section 3.1.). On the right side of the twin plot the x-axis shows values obtained with the reconstructed thermograms method (section 3.2.).

Overall a strong agreement was also observed between measured temperature parameters and calculated with the RT method (Supplementary Table 2). Nonetheless, the RT calculation method resulted in a less strong correlation and slightly biased relationship for $T90_{HC\_PYR}$ (R^2_{RT} = 0.78, $BIAS_{RT}$ = −3.6 °C, NRMSE = 14%; Fig. 31e). $T90_{HC\_PYR}$ was underestimated by the RT calculation method for the same two plots, for which the HI overestimation was observed and the same pattern, with the prediction bias increase towards the 50:50 mixtures was evident.

Finally, all three ways of calculating the expected stable SOC proportion of soil mixtures using the $PARTY_{SOC}v2.0_{EU}$ model performed well. Values of stable SOC proportion calculated as a weighted average of $PARTY_{SOC}v2.0_{EU}$ predictions obtained for layer A and layer B were relatively close to the actual stable SOC proportion predictions obtained using the measured Rock-Eval® parameters as predictors in the $PARTY_{SOC}v2.0_{EU}$ model (R^2 = 0.85, BIAS = 0.03, NRMSE = 13%; Supplementary Table 2). The relationship between theoretical and observed stable SOC proportions improved when Rock-Eval® parameters calculated with the AP method or with the RT method were used as predictors in the $PARTY_{SOC}v2.0_{EU}$ model (R^2_{AP} = 0.94, $BIAS_{AP}$ = −0.01, $NRMSE_{AP}$ = 7%, R^2_{RT} = 0.94, $BIAS_{RT}$ = −0.02, $NRMSE_{RT}$ = 9%). An overview of summary statistics of Rock-Eval® parameter values measured on “natural” samples is presented in Supplementary Table 3.

4.2. Clay content difference between sublayers explains the poorer agreement between calculated and measured HC_PYR thermogram-derived parameters

Our results suggested a strong agreement between calculated and measured values for all Rock-Eval® parameters of soil mixtures except those derived from the HC_PYR thermogram. Calculated values of S2, HI (obtained with the AP calculation approach) and temperature parameters of the HC_PYR thermogram (obtained with the RT calculation approach) were either positively or negatively biased, leading to a slight overestimation of S2 and HI, and an underestimation of HC_PYR temperature parameters respectively (Fig. 31c and e; Supplementary Fig. 2; Supplementary Table 2).

The overestimation bias observed for S2 was magnified for HI, which can be explained by the mathematical definition of the latter parameter (HI = S2/TOC × 100). We investigated the relationship of the overestimation of S2 and HI with various soil characteristics of each plot presented in Table 2. The mean difference between measured and calculated S2 and HI values was averaged per plot and it was compared to the difference in various soil characteristics between layer A and layer B (Table 2). This comparison revealed a strong correlation between overestimation of S2 and HI and difference in clay content between soil layers (Spearman rho = 0.90 and R^2 = 0.71 for S2, and Spearman rho = 0.85 and R^2 = 0.84 for HI; Fig. 32a and b). The bias was the highest for the two plots of the Parroy site, where the clay content difference between layer A and B was the highest (clay content difference > 20 wt.%). Investigation of the relationship of this bias with other soil characteristics, such as soil pH, and TOC content did not explain any of the observed variation. The clay content difference between the two layers was also negatively correlated with the bias observed for $T90_{HC\_PYR}$ (not shown here).

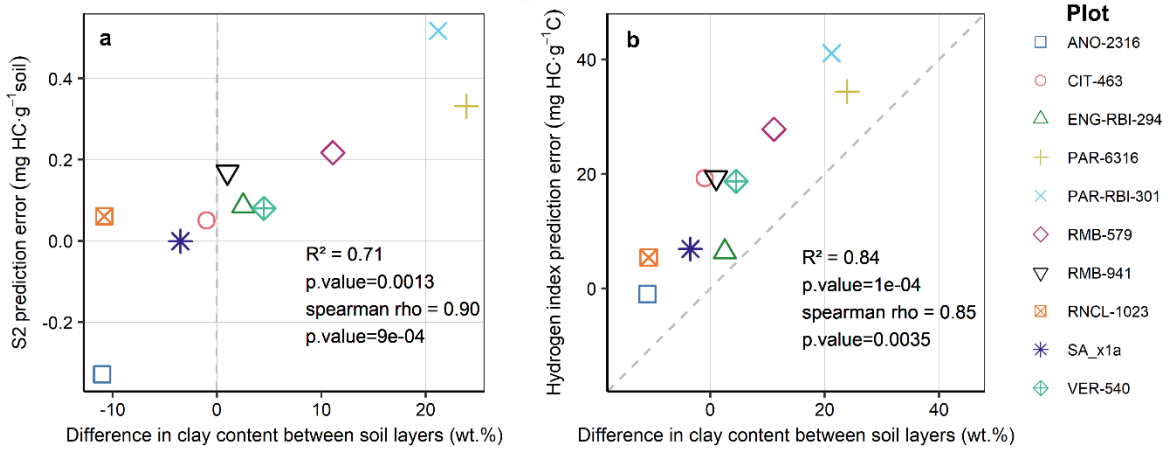


Figure 32: Relationship between prediction bias in Rock-Eval® parameters (S2 and HI) and difference in clay content between soil layers (i.e., clay content of the surface layer – clay content of the subsurface layer) for each plot. Prediction error shown on the y-axis is calculated as the mean difference between calculated Rock-Eval® parameter values (obtained with the acquisition parameters-based method) and measured values per plot.

5. Discussion

5.1. Recommendations for inferring Rock-Eval® parameters of soil mixtures

According to the type of parameter, two calculation methods are discussed for combining Rock-Eval® analysis results of different sublayers to obtain the Rock-Eval® signature of a soil layer. The first method based on acquisition parameters (AP method; see section 3.1.) is fast and easy to implement, and perfectly adequate, when classic Rock-Eval® parameters (such as PC, TOC, OI_{RE6}, MinC, etc.) or results of the $PARTY_{SOC}v2.0_{EU}$ model are of interest. Even though this method is simple and effective, attention must be paid when applying weighting on parameters representing ratios (e.g., HI, OI_{RE6}; Perdue & Koprivnjak, 2007) or more complex terms such as logarithms (I-index). One drawback of this method is related to its use to predict temperature parameters. Although empirically in this work the AP method performed well for the calculation of temperature parameters, from a purely mathematical point of view it cannot be recommended. The accuracy of this method for the calculation of temperature parameters specifically is very sensitive to mineral-induced changes on the Rock-Eval® signal. Temperature parameters do not represent surface areas (like S1, S2, etc.) but instead they mark a limit on the x-axis of a thermogram (time axis), defining the moment (translated into the corresponding temperature) when a specific proportion of a thermogram surface area was reached. Therefore, if an interaction causes a change in the shape of a thermogram, this will not be reflected in the AP based calculation method.

A second approach is presented (Section 3.2.) for the calculation of temperature parameters, based on signal processing. According to this method, the expected Rock-Eval® thermograms of a soil layer can be calculated based on the thermograms obtained on its sublayers (RT method; Supplementary Fig. 2). Rock-Eval® temperature parameters representative of a soil layer can subsequently be calculated on the reconstructed thermograms. This method might be considered as more demanding in terms of manipulation but it is also more suitable for the calculation of temperature parameters. In soils where a mineral matrix effect might be expected and in cases where precise information on the hydrocarbon signal is important, the RT method is more appropriate. However, in cases where the mineral matrix effect is expected to be low, or if the hydrocarbon signal is not particularly of interest, this approach might be unnecessarily cumbersome since its performance is expected to be equally good as the AP method.

For the variety of soil types and the range of sampling depths studied here, the most appropriate way to combine Rock-Eval® results of sublayers to obtain the stable SOC proportion of a soil layer, is to follow the AP method to calculate values of Rock-Eval® parameters to be used as predictors by the $PARTY_{SOC}v2.0_{EU}$ model.

Regardless of the chosen calculation method, one additional aspect to consider when adding information obtained on sublayers, is the content of fine fraction in each sublayer. This study focuses on linear mixing regarding the concentration of SOC and other parameters measured on a given soil mass. However, when averaging existing data, a correction will have to be undertaken to account for the soil mass contribution of each soil layer to the final sample. Thus, information on the fine fraction content, and the density difference between layers when calculating stocks of elements, will be required.

5.2. Results should be considered with caution when mixing samples with contrasted clay contents

Our initial hypothesis, that classic Rock-Eval® parameters will be independent of the sampling strategy for temperate forest soils was confirmed, with one exception. Introducing a clay rich soil layer to a coarser soil layer through the mixing process caused changes in the hydrocarbon signal (Supplementary Fig. 2), and especially on the S2, HI and $T90_{HC\_PYR}$ parameters. This effect was most pronounced for the two plots from Parroy, where the difference in clay content varied by more than 20% between layer A and layer B. After their release in the pyrolysis oven, hydrocarbon gases might be protected through retention by clay minerals (Espitalié et al., 1980). The observed increase in $T90_{HC\_PYR}$ is an indication of a bulk increase in hydrocarbons thermal stability that could be explained by retention of light hydrocarbons on mineral surfaces causing them to evaporate at a higher temperature (Rahman et al., 2017). The decrease in S2 and HI however would require a sink in the total fraction of carbon released as HC, with a shift of the lost amount of carbon to the CO or CO_2 signal, since TOC remains constant. However, due to the relatively low fraction of carbon released as HC compared to CO_2 (in average ten times less) and the very low S2 overestimation bias, the difference is not detectable in any of the other signals. Previous studies involving Rock-Eval® analysis of soils before and after removal of the mineral matrix through acid hydrolysis support the hypothesis that clay particles may

influence the detection of pyrolysis effluents (Zegouagh et al., 2004; Czirbus et al., 2016). Both studies showed that hydrocarbon yield increased in the absence of minerals, causing up to an 18-times increase in S2 (Zegouagh et al., 2004). Although qualitatively interesting, quantitative estimation of the mineral effect is problematic when the acid hydrolysis method is applied. While efficient for elimination of minerals, the acid attack also causes some loss of organic matter (e.g., 17% in Zegouagh et al. (2004)). Studies have shown that the occurring SOM loss is preferential, influencing the bulk composition of SOM (Rumpel et al., 2006; Spaccini et al., 2013). Future experimental work should focus on quantifying the effect of different minerals on the Rock-Eval® signal of SOM, using simple model systems. To better understand the interactions between SOM and minerals it would be of interest to examine how different minerals and different degrees of interaction might influence the detection of SOM.

For temperate forest soils, we can consider the difference in clay content (20%) observed in this study, as a first empirical threshold value, up to which the mineral matrix effect is insignificant. As the application range of the Rock-Eval® method extends to new climates and new soil types, it is important to know that the linear mixing assumption holds true in temperate forest soils. However, different guidelines might be necessary in tropical soils where, as shown by previous studies conducted on kerogens, the effect of more reactive clay minerals and iron oxides during pyrolysis of organic matter is expected to be stronger (Huizinga et al., 1987; Ma et al., 2018).

6. Conclusion

Our study shows that the Rock-Eval® parameters of a soil layer, including the centennially stable SOC proportion predicted by the $PARTY_{SOC}v2.0_{EU}$ model, can be obtained on Rock-Eval® results of its composing sublayers. This observation has significant practical implications regarding the comparison of available Rock-Eval® soil data from studies applying different sampling depths. According to the parameter type, one of the two methods of calculating expected values for the Rock-Eval® parameters of composite samples can be recommended: (1) as the weighed mean of acquired Rock-Eval® parameters for classic and some extended parameters (i.e., S1, S2, S3, S3', S3CO, S3'CO, PC, TOC, MinC, HI, OI_{RE6}, Pseudo S1, S2/PC, PC/TOC, HI/OI_{RE6}, I-index, R-index, TLCH-index, centennially stable SOC proportion), and (2) based on a signal reconstructed using Rock-Eval® thermograms measured on their composing endmembers for temperature parameters (i.e., $T70_{HC\_PYR}$, $T90_{HC\_PYR}$, $T30_{CO2\_PYR}$, $T50_{CO2\_PYR}$, $T70_{CO2\_PYR}$, $T90_{CO2\_PYR}$, $T70_{CO\_OX}$, $T50_{CO2\_OX}$, $T70_{CO2\_OX}$, $T90_{CO2\_OX}$). In the context of the ongoing work towards harmonization of databases it is important to keep in mind that in temperate soils, the linear mixing hypothesis is valid. However, in soils with pronounced clay content heterogeneities between sublayers (clay content difference > 20 wt.%) averaging HC_PYR-derived parameters over a soil profile might lead to biases. Although the reconstructed thermogram method is more demanding, it is more appropriate for the calculation of temperature parameters and it can offer valuable insight on mineral-induced changes in the Rock-Eval® signal. This work can be considered as an empirical guide, paving the way towards harmonized Rock-Eval® databases for soil studies.

Acknowledgments

The authors would like to thank the French ANR (project StoreSoilC; ANR-17-CE32-0005; PhD thesis of Eva Kanari), ADEME (Agence de la transition écologique; piCaSo project and PhD thesis of Eva Kanari) and the French Ministère de la transition écologique et solidaire (DEB; GNB-sol project) for the financial support during this work. We would also like to thank Jean-Maxime Gonzalez, Sébastien De Danieli, Thomas Goïtré, Nicolas Eisner, Audrey Bonnelye, Yoan Paillet, Charlène Heiniger and Isabelle Bilger for their help in sampling and processing soils. We deeply thank Julien Barré for his help in mathematical analysis. Finally, the authors thank the anonymous reviewers and the editor for their valuable time and input.

CHAPTER 3

Opening the back box: Why is thermal stability a suitable proxy for biogeochemical persistence of soil organic matter?

Earlier in this manuscript, both the complex nature of SOM with the associated difficulty in estimating its persistence and the potential of Rock-Eval® thermal analysis for a use as a proxy were thoroughly discussed. Although the value of this technique for estimating persistence and advancing SOC modelling is so far undisputable, our mechanistic understanding of the relationship between thermal and biogeochemical stability at the centre of this approach is still very limited.

In this chapter, based on a simple experimental setup and a model system, we focused on the effect of organic matter adsorption on minerals as a potential protection mechanism during thermal degradation of SOM. We analysed pure biochemical compounds from multiple groups (lignin, humic acid, proteins, carbohydrates, lipids) to obtain their Rock-Eval® signal. We prepared organo-mineral mixtures following a simple batch sorption protocol to study the resulting changes on thermal stability and Rock-Eval® parameters. As an intermediate step, we evaluated the effect of the presence of minerals in dry simple mixtures. We used pure artificial biochemical compounds and minerals as well as organic matter and mineral matrices encountered in natural soils.

We hypothesized that adsorption degree and strength would be dependent on organic compound functional groups as well as pH, mineral specific surface area and reactivity. We expected increasing association between organic matter and minerals to have a stronger effect on Rock-Eval® parameters (previously empirically correlated to SOC biogeochemical stability). We also anticipated interactions between reactive minerals and pyrolysis effluents resulting in a strong effect of the presence of minerals on the Rock-Eval® signal even in simple dry mixtures.

The choice of working with pure compounds for the sake of experimental model simplicity, revealed an important drawback of the Rock-Eval® method when applied to certain compounds, namely a deficiency in carbon yield of oxygenated pure compounds.

This issue is currently neglected in literature as the technique is mostly used for natural soil and sediment samples containing organic matter, whose Rock-Eval® analyses present good (90-100%) carbon yields. Furthermore, carbon detection efficiency was influenced by the presence of goethite in carbon poor mixtures (TOC≤1% wt.). We argue that this deficiency is due to the functioning principle of the FID, and we present evidence obtained using an experimental configuration of Rock-Eval® supporting that the missing carbon yield is associated to the pyrolysis step. For compounds such as organic polymers and particulate organic matter the carbon yield of Rock-Eval® was very good (carbon yield>95%).

Batch sorption experiments were successful for one of the proteins, bovine serum albumin, and all minerals (kaolinite, montmorillonite, goethite, and three natural soil matrices). We discuss possible explanations for this outcome, and for the fact that for the other two compounds used here (humic acid and cysteine) the adsorption efficiency was very low.

Dry mixtures of bovine serum albumin with minerals showed that some pure minerals can have a strong effect on the Rock-Eval® signal. We observed a strong oxidation of pyrolysis effluents by goethite, and a retention of effluents by montmorillonite. Although much less intense, some retention and oxidation of effluents was observed as well for natural soils. Amongst the different thermograms, the hydrocarbon signal was the most sensitive one to the presence of minerals. An important source of uncertainty questioning the origin of these observations is the variability in carbon content across mixtures.

The effect of adsorbed mixtures on the Rock-Eval® parameters was very similar to the one of corresponding dry mixtures. The cumulative effect of the changes caused on the Rock-Eval® signal led to slightly higher values of centennially stable SOC predicted by $PARTY_{SOC}$ for adsorbed mixtures. However, dry mixtures had a significant impact on predictions provided by $PARTY_{SOC}$ as well. Compared to the pure compound detection the difference caused by adsorption was much smaller than the difference caused by the presence of minerals. We attribute this lacking importance of adsorption to the specific choice of compounds used in this model system and we hypothesize that successful adsorption of smaller organic molecules might lead to a more efficient protection against thermal degradation.

An immediate research perspective is to homogenize as much as possible the ratio of organic compound to available mineral surface mixtures with a constant carbon content. Additional steps that could improve the current experimental protocol by providing more detailed information during the different stages of the experiment include: optimization of adsorption conditions by pH regulation, extended elemental analysis (C, H, N, S, O) of the mixtures before and after adsorption, and analysis of organic compounds in rinsing water to ensure that there is no preferential removal of some parts of the macromolecule during rinsing.

Future work should focus on understanding these associations at smaller scales, for example by studying the distribution of organic matter in adsorbed and dry mixtures using scanning or transmission electron microscopy. Similar composed mixtures could be analysed by differential scanning calorimetry to obtain information on the difference in energy gain when organic matter is adsorbed. Further research could also focus on the impact of other mechanisms potentially influencing thermal stability, such as chemical recalcitrance.

Understanding the influence of organo-mineral interactions on organic matter thermal stability expressed by Rock-Eval® parameters

The work presented in this chapter was conducted in collaboration with François Baudin, Pierre Barré, Tracy Boucher, Nicolas Bouton, Laure Soucémarianadin, Florence Savignac and Lauric Cécillon

Abstract

Thermal analysis has recently gained popularity as a quick and simple way to assess biogeochemical stability of soil organic matter (SOM). However, an important question remains: How can we explain the correlation between the processes of thermal and biogeochemical degradation of SOM? The objective of this work was to improve our current understanding of this relationship using a simple model system focusing on the influence of organo-mineral interactions.

First, pure biochemical compounds including lipids, proteins, carbohydrates and organic polymers were analysed by Rock-Eval® in order to register the Rock-Eval® signal of each of these substances. Then, we composed organo-mineral complexes by a batch sorption experiment using a selection of minerals (kaolinite, montmorillonite, goethite, and three natural soil matrices) and organic compounds (cysteine, bovine serum albumin and humic acid). Successfully adsorbed compounds and their respective simple dry mixtures were analysed by Rock-Eval® and the effect of organo-mineral associations as well as the pure mineral matrix effect were evaluated as the observed changes in Rock-Eval® parameters.

The main results of this chapter include: (i) an evaluation of the detection efficiency of pure biochemical compounds by Rock-Eval® that strongly varied according to the type of organic compound, with the highest yield for organic polymers (e.g., lignin) and the lowest yield for

lipids (e.g., palmitic acid), (ii) a first estimation of the effect of the presence of minerals during analysis of organic matter, with reactive minerals such as goethite and montmorillonite having a strong oxidizing and protecting effect respectively during the pyrolysis step, while natural soil matrices showed weaker interference, and (iii) an estimation of the influence of pre-existing adsorption between one organic compound (bovine serum albumin) and various minerals, namely causing a slight to negligible increase on the overall thermal stability assessment. The insights presented in this chapter are a step towards understanding the influence of different mechanisms on thermal stability and very importantly they reveal limitations and strengths of the Rock-Eval® technique.

1. Introduction

The importance of SOM for ecosystem and human well-being has been discussed multiple times throughout this manuscript (General introduction Sect. 1.3.). We also discussed the wide encountered spectrum of SOM persistence, a critical quality for understanding and predicting the evolution of this reservoir in the future which has proven to be particularly complex to estimate (General introduction Sect. 1.8. and 1.11.). Efforts to understand the processes behind this wide spectrum of SOM persistence have come to a consensus that interactions with plants, microorganisms and minerals are what is causing SOM ―that should decay fast in a soil environment according to thermodynamics― to persist in soils for millennia (Trumbore, 1997; Lehman and Kleber, 2015). Influenced by many biotic and abiotic factors, SOM persistence is best described as an ecosystem property (Schmidt et al., 2011; Kleber et al., 2015). Sorption onto mineral surfaces leading to chemical, biological or physical protection is accepted as a crucial mechanism for SOM stabilization (Six et al., 2002; Eusterhues et al., 2003; Kaiser and Guggenberger, 2003, Mikutta et al., 2006), whereas the role of soil mineralogy is emphasized (Doetterl et al., 2015).

To date, efforts to understand the influence of organo-mineral associations on SOM stability either focus on the overall effect of conceptual interactions, leading to conceptual fractionation protocols (e.g., Lugato et al., 2021) that however do not achieve homogeneous separation of stable SOM (Soucémarianadin et al., 2019; Chassé et al., 2021), or they are focusing on understanding organo-mineral interactions at the microscale. Batch sorption and desorption

experiments for example offer indirect interpretations of dominant organo-mineral associations (Mikutta et al., 2007; Wagai and Mayer, 2007; Hayakawa et al., 2018). Molecular level imaging and mapping techniques offer impressive insight into SOM distribution at the nanoscale (e.g., NanoSIMS; Vogel et al., 2014). Dynamic force spectrometry (Newcomb et al., 2017) presents a great opportunity at understanding the importance of the energetic balance at the molecular level. However, these techniques are still far from being able to provide quantitative information and they are particularly problematic regarding upscaling.

Earlier in this manuscript we discussed the potential of Rock-Eval® to characterize persistent SOM (General introduction Sect. 1.12.; Gregorich et al., 2015; Barré et al., 2016; Soucémarianadin et al., 2018; Poeplau et al., 2019) and providing parameters that are used as input variables in a machine learning model ($PARTY_{SOC}$) offering quantitative predictions of centennially persistent SOC (Cécillon et al., 2021; in ANNEX 1). In chapter 1, we discussed the independent validation of the $PARTY_{SOC}$ approach together with its usefulness for improving the accuracy of predictions of SOM dynamics. As this technique, initially developed for use on kerogens, is becoming increasingly applied on soils, a fundamental question becomes more pressing, namely to gain a better mechanistic understanding of the observed link between thermal and biogeochemical degradation of SOC.

The work presented in this chapter involves thermal analysis and a batch sorption experiment combined in an effort to decipher the importance of organo-mineral interactions during Rock-Eval® analysis. Specifically, we aim at studying the effect of pre-existing association as a parallelism to sorption of organic matter in soils and as a potential protection mechanism that may or may not be reflected on the Rock-Eval® signal.

We developed a simple protocol involving pure organic and mineral compounds of known composition as well as natural soil organic matter and soil minerals. Although we worked with a simple model system, we expected multiple simultaneous interactions to have a complex effect on the detection and the thermal stability of analysed organic matter.

We expected the adsorption efficiency of the different mineral matrices to increase with increasing specific surface area and cation exchange capacity. Also, we anticipated that reactive pure minerals would cause a change on Rock-Eval® parameters through interaction with pyrolysis effluents. We hypothesized that the hydrocarbon signal would be particularly sensitive to the presence of minerals and that pre-existing association (i.e., adsorption) between organic compounds and minerals would be reflected on the obtained signal. Finally, we expected thermal stability to increase with increasing degree of association.

To test these hypotheses, multiple steps were necessary, given the very limited use of Rock-Eval® for analysis of pure biochemical compounds (Carrie et al., 2012) and the extensive literature on evidence for strong mineral matrix effect during thermal analyses (Espitalié et al., 1980; Davis and Stanley, 1982; Espitalié et al., 1984; Alcañiz et al., 1989; Czirbus et al., 2016; Bu et al., 2017; Ma et al., 2018; Pan et al., 2010). To overcome these complications, we first evaluated the detection efficiency of Rock-Eval® for pure compounds. Second, we produced adsorbed mixtures of organics and minerals and third we composed and analysed simple dry mixtures of the same compounds to quantify the effect of pre-existing associations independently of the "dry" mineral matrix effect.

In this chapter the terms "association" and "interaction" are used repeatedly, thus it is important to underline that they are not synonymous. First, interaction is used as a more general term that describes any reaction taking place between organic matter or organic effluents and minerals, including the existence of bonds, or the spontaneous retention or oxidation of effluents by minerals. Association on the other hand is more specific and indicates precisely the existence of a bond (presumably adsorption) between organic compounds and minerals, generated by the batch sorption experiment. To be considered as "associated", the organic compound should have not been leached out after several rinsing steps using deionized water.

2. Materials and methods

2.1. Materials

2.1.1. Organic compounds

We selected 10 pure biochemical compounds, including carbohydrates (glucose monohydrate, starch and cellulose), lipids (cholesterol and palmitic acid), proteins (cysteine, reduced glutathione and bovine serum albumin), lignin and humic acid, which were all purchased from Sigma-Aldrich® in dry powder form. In addition, we used particulate organic matter (POM), extracted from a surface horizon (0–5cm) of a grassland soil collected next to the "Les Closeaux" Field Experiment in Versailles, France (Fernández-Ugalde et al., 2013; Table 3). Following a simple protocol, 60 g of soil were immersed in milli-Q water and shaken overnight in a container with five 1cm-diameter glass beads to break apart aggregates. The organic matter fraction not associated to minerals was separated through wet sieving using a 200 µm sieve. The obtained material was dried at 40 °C and ground to obtain a homogeneous powder.

2.1.2. Mineral compounds

Our mineral selection included four pure minerals: kaolinite, montmorillonite, goethite and sand, and three naturally occurring SOC-poor soils (TOC<0.2% wt.). The low SOC content was sought so that the Rock-Eval® signal due to this organic matter can be considered as negligible compared to that of the added compounds. All pure minerals were purchased by Sigma-Aldrich® except sand, which was collected from the Fontainebleau quarry in France (~98% wt. SiO_2, TOC<0.2% wt.). The three natural soil samples (soil 1: HET 03, soil 2: CHS 27, and soil 3: CHS 72) collected from deep soil layers (0.8–1.0m) of temperate forested sites (RENECOFOR network, the French part of the European ICP Forests Level 2 network; Nicolas et al., 2018) were selected according to their texture and clay mineralogy. Soil 1 had a lower clay content (11% wt.) compared to the other two (~30% wt.) and contained a large fraction of sand (60% wt.). According to XRD analysis of extracted clays from the three soil matrices, the samples shared a similar mineralogy, with mainly kaolinite, chlorite and illite. Soil 2 and soil 3 showed some illite-smectite interstratification as well. The clay minerals in soil 1 had a particularly well defined crystal form. In soil 3 the influence of kaolinite was more important than in the other two soils (Table 3; Fig. S1).

Table 3: Chemical and textural (where applicable) characteristics of the organic compounds and minerals used in this study.

Organic compounds				Minerals	
Name	**Formula**	**Theoretical C content (% wt.)**	**Solubility in H_2O, 25°C (mg/ml)**	**Name**	**Formula/Composition Soil type (WRB, 2014)**
Bovine serum albumin (BSA)	—	51.1 †	≥40 ‡	**Sand**	SiO_2
L-Glutathione reduced	$C_{10}H_{17}N_3O_6S$	39.1 ∗	≥1000 §	**Kaolinite**	$Al_2Si_2O_5(OH)_4$
L-Cysteine	$C_3H_7NO_2S$	29.8 ∗	≥1000 §	**Montmorillonite**	$(Na,Ca)_{0.3}(Al,Mg)_2Si_4O_{10}(OH)_2 \cdot n H_2O$
Cellulose	$(C_6H_{10}O_5)n$	44.4 ∗	none ‖	**Goethite**	FeOOH
Starch (wheat)	$(C_6H_{10}O_5)n$	44.4 ∗	none ‡	**Soil 1 (HET 03)**	Clay/Silt/Sand: 11/29/60 (% wt.) Dystric Cambisol
D(+)-Glucose Monohydrate	$C_6H_{12}O_6 \cdot H_2O$	36.4 ∗	1000 ‡		
Cholesterol	$C_{27}H_{46}O$	83.9 ∗	<0.1 §	**Soil 2 (CHS 27)**	Clay/Silt/Sand: 29/63/8 (% wt.) Luvisol
Palmitic acid	$C_{16}H_{32}O_2$	75.0 ∗	<0.1 §		
Humic acid	—	42.4 †	none ¶	**Soil 3 (CHS 72)**	Clay/Silt/Sand: 35/33/32 (% wt.) Luvisol
Lignin	—	47.9 †	<0.1 #		
Particulate organic matter (POM)	—	13.9 †	none (by definition)		

∗ According to the chemical formula, † According to elemental analysis, ‡ Product safety data sheet [Sigma, 2021; SigmaAldrich, 2021a, b]
§ Haynes et al., 2017, ‖ Alves et al., 2016, ¶ Brigante et al., 2007, # Thakur et al., 2014

2.2. Sample preparation

2.2.1. Adsorption experiment

For the adsorption experiment we used a limited selection of organic compounds that were previously used to form organo-mineral complexes: cysteine (Faghihian and Nejati-Yazdinejad, 2009; Vieira et al., 2011; Hu et al., 2020), bovine serum albumin (Chevallier et al., 2003; Phan et al., 2015) and humic acid (Feng et al., 2005; Chen et al., 2017). Glucose was also used due to its high solubility, even though it is not expected to sorb on minerals. First, a stock solution of each organic compound ($2 g \cdot l^{-1}$) and a background electrolyte (0.01M $CaCl_2$) was prepared. Only for the stock solution of humic acid, we regulated the pH at 12, by adding 0.1M NaOH and we avoided the use of $CaCl_2$ to ensure complete dissolution (Brigante et al., 2007).

Second, 30ml of stock solution were added to 1g of each of the selected mineral compounds, resulting in a constant theoretical initial ratio of 60mg organic compound per 1g mineral, assuming a homogeneous stock solution. Consequently, the theoretical initial TOC of

composed mixtures was in the range of 1.8–3.1% wt., with cysteine on the lower and BSA on the higher end. The mixtures were shaken for 24 hours in the dark and then centrifuged for ten minutes at RCF 4300×g. The supernatant was collected and after an addition of 5µl of 85% H_3PO_4 solution (to prevent DOC microbial degradation) it was set aside for dissolved organic carbon (DOC) analysis. Five rinsing and centrifuging steps per mixture were conducted in total using a simple 0.01M $CaCl_2$ solution (or pure milli-Q water for the humic acid mixtures) and all rinsing liquids were collected and set aside. Considering the important role of pH for the formation of adsorption, a measurement of pH was taken in the first and last rinsing liquids from each mixture (Table S1). Finally, the obtained pellets were dried for 48 hours at 30 °C and ground by hand to a homogeneous powder.

2.2.2. Dry mixtures

Simple mixtures were composed using bovine serum albumin powder and all the mineral matrices used in the adsorption experiment, except sand. The mixing ratio was adapted to match the end C content of each mixture after the adsorption experiment (1–2 % wt. TOC; see below Sect. 3.2.; Fig. 32). The weighted components were ground by hand in an agate mortar to ensure a homogeneous mixing.

2.3. Analytical techniques and calculation of parameters

2.3.1. Rock-Eval® thermal analysis

Organic compounds underwent no pre-treatment before thermal analysis. Samples were prepared simply by adding 5mg of compound into the crucible, covered by approx. 50mg of pure SiO_2 sand. For all dry mixtures and adsorbed mixtures, we used a constant quantity of 60mg of sample for Rock-Eval® analysis. A Rock-Eval 6® Turbo apparatus (Vinci Technologies) was used with the same heating routine as described in Chapters 1 and 2 (see Chapter 1 Sect. 2.3.). Briefly, pyrolysis was conducted under inert (N_2) atmosphere starting at 200 °C with a three-minute isotherm followed by a temperature increase of 30 °C·min^{-1} up to 650 °C.

Oxidation temperature started at 300 °C with a one-minute isotherm and then increased by 20 °C·min^{-1} min up to 850 °C with an additional isotherm of five minutes at the end (Behar et al.,

2001; Disnar et al., 2003; Baudin et al., 2015). A detailed description of the obtained Rock-Eval® thermograms and parameters can also be found in the previous sections (e.g., General introduction Sect. 1.13.). Here we focused on classic Rock-Eval® parameters, such as total organic carbon (TOC in % wt.), pyrolyzed carbon (PC in $gC \cdot kg^{-1}$—the sum of organic C released during pyrolysis), hydrogen index (HI in $mgHC \cdot gC^{-1}$—the ratio of released hydrocarbons to TOC) and oxygen index (OI_{RE6} in $mgO_2 \cdot gC^{-1}$—the ratio of oxygen of organic origin to TOC). In addition, further parameters were calculated such as PseudoS1 (in $mgC \cdot g^{-1}$—the sum of carbon released during the first 200 seconds of pyrolysis; Khedim et al., 2021), S2/PC (unitless—the ratio of the amount of hydrocarbons released after the first 200 seconds of pyrolysis to pyrolysed carbon; Poeplau et al., 2019), the ratio PC/TOC (unitless), the ratio HI/OI_{RE6} (in $mgHC \cdot mgO_2^{-1}$) and ten temperature parameters (i.e., $T70_{HC\_PYR}$, $T90_{HC\_PYR}$, $T30_{CO2\_PYR}$, $T50_{CO2\_PYR}$, $T70_{CO2\_PYR}$, $T90_{CO2\_PYR}$, $T70_{CO\_OX}$, $T50_{CO2\_OX}$, $T70_{CO2\_OX}$, and $T90_{CO2\_OX}$ in °C) that describe evolution steps, e.g., at which temperature 30, 50, 70 and 90% of a given gas (HC, CO and CO_2) was released during which step of the analysis (pyrolysis and oxidation). These parameters were calculated for all analysed samples.

2.3.2. The machine-learning $PARTY_{SOC}$ model

We used a selection of 18 of the obtained Rock-Eval® parameters as predictors for the European version of $PARTY_{SOC}$ ($PARTY_{SOC}v2.0_{EU}$; Cécillon et al., 2021; in ANNEX 1) available on Zenodo (https://doi.org/10.5281/zenodo.4446138) to generate predictions of centennially persistent SOC proportion for pure compounds, and composed dry and adsorbed mixtures. Here, we were interested in investigating the effect of organo-mineral interactions on the predictions and the sensitivity of the model rather than the absolute values themselves.

2.3.3. Emelental analysis

The carbon content of all samples including pure compounds, and prepared adsorbed and dry mixtures was verified using a Thermo Scientific™ Flash2000™ elemental analyser, at the laboratory of ISTeP, Sorbonne University. Carbon content obtained with elemental analysis (C_{EA}) was used to evaluate the carbon yield of Rock-Eval®.

2.3.4. Carbon yield calculation

One particularity of this study regarding the calculation of the carbon yield of Rock-Eval® for the analysed organic compounds and mixtures is that all of the carbon released during Rock-Eval® analysis (TOC+MinC) was taken into account. For all comparison of Rock-Eval® and elemental analysis the carbon yield was calculated as: C-yield=(TOC+MinC)/C_{EA}.

2.3.5. Dissolved Organic Carbon analysis

The concentration of DOC in the rinsing liquids was measured by a TOC-V CSH (Shimadzu) analyser, at the laboratory of IEES, Sorbonne University. After thermal oxidation of the sample at 680°C the produced CO_2 gas was detected by gas chromatography combined with an NDIR (nondispersive infrared) detector. Results confirmed that the applied rinsing steps were sufficient to remove any carbon not in association with minerals (Fig. S2) since a plateau value close to the detection limit was reached after the first two to three rinsing steps for all mixtures.

2.3.6. Standardized relative change in Rock-Eval® parameters in the presence of minerals

Parameter values were normalized by the range recorded for each parameter to set the importance of the change independent of the parameter unit (Eq. 1). Finally, the change caused in Rock-Eval® parameter values was evaluated in comparison to the respective parameter value recorded during pure compound analysis (Eq. 2).

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

$$x_{change} = x_{normalized} - x_{pure\ compound} \quad (2)$$

where $x_{normalized}$ indicates a Rock-Eval® parameter normalized by its range, x is a recorded Rock-Eval® parameter value for the analysed mixtures, x_{min} and x_{max} are the minimum and maximum values recorded for the same parameter, x_{change} is the change in a parameter value compared to the value recorded during pure compound analysis, $x_{pure\ compound}$ (also normalized).

3. Results

3.1. Detection of organic matter by Rock-Eval® depends on the type of organic compound

The overall carbon yield of the Rock-Eval® method for the selected group of organic compounds was correlated to the carbon yield obtained by classic elemental analysis (R^2=0.52, mean carbon yield=79%; Fig. 33). The compound-specific carbon yield of Rock-Eval® varied strongly according to the type of organic compound. The detection of POM and lignin was nearly perfect (carbon yield=97% compared to elemental analysis). The carbon yield was also very satisfactory for glutathione, humic acid, bovine serum albumin, glucose and cholesterol (≥85%), and it was acceptable for cysteine (>75%). It was not satisfactory for starch and cellulose (65% and 58% respectively) and it was particularly poor for palmitic acid (37%). It is noteworthy that consideration of palmitic acid as an outlier and its removal from the sample set significantly improved the correlation between the two methods, increasing the coefficient of determination (R^2=0.89), bringing the intercept close to zero (b=0.81) and the slope close to one (a=0.85). Compounds used in the next part of the experiment all had an acceptable carbon yield, above an arbitrary threshold value of 75%.

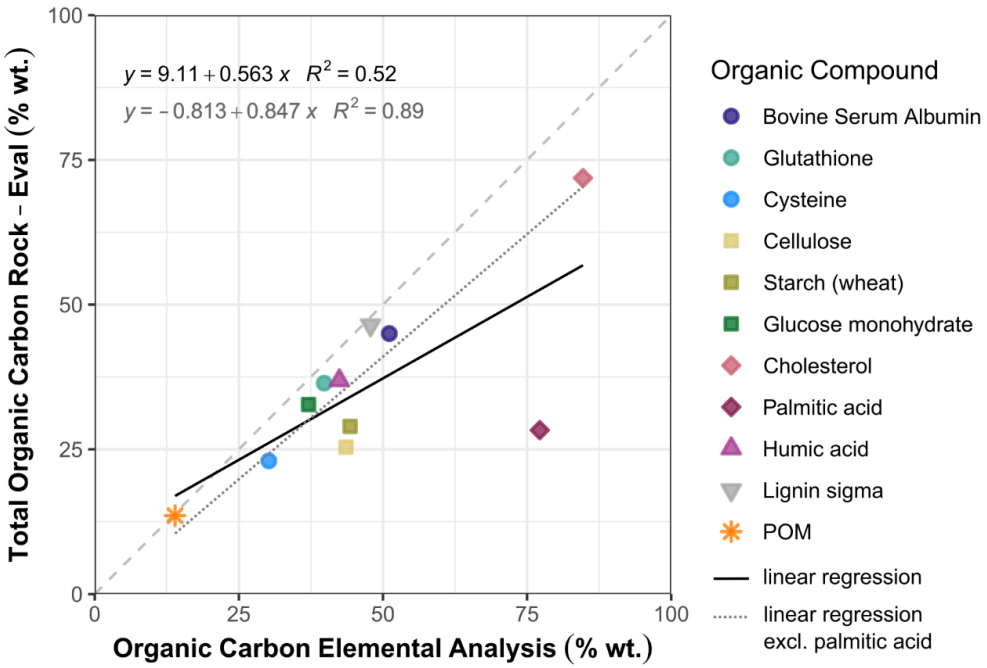


Figure 33: Performance of Rock-Eval® to detect pure compounds (TOC+MinC) compared to classic flush combustion elemental analysis. Points represent mean values based on duplicate analysis. The equations in black and in grey represent the linear regressions between x and y values, with and without palmitic acid, shown by the black and grey lines, respectively.

3.2. Efficiency of adsorption formation depends on the type of organic compound

Four compounds were used in the batch sorption experiment: glucose, cysteine, bovine serum albumin, and humic acid. Total carbon content (C_{EA}) was measured by elemental analysis for all mixtures (n=28) composed with the sorption experiment to evaluate the success of compound adsorption. Mixtures involving sand and glucose for which organo-mineral formation was not expected had indeed a negligible carbon content (C_{EA}<0.15% wt. and C_{EA}<0.25% wt., respectively; Fig. 34). These results combined with the plateau value close to zero reached in DOC concentration of rinsing solutions (Fig. S2) confirmed that our experimental design was appropriate for eliminating carbon not associated with minerals.

Out of the four organic compounds used here, only organo-mineral mixtures composed with bovine serum albumin contained a significant amount of organic carbon, in the range of 1–2% wt. (Fig. 34). The amount of carbon retained in bovine serum albumin mixtures varied in the following decreasing order: kaolinite, soil 3, montmorillonite, soil 2, soil 1 and goethite. All humic acid mixtures had a very low carbon content (C_{EA}<0.25% wt.). Carbon content of cysteine mixtures was slightly higher (C_{EA}<0.5% wt.), whereas the cysteine-goethite mixture contained 1% wt. carbon.

Evaluation of the carbon detection of these samples with Rock-Eval® compared to elemental analysis showed a strong correlation between the two methods (R^2=0.9; Fig. 34). Moreover, it revealed that goethite-containing samples had a particularly poor carbon yield (56% for bovine serum albumin and 34% for cysteine). Thus, the cysteine-goethite mixture was excluded from further consideration, along with the rest of the mixtures containing less than 1% wt. carbon, resulting in a final selection of six successfully adsorbed and adequately detected samples (i.e., bovine serum albumin mixtures with goethite, kaolinite, montmorillonite, soil 1, soil 2 and soil 3). While the bovine serum albumin-goethite mixture was retained for further consideration, related results were treated with great caution and only as a first qualitative impression because of the low Rock-Eval® carbon yield for this sample.

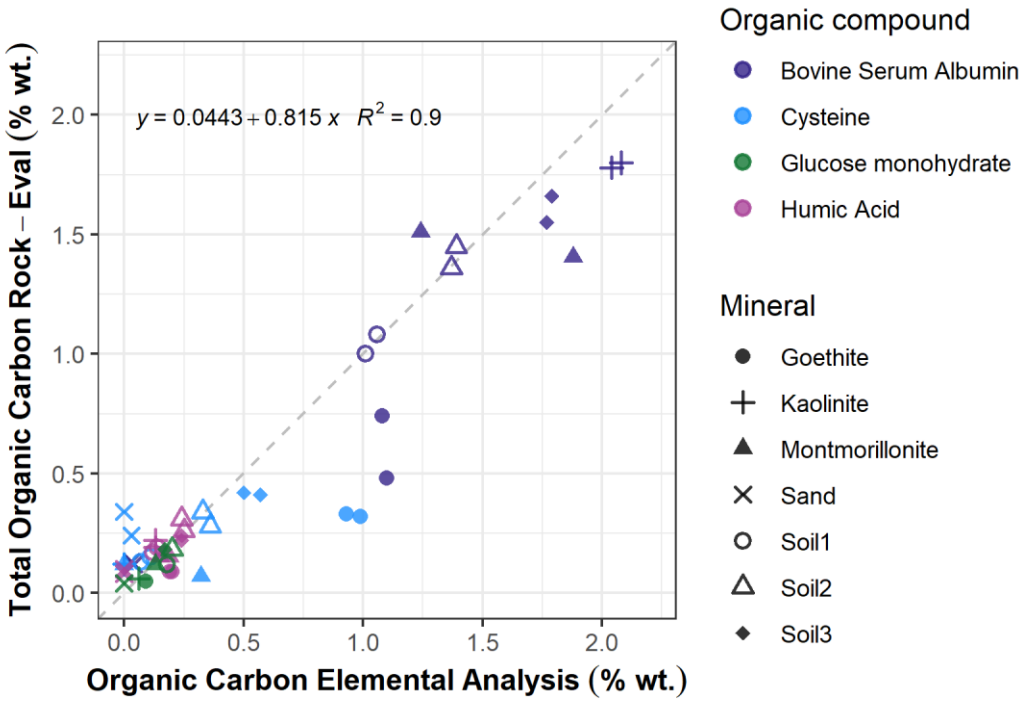


Figure 34: Evaluation of the adsorption formation efficiency in organo-mineral mixtures according to their carbon content and comparison of carbon yield obtained with elemental analysis and Rock-Eval® (TOC+MinC). Points represent individual measurements of duplicate analysis. The equation represents the linear regression between x and y values.

3.3. Effect of organo-mineral interactions on Rock-Eval® parameters

3.3.1. Satisfactory detection of bovine serum albumin by Rock-Eval® in the presence of minerals

Before investigating the effect of adsorption on Rock-Eval® parameters, we evaluated the impact that the simple presence of minerals can have on effluent detection. First, bovine serum albumin detection by Rock-Eval® in the presence of minerals was evaluated. We found that the Rock-Eval® carbon yield was close to the theoretical carbon content of the mixtures (known mixture composition) and strongly correlated with the carbon yield of elemental analysis (R^2=0.93; Fig. 35). Although Rock-Eval® tended to slightly underestimate the carbon content of the samples, the mean carbon yield of Rock-Eval® compared to elemental analysis was satisfactory (93%) and didn't show significant variation for the different minerals.

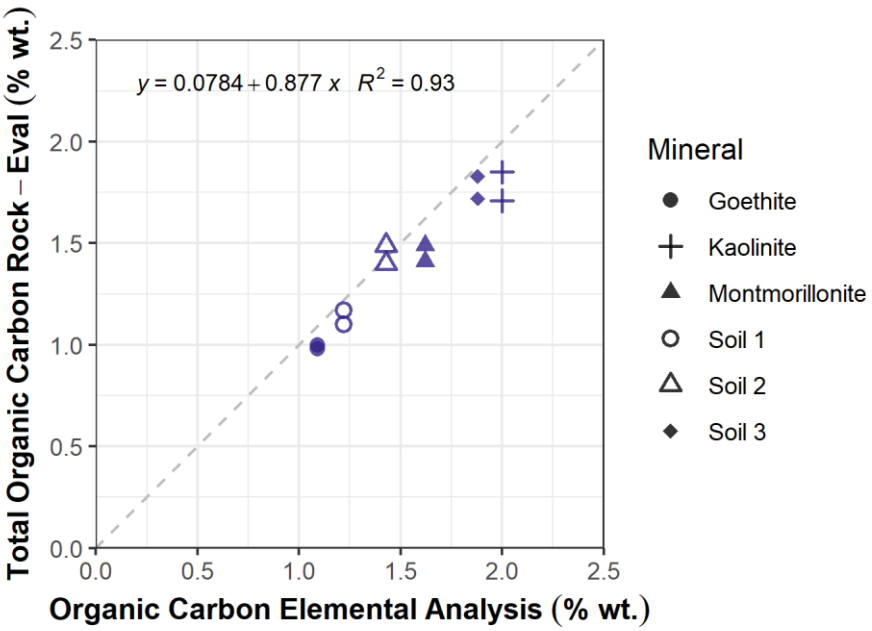


Figure 35: Performance of Rock-Eval® for detecting carbon (TOC+MinC) in simple dry mixtures compared to classic flush combustion elemental analysis. Points represent duplicate measurements. The equation represents the linear regression between x and y values.

3.3.2. Shifts in Rock-Eval® parameters in dry and adsorbed bovine serum albumin mixtures

The strongest mineral influence in dry mixtures was observed for goethite. Goethite strongly decreased the HI and increased the OI_{RE6} (Fig. 36a and b). All minerals in dry mixtures except kaolinite caused some decrease in HI, which was observed as well, slightly weaker nevertheless, for the adsorbed mixtures. In contrast, an increase in HI was observed in the kaolinite adsorbed mixture (Fig. 36a). OI_{RE6} was less affected by the presence of minerals in dry and adsorbed mixtures. The change recorded in this parameter as well was less pronounced for the adsorbed mixtures than for the dry mixtures (Fig. 36b).

All minerals in dry mixtures had an increasing effect on PseudoS1, which was stronger in adsorbed mixtures with the three soil matrices and montmorillonite, for which the increase was particularly strong (Fig. 36c).

All minerals had an increasing effect on temperature parameters of the pyrolysis step (Fig. 36d and e). The effect on the HC_PYR thermogram was stronger in simple dry mixtures than in adsorbed mixtures of the same minerals (Fig. 36d), whereas the increase in $T70_{HC\_PYR}$ was particularly strong for montmorillonite mixtures. The opposite was true for the CO2_PYR thermogram, where adsorption caused an even stronger increase in the $T70_{CO2\_PYR}$ parameter compared to dry mixtures (Fig. 36e). All minerals in dry mixtures except goethite caused an increase in temperature parameters related to the oxidation step, whereas this effect was generally weaker for adsorbed mixtures (Fig. 36f and g).

Finally, although the change caused in the value of centennially stable SOC proportion predicted by $PARTY_{SOC}$ was slightly higher for adsorbed mixtures, it was similar to that caused by the simple dry mixtures (Fig. 36h). The increase in the value of centennially stable SOC proportion observed for mixtures with the various minerals was almost exactly inversely proportional to the amount of carbon retained in each mixture (i.e., increase was higher for goethite > soil 1 > soil 2 > soil 3 > montmorillonite > kaolinite, whereas retention increased in the order goethite > soil 1 > soil 2 > montmorillonite > soil 3 > kaolinite).

Figure 36: Effect of organo-mineral interactions on Rock-Eval® parameters. Orange bars represent dry mixtures of bovine serum albumin with different minerals, whereas blue bars represent adsorbed mixtures of the same compound. Parameter values shown here present the change relative to the pure compound signal (sand) normalized by the range of each parameter (see Sect. 2.3.6).

4. Discussion

4.1. Detection of organic compounds by flame ionization detector

4.1.1. Deficiency of FID for oxygenated compounds

Evaluation of the carbon yield obtained with Rock-Eval® compared to classic elemental analysis showed particularly poor performance for some of the compounds such as cellulose, starch and palmitic acid. This deficiency was ignored in the only study known to us analysing pure compounds with Rock-Eval® (Carrie et al., 2012), which calls for revision of the results presented in that study. The carbon yield was also poor for the goethite adsorbed mixtures with cysteine and bovine serum albumin. A common attribute of these samples is the presence of oxygen, either due to the composition of the organic compounds or through the addition of an iron oxide. A known limitation of the FID method directly associated with its functioning principle, is that it is only able to detect specific moieties, namely H-C bonds. A possible explanation of the detection deficiency could be the higher presence of heteroatoms in pyrolysis effluents in samples enriched in oxygen. This is a common issue underlined in many studies focusing on the ability of FID to detect oxygenated organic compounds and leading to the consensus that a calibration for the specific compound analysed is needed, known as the response factor (Dewar, 1961; Dietz, 1967; Maduskar et al., 2015).

4.1.2. A quick look into the FID method

A closer look at the principle of the FID method is required for better understanding why an appropriate response factor is crucial. Starting with a hydrogen flame, effluents are burned to produce ions transported between two electrodes where a difference of potential is occurring (Fig. 37). The presence of the produced ions generates an electric current which is directly measured, and translated into amount of organic moieties using the compound-specific response factor (The Flame Ionization Detector, 2005). Dietz (1967) showed that although hydrocarbons are generally well detected, with response factors between 0.98–1.12, correction is much more important for non-hydrocarbons, e.g., alcohols have response factors between 0.23–0.85, and acids between 0.01–0.65.

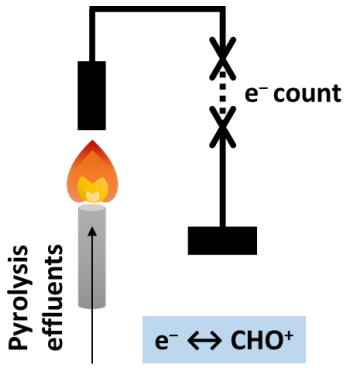


Figure 37: Functioning principle of a flame ionization detector, starting with a hydrogen flame, through which the pyrolysis effluents are burned to create CHO^+ ions, creating an electric current that is then detected and translated back into amount of organic moieties.

4.1.3. The FID is a suitable method for analysis of kerogens and SOM

For the typical use of Rock-Eval® on kerogens and even SOM the FID is an appropriate detector. Moreover, in soils the use of this method is not influenced by the performance of the FID (Disnar et al., 2003). Our results are in agreement with literature since POM analysis resulted in a perfect yield and soil matrices did not cause a problem in carbon detection. The only mineral interfering with the detection of carbon was goethite and only in adsorbed mixtures. This might be an important warning to keep in mind for the analysis of tropical soils rich in reactive iron minerals.

4.1.4. Explaining the carbon yield deficit during the pyrolysis step

The hypothesis that carbon yield deficiency comes from the pyrolysis step was tested by an experimental configuration of Rock-Eval®, called the “elemental analyser mode”. This configuration (currently under patent protection) allowed for complete detection of carbon in pyrolysis effluents and resulted in an excellent carbon yield for all compounds (Fig. S4), verifying our hypothesis that some carbon passes undetected during pyrolysis.

4.2. Insights on optimizing adsorption conditions

Our batch sorption experiment resulted in successful adsorption of bovine serum albumin on kaolinite, montmorillonite, goethite, and the three soil matrices. On the contrary, the adsorption of humic acid and cysteine was below our expectation.

The amount of detected carbon in final mixtures of bovine serum albumin ranged from 1–2% wt., corresponding to amounts of adsorbed compound between 19.7–39.1 mg bovine serum albumin per 1 g of mineral assuming the carbon content and molecular mass of this compound remained unchanged throughout the adsorption experiment. A study using a similar batch sorption protocol (replacing $CaCl_2$ with a phosphate buffer and setting pH at 7) mentioned adsorption capacity of up to 240 mg bovine serum albumin per 1 g of extracted clay fraction (Chevallier et al., 2003). However, this value might be an overestimation of the adsorption potential of bovine serum albumin on clay, due to the presence of organic matter in the initial clay fraction (13.9 gC kg^{-1}), shown to promote adsorption of proteins (Calamai et al., 2000). The ability of bovine serum albumin to form adsorption on various surfaces can be explained by its complex composition (Phan et al., 2015). In general, solution pH affects the formation of adsorption since at pH values above or below the isoelectric point of compounds (IEP) or the point of zero charge (PZC) of minerals their surface charge changes sign (Sposito, 2008), thus favouring or hindering electrostatic attraction. The bovine serum albumin molecule however, is characterized by the occurrence of both negatively charged amino acids and positively charged residues, resulting in an ability to adsorb to both positively and negatively charged surfaces (Servagent-Noinville et al., 2000; Kudelski, 2003). This composition renders the molecule highly stable, soluble in water and appropriate for sorption studies (Kudelski, 2003; Kopac et al., 2008).

For cysteine and humic acid the importance of pH is more important than for bovine serum albumin. Cysteine, as all amino acids, is distinguished by the occurrence of two functional groups an amine group ($-NH_2$) and a carboxyl group (-COOH) in the same molecule (UK: CHE 103 - Chemistry for Allied Health, 2019). At solution pH values above its isoelectric point (IEP=5.02), available OH^- ions in excess bind with the amine group of cysteine, forming NH_2 and water, and leading to dominance of the negative charge of the COO^- group, and reversely at solution pH values lower than the IEP of cysteine.

In our sorption experiment the solution pH exceeded the IEP of cysteine only in the case of the goethite mixture. For the cysteine-goethite mixture at a pH (7.6; Table S1) higher than the IEP of cysteine (5.02) and the PZC of goethite (6.4; Sposito, 2008) the surface charge of cysteine should be negative and that of goethite as well. The observed retention of cysteine in this mixture in the order of C=1% wt., corresponding to 33.11 mg g^{-1} could be explained by the presence of free Ca^{2+} (introduced by the use of background electrolyte) favouring electrostatic attraction. In the case of clays (kaolinite and montmorillonite) in our experiment the conditions were not ideal. As the naturally occurring solution pH (4.3 and 3.3, respectively; Table S1) in both cases was lower than the IEP of cysteine and lower than the PZC of the minerals (5.2 for kaolinite (Sposito, 2008) and between 3.4-5.9 for montmorillonite (Helmy et al., 1994; Ijagbemi et al., 2009). This would indicate that at these low pH values both cysteine and the mineral surfaces were always positively charged, hindering electrostatic attraction. Maximal values of cysteine adsorbed on clays reported in the literature were in the order of 43.56 mg/g on bentonite clay (Faghihian and Nejati-Yazdinejad, 2009) and 57.87 mg/g on K-Montmorillonite (Hu et al., 2020). Adsorption efficiency was very strongly dependent on pH with the maximum reached at pH=4 and maintained only for a very short pH range (3.5–4.5) in the absence of background electrolyte for the bentonite clay. This observation is in agreement with the PZC theory and with the values mentioned above. Reaching this optimal pH range can be challenging however, especially in the case of the cysteine-kaolinite pair. An alternative strategy could be to imitate the example of the cysteine-goethite pair described above, by increasing the pH in order to obtain negatively charged materials and use Ca^{2+} to form cation bridges.

Humic acid is by definition negatively charged, except at very low pH values (Sposito, 2008). In this experiment, conducted at pH~12 well above the point of zero charge of goethite (PZC=6.4) and kaolinite (PZC=5.2), both minerals were negatively charged. Thus, electrostatic repulsion could be an explanation, hindering humic acid adsorption. Efforts to conduct this experiment at lower pH (~7) resulted in incomplete dissolution (not shown here), as described in literature (humic acid dissolves at pH>10; Brigante et al., 2007). The authors also underline the decrease in solubility caused by even low concentrations of Ca^{2+} and the increase in flocculation of dissolved humic acid at Ca^{2+} concentrations of $3 \cdot 10^{-3}$M or higher. Despite these fundamental insights on the behaviour of humic acid, many studies focusing on the protection of SOM by minerals present batch sorption experiments conducted at low pH and using $CaCl_2$ as background electrolyte. Values of humic acid adsorption on clays mentioned in Feng et al.

(2005) were in the order of C=0.26 % wt. for kaolinite and C=0.43 % wt. for montmorillonite in a 0.01M $CaCl_2$ solution and at pH=4. The authors suggest that the use of $CaCl_2$ favours bridging the repulsive (negative) charges between humic acid and mineral surfaces. Chen et al. (2017) investigating the adsorption of humic acid on kaolinite and montmorillonite in the presence of 0.1 M KCl at varying pH (4–8), report values in micromole of humic acid per g mineral between 4–9 μmol/g, corresponding to C≈7–17 % wt., while maximum values were obtained at pH=4. These contradictory results suggest that it might be necessary to measure and report dissolution efficiency or alternatively, use control treatments to ensure that incomplete dissolution or flocculation is not contributing to the final concentration of humic acid.

4.3. On the effect of dry or adsorbed minerals on the Rock-Eval® signal and why it should be carefully considered

As organo-mineral complexes were successfully formed for bovine serum albumin only, this compound was further considered to study the effect of adsorption on the Rock-Eval® signal. In the effort to estimate the effect of adsorption on Rock-Eval® parameters, it is important to remember that even the simple presence of minerals in the pyrolysis oven can cause strong changes on the signal. As presented in Sect. 3.3.2. of this chapter, a similar change in a parameter such as HI might have an entirely different origin for example (i) decrease through oxidation of effluents by iron minerals, (ii) decrease through retention of effluents by reactive clays, or (iii) decrease through adsorption of the organic matter on mineral surfaces.

Careful observation of the thermogram shapes is helpful in interpreting the resulting changes on parameter values. First, the thermogram intensities revealed that most of carbon was released as HC during pyrolysis and as CO_2 during oxidation, with the two thermograms (HC_PYR and CO2_OX) sharing approximately equal amounts of carbon (Fig. S3), while the remaining three thermograms (CO2_PYR, CO_PYR, CO_OX) combined accounted for approximately ten times less the amount of carbon. For goethite, shifts between pyrolysis thermograms took place (e.g., sink in HC_PYR and increase in CO2_PYR), explaining the simultaneous decrease in HI and increase in OI_{RE6}. Also, an interesting observation was the peculiar shape caused by montmorillonite mixtures, which had a “flattening” effect on the HC_PYR signal, formed by

lowering the intensity of the S2 peak and by shifting the detection of HC towards a more advanced analysis stage compared to the rest of the minerals (Fig. S3). The opposite was observed for kaolinite mixtures, where the combined effect of more carbon present in this mixture and the presumably lower capacity of kaolinite to adsorb hydrocarbon effluents led to the formation of a particularly high and "pointed" S2 peak. Visual checking of the thermogram shapes confirmed that goethite had the tendency to shift the CO2_OX signal towards earlier release, in line with the observed decrease in CO2_OX-derived temperature parameters. However, the decrease in CO_OX temperature parameters caused by goethite might have a different origin since for both dry and adsorbed mixtures, the CO_OX signal decreased significantly to the order of the baseline level (Fig. S3).

4.3.1. The strong effect of reactive clay minerals and iron oxides during thermal analysis

The main trends in the effect of the mineral matrix were in agreement with literature. Clay minerals are known to cause retention of hydrocarbons, partially attributed to their large specific surface area (Espitalié et al., 1984; Spiro, 1991; Hayakawa et al., 2018; Ma et al., 2018; Kanari et al., 2021). Studies restricted to pyrolysis (without oxidation) of kerogens, support that retention of pyrolysis products on clays caused a lower amount of consumed organic carbon and stabilisation of organic matter (Espitalié et al., 1984; Spiro, 1991). Similarly, in our study we observed a decrease in the ratio of carbon released during pyrolysis to total organic carbon (PC/TOC; not shown here) in the presence of montmorillonite and a sustained increase in temperature parameters indicating resistance to thermal degradation. Several studies mention a catalytic effect of montmorillonite during the generation of light hydrocarbons, explained by its solid acidity (Bu et al., 2017). This is in line with the increase in PseudoS1 (due to a shift in the hydrocarbon signal; Fig. S3) observed in the adsorbed montmorillonite mixture.

Considering the implications of these results for soil studies, the fact that reactive minerals such as goethite can lead to extremely high OI_{RE6} values and even affect carbon detection, while causing higher centennially persistent SOC predictions is partially alarming.

For testing the limits of the approach and improving our understanding of the importance of the effect of minerals, estimating a threshold value of iron oxide content or carbon to iron oxide ratio above which this effect is problematic would be an interesting objective. However, in temperate soils the scenario of such a high amount of reactive oxides is quite unrealistic thus the results do not raise any red flags regarding the use of the Rock-Eval® method. Even for tropical soils the limitations caused by soil mineralogy would have to be investigated, preferably using natural soil samples.

4.3.2. The low effect of pre-existing adsorption on the Rock-Eval® signal

Our final hypothesis of an increased protection against thermal degradation with an increasing degree of association (namely in the adsorbed mixtures compared to the dry ones) could not be validated here. Based on the only successfully adsorbed compound, bovine serum albumin, the impact of adsorbed and dry mixtures on the Rock-Eval® signal and its effect on the thermal parameters was similar. According to our results, in this model system the effect on the Rock-Eval® signal is mostly due to the presence of the minerals and much less due to the existence of adsorption. Although, the combined effect on the most important parameters led to predictions of higher centennially persistent SOC proportion values for adsorbed mixtures compared to simple dry mixtures, and was thus in favour of our initial hypothesis, the difference was still very small compared to the effect of the dry minerals.

The unexpected low impact of pre-existing adsorption on the Rock-Eval® signal could be related to the large size of the bovine serum albumin molecule. Assuming that only specific sites of this large and complex protein are capable of building bonds with a given mineral surface, this would mean that although a small part of the molecule is adsorbed, and might require more energy in order to be thermally degraded, the remaining larger part of the molecule can still react and be pyrolyzed or oxidized without delay. Testing this hypothesis would require repeating this experiment and successfully forming associations between smaller organic molecules and minerals.

4.4. The role of carbon content variations across mixtures and perspectives for future work

An important source of uncertainty remains, associated with the observations presented in this chapter, namely the difference in carbon concentrations across mixtures. The different ratio of organic compound per mineral surface present in each mixture does not allow us to exclude the hypothesis that the observed changes in Rock-Eval® parameters and centennially persistent SOC proportion predictions between mixtures with different minerals are not an effect of the variation in carbon content. Even though the majority of parameters used in this study are considered relatively to the total organic carbon content of each mixture, previous studies showed that the importance of the mineral effect can increase with decreasing carbon content, especially when TOC < 3% wt. (Dahl et al., 2004; Zegouagh et al., 2004). As a result, the homogenization of carbon content across mixtures is a clear and immediate perspective that would help increase our confidence in the results presented in this study.

Other perspectives to complement the results presented in this study include conducting differential scanning calorimetry measurements on adsorbed and dry mixtures and evaluating the change in energetic balance due to the existence of organo-mineral associations. Some precise elemental analysis of all elements present in the organic compounds used (e.g., C, H, N, S, and O for cysteine) before and after adsorption could offer useful information on potential stoichiometric shifts caused during the adsorption experiment. Additionally, imaging and chemical mapping techniques such as scanning and transmission electron microscopy could provide an idea of the distribution of organic matter in the different mixtures.

Finally, the relevance of these results for the natural soil environment should not be over interpreted, due to the orders of magnitude of difference in experimental conditions such as water content, biological activity, temperature, time scale and degree of complexity.

5. Conclusions

In this study we analysed pure biochemical compounds, and adsorbed and dry model mixtures using Rock-Eval®. Evaluation of the detection efficiency of pure compounds by Rock-Eval® showed that especially for oxygenated compounds such as fatty acids, the carbon yield is not satisfactory. We recommend the application of a correction factor on the FID signal according to the type of organic compound and call for revision of former studies dealing with similar compounds and ignoring this effect. We underline that this limitation does not apply for analysis of soil organic matter or kerogens.

Our main insights from the analysis of simple dry organo-mineral mixtures were (i) the observation of a strong oxidizing effect of goethite, (ii) interfering with the FID detection and causing a carbon yield deficiency and (iii) the retention of hydrocarbons as well as (iv) cracking of light hydrocarbons by montmorillonite.

Adsorption of bovine serum albumin on mineral surfaces only slightly increased its thermal stability, showing that the produced sorption was not a particularly efficient mechanism and that its influence of the Rock-Eval® signal was negligible for this specific compound. Changes in Rock-Eval® parameters caused by the existence of adsorption between minerals and bovine serum albumin were similar to those caused by the simple presence of dry minerals in the crucible. These included decrease in hydrogen index, increase in temperature parameters (negatively and positively correlated to biogeochemical SOC stability, respectively) and an increase in the values of centennially persistent SOC proportion predicted by the machine-learning $PARTY_{SOC}$ model. This study provides a first idea about the part of thermal stability attributed to the mineral matrix effect and to organo-mineral associations. More effort should be devoted to create stronger associations and further examine their influence on thermal stability. Other mechanisms at interplay that might influence the link between thermal and biogeochemical stability such as recalcitrance and the balance between energy gain and activation energy should also be studied.

3. General discussion

In the context of food security and climate change mitigation, appropriate land management policies are crucial. Due to the ability of soils to provide numerous vital ecosystem services and the importance of soil organic carbon (SOC) for soil health, understanding the dynamics controlling the evolution of this reservoir is key. In this work we discussed the advancements regarding the understanding of processes controlling the fate of SOC, as well as the development of modelling approaches leading to the current ability of the soil scientific community to predict SOC changes. We underlined the naturally occurring significant variations in SOC persistence and we identified the lack of appropriate methods for quantifying it as a major knowledge gap as it is an important source of uncertainty for SOC projections. We tackled this challenge with the use of a thermal-analysis-based machine-learning approach, providing information on SOC characterization and SOC persistence, combined with a well-calibrated simple SOC dynamics model. Considering the clear value of thermal analysis for assessing SOC persistence, we conducted experiments to progress our understanding of the potential and limitations of this technique. Using natural soil samples and pure biochemical and mineral compounds we tried to answer questions such as the comparability of available thermal analysis data and the role of organo-mineral interactions during thermal analysis.

3.1. Using thermal analysis to improve SOC simulations

3.1.1. Thermal analysis can estimate the size of the stable-SOC pool in unknown sites

In the first chapter of this work we examined the potential of PARTY$_{SOC}$ (Cécillon et al., 2021; presented in ANNEX 1), a thermal-analysis-based machine-learning model, to estimate the centennially persistent carbon proportion of independent soil samples, meaning samples from sites not included in its training set. We used archived and recent soil samples from long-term experimentation (LTE) sites with known land management practices and a carefully monitored SOC stock evolution over several decades (median LTE duration 21 years). The observed SOC stock changes at these sites were previously simulated using AMG, a simple multi-compartmental model of SOC dynamics (Clivot et al., 2019; Levavasseur et al., 2020) considering two SOC pools: stable SOC and active SOC. The detailed monitoring information (C inputs, climatic data and soil characteristics) available at these sites offer the opportunity of conducting inverse modelling, i.e., fitting the model to the observed SOC evolution to obtain the mathematically optimal pool partitioning at the onset of the simulation period (Clivot et al., 2019).

3.1.1.1. Independent validation of the PARTY$_{SOC}$ model

To evaluate the performance of the PARTY$_{SOC}$ model on this independent dataset we compared its predictions with post optimized pool partitioning obtained with the AMG model. We showed that there was a strong correlation between the estimations of SOC stable at the scale of a century obtained by the two methods (R^2=0.63 for stable SOC expressed as proportion and R^2=0.95 for stable SOC expressed as concentration). Moreover, compared to the AMG default pool partitioning (depending only on long-term land-use history; here 0.65 at all sites as they were former croplands) the PARTY$_{SOC}$ model provided more variable predictions (0.44–0.74). These were in the same range as optimal values, and they reflected which sites were far from SOC equilibrium (e.g., Tartas, Kerbernez). Although the default pool partitioning was appropriate for sites with exceptionally uniform land-management history (e.g., Boigneville), in cases where “recent” (~decades) disturbances occurred, values predicted by the PARTY$_{SOC}$ model revealed the legacy effect of land management practices at the different sites.

More specifically, at the site of Tartas the intense application of poultry manure amendment before the onset of the LTE led to an increase of SOC in the active pool, lowering the proportion of stable SOC. Similarly, in Kerbernez the low stable SOC proportion values could be explained by the fact that this site was turned from grassland to cropland 20 years before the onset of the LTE.

3.1.1.2. The PARTY$_{SOC}$ model trained on bare fallows can be successfully used on soils with vegetation cover

A different way to evaluate the suitability of the PARTY$_{SOC}$ model for use outside its calibration set would be to compare its performance on these new independent sites to its internal validation conducted in Cécillon et al. (2021). Using a conservative error estimation by a leave-one-site-out assessment method, the authors calculated a root mean square error of predictions (RMSE) and a relative root mean square error (RRMSE; RMSE divided by mean of observations) in the order of 0.15 and 27%, respectively. In our study, both estimations resulted in lower values, namely RMSE=0.06 and RRMSE=10%. This is particularly encouraging for the use of the PARTY$_{SOC}$ model in French and eventually European croplands, since it could indicate that (i) despite the use of long-term bare fallow plots in the calibration set of PARTY$_{SOC}$ it can successfully be used on sites with a vegetation cover and (ii) that since the current training set of the PARTY$_{SOC}$ model is able to account for the pedoclimatic variability of the sites used here, the same is probably true for most of Europe (training sites from France, Germany, Sweden, Denmark and England; Cécillon et al., 2021).

3.1.1.3. Absolute values of centennially persistent SOC concentration

Other methods used to estimate the proportion of SOC that is stable at the scale of a century, such as radiocarbon ages assessed it at 35–60% of total SOC (*e.g.*, Mills et al., 2014). The natural ^{13}C abundance method applied at LTEs with vegetation change chronosequences produced a mean estimation of persistent SOC at 50% (Balesdent et al., 1987). Centennially stable SOC proportions derived from decay modelling at long-term bare fallow sites were between 30–70% (Barré et al., 2010), while inverse modelling based on detailed long-term monitoring of SOC dynamics at agronomical experiments resulted in a similar range, namely 40–75% (Clivot et al., 2019).

Expressed as concentration, centennially persistent SOC in North-western Europe ranges roughly between 5–15 gC kg^{-1}. Our results obtained based on characterization of SOC by means of thermal analysis are in agreement with the above methods since the percentage of centennially stable SOC was between 44–74%, while its concentration varied amongst sites between 2–12 gC kg^{-1}.

3.1.1.4. Inferring information on the sensitivity of the stable SOC faction using $PARTY_{SOC}$ on archived samples

The availability of physical soil samples not only from the onset of the LTEs but also from the end and in some cases intermediate dates of the simulation period, offered us the opportunity to study the evolution of the persistent SOC fraction with time. This assessment showed that $PARTY_{SOC}$ predictions of persistent SOC concentration based on samples taken at different dates of an LTE resulted in a similar value. Even though the amount of active SOC changed during the LTE (and thus the proportion of persistent SOC) due to SOC gains or losses associated to management practices, the amount of persistent SOC remained constant. Even at LTEs covering approximately the last 50 years (i.e., Boigneville) the climatic change (temperature increase since the 1970's) had no influence on the concentration of persistent SOC. This is a particularly important insight in favour of the validity of the concept of a centennially stable SOC pool, as well as regarding its limits. Another important indication of the constant persistent SOC concentration with time, is the possibility to use recent samples more easily available than archived ones to initialize the persistent SOC content at the onset of SOC simulations.

The good overall performance of $PARTY_{SOC}$ encouraged us to use this approach to initialize SOC simulations with exceptional advantages compared to traditional initialization methods.

3.1.2. Precise SOC simulations obtained with PARTY$_{SOC}$-based initialization

The second objective of Chapter 1 was to use the PARTY$_{SOC}$ predictions of centennially persistent carbon to initialize the pool partitioning of AMG (Andriulo et al., 1999; Clivot et al., 2019). The improvement brought to SOC simulations compared to the AMG default initialization was evaluated for 32 treatments from nine agricultural LTEs. The performance of the AMG model improved significantly ($RMSE_{default}$=5.95 tC ha^{-1}; $RMSE_{PARTYsoc\text{-}based}$=3.6 tC ha^{-1}) and the model was able to reproduce the observed changes in SOC stocks at the 32 treatments with high precision ($BIAS_{PARTYsoc\text{-}based}$=0.06 tC ha^{-1}). The AMG model initialized using the default and the PARTY$_{SOC}$ pool partitioning had a similar performance for sites with optimal pool partitioning close to the default value (0.65; e.g., Boigneville). An important improvement was observed with the PARTY$_{SOC}$-based initialization for sites with complex land-use or land management history (i.e., Tartas and Kerbernez as mentioned above) due to the inaccuracy of the pool partitioning by default in these cases. For these sites, the AMG model initialized using the default pool partitioning underestimated the observed decrease in SOC stocks ($BIAS_{default}$=2.33 tC ha^{-1}). Finally, the performance of the AMG model initialized with PARTY$_{SOC}$ was comparable to the optimal pool partitioning runs ($RMSE_{optimal}$=2.12 tC ha^{-1}). It is interesting to note that the error of the AMG model initialized using PARTY$_{SOC}$ was of the same order of magnitude as the measurement error of SOC stocks. Therefore, on the studied dataset, the possibilities of improving the simulations of SOC evolution are marginal to zero.

3.1.2.1. Advantages of PARTY$_{SOC}$ compared to current initialization methods

In this section we discuss one of the most important highlights of the first chapter, the independent and complete validation of the PARTY$_{SOC}$ model as an appropriate pool partitioning initialization method.

3.1.2.1.1. Escaping the unrealistic SOC equilibrium hypothesis

First, the PARTY$_{SOC}$ approach avoids the unrealistic SOC equilibrium assumption made by definition by all spin-up initialization approaches. As information on SOC persistence is difficult to obtain, modelling studies often assume that after a long enough period under a given land use has passed an equilibrium will be reached where SOC inputs will equal SOC outputs (Wutzler and Reichstein, 2007).

However, this assumption is unrealistic even at sites with decades of continuous land use (Wutzler and Reichstein, 2007; Poeplau et al., 2011; Oberholzer et al., 2014; Herbst et al., 2018; Clivot et al., 2019). Another important drawback of spin-up methods as explained in Chapter 1 is associated to the data required by this approach (such as C inputs and climatic data) covering thousands of years, which are impossible to know with precision. Evidently, $PARTY_{SOC}$ does not depend on such uncertain datasets, as it is a method exclusively built on empirical data.

3.1.2.1.2. Complex fractionation based on conceptual mechanisms *versus* automated characterization of *in situ* persistent SOC enrichment

Other studies recognizing the need for alternative initialization methods suggested a variety of physico-chemical fractionation protocols (Balesdent, 1996; Skjemstad et al., 2004; Zimmermann et al., 2007; Lavallee et al., 2020). Yet, the heterogeneous nature of SOM and the many mechanisms controlling its persistence, make the arbitrary limits set by these laboratory methods incapable of isolating homogeneous stability fractions. In essence, separated fractions (e.g., POM and MAOM; Lavallee et al., 2020) are in reality composed of different proportions of centennially labile and persistent SOC (Balesdent, 1996; Lutfalla et al., 2017; Sanderman and Grandy, 2020; Chassé et al., 2021). In contrast, the estimations provided by $PARTY_{SOC}$ have the advantage of being based on observations from sites where decade-long SOC input elimination and native SOC decay guarantee that only SOC that is biogeochemically persistent remains.

Moreover, it has been shown that even slight inconsistencies in the complex fractionation schemes as well as the varying susceptibility of different soils to the same procedure make these methods difficult to compare and transfer (Poeplau et al., 2013; Just et al., 2021). In comparison, the $PARTY_{SOC}$ model benefits from being a fully automated approach based on SOC characterization following standardized analysis steps.

3.1.2.1.3. The concept of proper complete validation of initialization approaches

Furthermore, in lack of pluri-decadal SOC monitoring field experiments SOC projections initialized according to fractionation methods are often irrationally validated against equilibrium runs instead of observed SOC evolution (e.g., Dondini et al., 2009; Leifeld et al., 2009a; Xu et al., 2011; Weihermüller et al., 2013).

Although understandable due to the enormous time, financial and human resource investment required to obtain real SOC stock time series, this remains an important issue. Without assessing the effect of an initialization method on the precision of SOC simulations, how can we evaluate its efficiency?

In the rare cases found in literature when fractionation schemes were used to initialize SOC simulations of observed SOC evolution, either (i) no significant improvement was found compared to spin-up or default initializations (Leifeld et al., 2009b; Nemo et al., 2016; Cagnarini et al., 2019), or (ii) other model parameters (i.e., decomposition rate) needed to be adjusted in order to yield more accurate simulations, failing once again to confirm the independency and transferability of the method (Skjemstad et al., 2004; Luo et al.,2014).

As we showed in this Chapter 1, not only do the estimations of pool partitioning of $PARTY_{SOC}$ correspond well to observations from fully independent sites but we also demonstrate that its use has the potential to improve the accuracy of SOC simulations. The simple coupling of the two models ($PARTY_{SOC}$ and AMG) without any changes in their internal structure or parameters provides strong evidence not only of the effectiveness but also the transferability of the $PARTY_{SOC}$ method.

3.1.2.1.4. The particular improvement of SOC simulations at sites far from SOC equilibrium

As we note above and in the discussion of Chapter 1, in our study the largest improvement to SOC simulations due to the $PARTY_{SOC}$-based initialization was observed at sites far from SOC equilibrium. The fact that past changes in land use and soil management are common over a large yet unknown surface of arable land of France and Europe (Fuchs et al., 2015; Erb et al., 2017), highlights the importance of the ability of $PARTY_{SOC}$ to account for legacy effects of site history on SOC pool partitioning. We thus expect $PARTY_{SOC}$-based initialization of SOC models to result in significantly improved simulations of SOC dynamics at a national and even continental scale.

3.1.3. The power of SOC characterization information included in PARTYsoc

The rare long-term bare fallow experiments are only one out of the three pillars on which the PARTY$_{SOC}$ model was built. The archived samples available at these sites present an extraordinary case of enrichment of an undisturbed soil in persistent SOC and they provide the opportunity to characterize this change by laboratory methods. Rock-Eval® thermal analysis includes information on thermal fractionation but also on bulk chemistry of SOM through parameters such as oxygen index and hydrogen index. This characterization method allows PARTY$_{SOC}$ to include information such as the activation energy and redox state of SOM; key elements of new understanding of SOC dynamics (Dignac et al., 2017). In agreement with literature, persistent SOC is hydrogen depleted and thermally stable (Gregorich et al., 2015; Barré et al., 2016; Hemingway et al., 2019; Poeplau et al., 2019; Chassé et al., 2021). The third strong point of PARTY$_{SOC}$ is the well-built machine-learning algorithm transforming the correlation between thermal and biogeochemical stability into quantification of the centennially stable SOC fraction. Due to the combination of these three powerful components, PARTY$_{SOC}$ is able to estimate the pool partitioning based on experimental data, keeping assumptions at a minimum.

Other more straight-forward but also important advantages of PARTY$_{SOC}$ model compared to fractionation methods are its low cost and short analysis time making it easy to apply on large scale studies. Upscaling of this approach to a global scale remains however a challenging objective.

3.1.4. Extending the use of the PARTY$_{SOC}$ model to other pedoclimates

For the geographical expansion of PARTY$_{SOC}$ to new regions with different pedoclimates, a first step was taken from the development of the first version of the model (Cécillon et al., 2018) to the second (Cécillon et al., 2021). Namely, two new European sites with bare-fallow treatments, Askov (Denmark) and Bad Lauchstädt (Germany), and a tropical site with a C3–C4 vegetation change chronosequence from Colombia were added to the model training set. Similarly, more sites from new regions of the world will be necessary in order to increase the diversity of pedoclimates covered by the training set of the PARTY$_{SOC}$ model.

Sites will need to fulfil criteria such as (i) long-enough periods of monitoring and sampling allowing for (ii) *in situ* estimations of SOC persistence (e.g., through inverse modelling, decay modelling or isotopic methods) and (iii) archived samples covering the period of the field experiment will be necessary in order to be analysed with Rock-Eval® and train the $PARTY_{SOC}$ model using the information provided by thermal analysis. Then the updated version of the $PARTY_{SOC}$ model will have to be tested and validated on independent sites from the new pedoclimatic range, similar to the work presented here, before it can be used to initialize SOC simulations. For this final step, the importance of well calibrated SOC dynamics models able to reproduce with accuracy observed SOC stock changes is of major importance.

3.1.5. The need for well-calibrated operational SOC dynamics models

In the core of the independent validation concept, as well as for the evaluation of the improvement brought to SOC simulations by an initialization approach, is the assumption that SOC dynamics models correctly account for drivers of SOC change. The work presented in Chapter 1 relies heavily on the previous extensive parametrization work of the AMG model (presented in Sect. 1.10) bringing its accuracy to the highest level for French cropland (Martin et al., 2019). The precision of the AMG model for reproducing observed changes in SOC stocks when its pool partitioning was correctly initialized was in the margin of the error of SOC stock measurements in this study (mean confidence interval=10 tC ha^{-1}; see also Schrumpf et al., 2011).

We believe that this point should not be overlooked. On the contrary we consider this an important proof of the value of simple SOC dynamics models with kinetically defined pools against more complex process-based multi-compartmental models (Manzoni and Porporato, 2009; Campbell and Paustian, 2015). Simple models are often judged as outdated because they do not explicitly represent processes controlling SOC mineralization such as sorption of organic matter onto mineral surfaces and microbial activity. Conducting simulations at the pluri-decadal scale using AMG however shows that the representation of drivers controlling SOC mineralization seems to be sufficient to accurately reproduce SOC evolution at the plot to landscape scale (Martin et al., 2019; Bruni et al., 2021). Models like AMG were developed with the main purpose to simulate observations from the plot scale in the simplest mathematical way possible and with the highest computational efficiency (e.g., Mary and Guerif, 1994). Their mathematical simplicity is a great advantage regarding their transferability.

As there are less parameters to constrain it is easier to transfer these models to new conditions and zoom out to the ecosystem and the global scale, justifying their use in Earth System Models (Todd-Brown et al., 2014).

On the other hand, mechanistic models are conceptually interesting, since they advance our thinking and ideas about processes controlling SOC dynamics (Wieder et al., 2015; Abramoff et al., 2018; Zhang et al., 2021). Theoretically, if we could parametrize them correctly, they could provide valuable insights for assessing sensitivity of specific processes important for SOM evolution, e.g., under environmental changes. They provide a different perspective to the system, regarding the time and space dimension, since they focus on rapid processes (minutes – hours) occurring at the microscale. This complexity makes it more difficult to summarize all these detailed interactions to infer the reaction of the entire system. Not only are we currently missing the knowledge to constrain these models but also mathematically integrating them into ecosystem models would be extremely demanding.

Even though each model type has its own interest, when choosing a model for use as a predictive tool, model validation should be higher up the list of priorities (Campbell and Paustian, 2015). AMG might not represent detailed microbial processes but through the incorporation of variables such as pH, C/N ratio and clay content for the adjustment of the mineralization rate, the model indirectly accounts for aspects such as protection by clay minerals or type of organic matter input. Support to the idea that pedoclimatic characteristics can explain SOC variance is provided in a recent study from Luo and Viscarra-Rossel (2020) who showed that climatic and edaphic variables could account for 70–80% of the SOC variance observed amongst 141,584 worldwide soil profiles, whereas the potential of biotic factors was only at 5–10%. For further details on the ability of the AMG model to explain the variance of SOC stocks and the importance of the different pedoclimatic variables see ANNEX 2.

Going back to the objective of expanding the $PARTY_{SOC}$ initialization method to new pedoclimates and assuming that the simple AMG model could be used in new areas, still a significant amount of data would be required to run SOC simulations. The model requires climatic data at an annual time step (temperature, precipitation), soil characteristics (pH, C/N ratio, clay and carbonate content, bulk density) as well as detailed information on land cover and management practices (crop yields, irrigation, residue return or export).

An alternative to be considered would be to use some of the more popular models found in literature such as Century and RothC, with which SOC simulations have already been conducted all over the world and data sets to run the models are already available.

3.1.6. Adapting PARTY$_{SOC}$ to other SOC dynamics models

In the case of using other SOC dynamics models, the procedure regarding the calibration and validation of the PARTY$_{SOC}$ model would be similar. However, additional internal changes in the PARTY$_{SOC}$ model would be needed to adapt the format of its predictions to the pool structure of the chosen SOC dynamics model (3–5 pools instead of 2). The increase in number of pools causes a problem of equifinality, as not enough data is available to constrain 5 pools with confidence. However, similar to this work, initializing the size of the "stable" pool of other models could be a good starting point. A potential difficulty would be estimating the size of pools with residence times of 1000 years (i.e., IOM pool of RothC or passive SOC pool of Century), when the oldest monitoring sites are less than 100 years old. Moreover, finding SOC time series against which the precision of the model for projections this long could be evaluated would be challenging.

3.1.7. Transferring PARTY$_{SOC}$ to other land-uses

The ultimate challenge, creating a national or global map of SOC persistence will require covering areas other than cropland, such as managed and unmanaged grasslands and forests. Our ability to transfer SOC models to other land uses will depend on the advancement of our knowledge and understanding of SOC stock dynamics in these cases. Yet, the precious LTEs monitoring SOC stock evolution based on which all the necessary calibration and validation steps are executed are very rare for land uses other than cropland. Models currently used to simulate SOC evolution in forest soils do not provide any reliable projections (e.g., Yasso07 model; Mao et al., 2019). However, the development and increasing availability of databases presenting SOC stock changes in forest soils over several decades (Jolivet et al., 2006; Jonard et al., 2019) offer potential prospects for testing the improvement that PARTY$_{SOC}$ could bring to the accuracy of simulations of SOC stock changes in forests.

3.1.8. Modelling SOC stocks in deeper soil layers

All the studies presented above using LTEs to understand SOC dynamics focus specifically on the ~0–30cm topsoil layer. Recent work however underlined the significance of subsoil carbon (down to 2m) representing more than 50% of SOC (Jobbágy and Jackson, 2000; Batjes, 2016) for the soil-atmosphere C exchange cycle. This SOC present in subsoil is in exchange with the atmosphere in relevant temporal scales, as according to isotopic studies recently incorporated SOC (50 years) represents in average 13% of subsoil SOC (30–200 cm; Balesdent et al., 2018). Moreover, land management such as cropping and grazing have an effect on subsoil SOC concentration (Sanderman et al., 2017). As a result, subsoil SOC is considered important for the role of soils as a potential solution or hazard to climate change. A vertical expansion of SOC models is recommended to better account for the behaviour of this reservoir (Balesdent et al., 2018; Luo and Vsicarra-Rossel, 2020).

Finally, regarding the Rock-Eval® analytical technique and its limits, some questions we have to ask does the correlation between thermal stability and biogeochemical stability still hold true in other regions, in other land-uses and in deeper soil layers? Where are the limits of the Rock-Eval® as a SOM characterization technique and when does the interference of minerals become too important?

We try to partly answer these questions in the next part of the discussion according to the experimental work conducted during this thesis. We start by discussing the potential to combine information obtained on soil sublayers to predict parameters of the 0–50 cm soil layer.

3.2. Additivity of Rock-Eval® parameters

3.2.1. Predicting Rock-Eval® parameters of a soil layer based on its sublayers

The central question of Chapter 2 regarding the comparability of Rock-Eval® parameters of soil samples from different depths is a practical one, as it could support harmonization of available data. Using samples from 10 plots, located at 8 forest sites in France, we conducted a soil mixing experiment. We discuss two methods for calculating Rock-Eval® parameters of a soil profile (0–50 cm) by combining Rock-Eval® results recorded on its sublayers (0–30 and 30–50 cm). We show that it is possible to infer the Rock-Eval® characterisation of a soil layer from sublayer characteristics and we provide guidelines regarding the most appropriate calculation procedure according to the type of Rock-Eval® parameter. We draw attention to a mineral effect hindering the prediction of hydrocarbon-related Rock-Eval® parameters in soils with pronounced clay content difference between the two soil sublayers.

3.2.1.1. Calculating acquisition and classic Rock-Eval® parameters of a soil layer

Overall, Rock-Eval® acquisition parameters such as thermogram peak areas (S1, S2, S3, etc.), and classic Rock-Eval® parameters such as total organic carbon (TOC), pyrolyzed carbon (PC), oxygen index (OI_{RE6}), and hydrogen index (HI) were linearly additive for the temperate forest samples examined here. A weighted mean calculation was the simplest and most efficient way of predicting these parameters. Although this calculation method is easy and efficient, we underline the importance of the correct mathematical solution of the weighted mean equation when it comes to parameters representing ratios (e.g., HI, OI_{RE6}; Perdue and Koprivnjak, 2007) or more complex terms such as logarithms (I-index).

Although this calculation method resulted in empirically correct values for temperature parameters as well (e.g., $T70_{HC\_PYR}$, $T90_{HC\_PYR}$, $T30_{CO2\_PYR}$, $T50_{CO2\_PYR}$) for the samples used here we discuss below why this method is mathematically wrong and we suggest an improved approach for predicting temperature parameters.

3.2.1.2. Calculating Rock-Eval® temperature parameters of a soil layer

The main difference between acquisition parameters and temperature parameters is that the latter do not represent surface areas (like S1, S2, etc.) but instead they mark a limit on the x-axis of a thermogram (time axis), defining the moment (translated into the corresponding temperature) when a specific proportion of a thermogram surface area was reached (e.g., $T70_{HC\_PYR}$ is the temperature at which 70% of the total HC_PYR was released). The issue created due to this particular definition is that the arithmetic weighted mean calculation method cannot take into account the relative contribution of each endmember to the final signal of the mixture. For the example of a mixture composed by 50% of a thermally labile carbon-rich sample (endmember A) and 50% of a thermally stable carbon-poor sample (endmember B), the weighted mean calculation will consider the $T70_{HC\_PYR}$ as a simple average of the two samples. However, in reality the signal of this mixture will be mainly controlled by endmember A since this will release more gas, forming the majority of the detected signal. The reconstructed signal calculation has the advantage of accounting for the contribution of each endmember since it takes into consideration the intensity of their respective signals. Hence we recommend calculating temperature parameters based on the reconstructed Rock-Eval® thermograms. For the rest of the parameters mentioned in the last section (acquisition parameters and classic Rock-Eval® parameters) the reconstructed thermogram method is equivalent to the simple weighted mean calculation.

3.2.1.3. Three ways of calculating persistent SOC proportion of a soil layer based on its sublayer characteristics

Naturally, we were interested in the effect of the different calculation methods on the predictions of persistent SOC proportion of the complete 0–50 cm soil layer. We investigated three different ways of obtaining $PARTY_{SOC}$ predictions. First, by calculating the persistent SOC proportion of soil mixtures as a weighted average of $PARTY_{SOC}$ predictions obtained for the two sublayers. Second, by running the $PARTY_{SOC}$ model, using calculated Rock-Eval® parameters as predictors, obtained with the simple weighted mean method. Third, by running the $PARTY_{SOC}$ model using mostly the same weighted-mean-calculated predictors but replacing temperature parameters with ones obtained with the reconstructed-thermogram method. All three ways of calculating the expected persistent SOC proportion of soil mixtures using the $PARTY_{SOC}$ model performed well.

However, the most appropriate method was the second one, based on weighted-mean-calculated predictors. The observation that persistent SOC proportion is a parameter we can predict is particularly encouraging since it indicates the possibility to incorporate existing samples (potentially collected at different depths) to the $PARTY_{SOC}$ model. Regarding the expansion of the model (currently 0–30 cm) to deeper soils layers, the absolute values of persistent SOC predictions will have to be validated first.

3.2.1.4. Interference of clays introduced through soil mixing

The simple picture described above, validating our initial hypothesis, that classic Rock-Eval® parameters will be independent of the sampling strategy for temperate forest soils was slightly different at two out of the ten investigated plots. A disagreement (prediction error) was observed between calculated and observed hydrocarbon-signal(HC_PYR)-derived parameters at these plots. The prediction error was strongly correlated to the difference in clay content between the two sublayers (Δ_{CLAY}). We obtained a preliminary threshold value of Δ_{CLAY}=20% below which the prediction error was insignificant. We explain this effect as retention of hydrocarbon effluents on clay mineral surfaces during pyrolysis as this is a common observation in literature (Espitalié et al., 1980; Zegouagh et al., 2004; Czirbus et al., 2016; Rahman et al., 2017). At these plots, none of the two calculation methods could account for the mineral effect. Qualitatively, we could observe that introducing clay-rich subsoil in the mixture caused a decrease in S2 and HI while it increased HC_PYR-derived temperature parameters (e.g., $T90_{HC\_PYR}$). Studies attempting to quantify the mineral effect of clays using the acid hydrolysis method propose values in the order of magnitude of 18-fold increase in S2 in the absence of minerals (Zegouagh et al., 2004). Although qualitatively interesting, these quantitative results are questionable due to the important loss of organic matter caused by this method (e.g., 17% in Zegouagh et al., 2004), that is believed to be preferential, influencing the bulk composition of SOM (Rumpel et al., 2006; Spaccini et al., 2013).

It would be helpful to have correction factors able to account for the changes in the hydrocarbon signal caused by specific textural differences. However, it is important to differentiate between the effect caused by textural differences and the influence of carbon content and clay mineralogy. This objective was not possible to achieve in this study, since the clay effect was an issue only in two plots from a single site.

A more extensive sample set including samples with similar texture and carbon content but different mineralogy (or reversely but most importantly changing only one variable at a time) would be necessary to assess the real effect of clay texture or clay mineralogy.

Finally, this first estimate of a mild mineral effect observed in natural soils due to clays is only characteristic of temperate soils. In different pedoclimates (e.g., tropical, Mediterranean) the linearity and comparability of Rock-Eval® parameters could be influenced by other minerals such as reactive clays and iron oxides (shown to interfere strongly with pyrolysis effluents; Huizinga et al., 1987; Ma et al., 2018).

3.2.1.5. Importance of equivalent soil mass

One aspect very shortly discussed in Chapter 2 is the role of equivalent soil mass in the calculation of Rock-Eval® parameters based on sublayers. The particularity of our experiment compared to real conditions of combining soil sublayer characteristics to infer information on a soil layer, is that we control and know precisely the mass of each endmember in each mixture. In reality, when averaging existing data, a correction will have to be undertaken to account for the soil mass contribution of each soil layer to the final "mixture". Thus, information on the fine fraction content and the density difference between layers will be required when calculating stocks of elements (Ellert and Bettany, 1995).

3.2.1.6. Uncertainty of Rock-Eval® parameters and error propagation

A general drawback of soil studies using Rock-Eval® is the lack of information regarding the uncertainty associated with the different parameters. However, a statistical analysis of a large soil sample data set, would probably lead to a conservative estimation of uncertainty due to the inherent heterogeneity of the soil material. Especially in this study, we can imagine that a propagation of the analytical error could have been insightful since it would allow to compare the prediction error to the propagated uncertainty, especially for more complex parameters calculated on multiple measurements (e.g., HI, PC/TOC, I-index).

3.3. Opening the black-box of Rock-Eval®; studying the effect of organo-mineral interactions in model systems

In Chapter 3 we attempt to progress towards a more mechanistic understanding of the link between *in situ* biogeochemical SOC stability and thermal stability assessed with the Rock-Eval® method. Based on a simple experimental set-up combining Rock-Eval® thermal analysis and a batch sorption experiment, we investigate the role of adsorption as a potential protection mechanism during thermal degradation of SOM. For simplification and reproducibility purposes we used pure biochemical compounds from multiple groups (lignin, humic acid, proteins, carbohydrates, lipids), pure minerals (kaolinite, montmorillonite, goethite) and three naturally occurring SOC-depleted soil mineral matrices. We prepared organo-mineral mixtures following a simple batch sorption protocol to study the resulting changes on thermal stability and Rock-Eval® parameters. As an intermediate step, we evaluated the effect of the presence of minerals in dry simple mixtures.

3.3.1. Detection of pure compounds using Rock-Eval®

Working with pure compounds revealed an important drawback of the Rock-Eval® method, namely a deficiency in carbon yield of oxygenated compounds (e.g., starch, cellulose, palmitic acid). We argue that this deficiency is due to the functioning principle of the FID as shown before (Dewar, 1961; Dietz, 1967; Maduskar et al., 2015), and we present evidence obtained using an experimental configuration of Rock-Eval® supporting that the missing carbon yield is associated to the pyrolysis step. This major issue is currently neglected in the only case we could find in literature that uses this technique to analyse pure compounds (Carrie et al., 2012). Very importantly, although this failure is crucial for our study it does not question the use of Rock-Eval® for soils as particulate organic matter carbon yield is >95%.

Ideally, to continue using Rock-Eval® for analysis of pure compounds, correction factors should be calculated to account for the missing yield (Dietz, 1967). Yet, valuable information such as the shape of the thermograms and the calculation of temperature parameters would still be lost.

3.3.2. Aspects defining the success of batch sorption experiments

Keeping only compounds with sufficiently good detection (>75%) we attempted to produce organo-mineral associations through a batch sorption experiment between each of three organic compounds (bovine serum albumin, cysteine, humic acid) and each of six minerals (kaolinite, montmorillonite, goethite, three natural mineral matrices). Our effort was successful for one of the proteins, bovine serum albumin, and all minerals (1>TOC>2 % wt.). In Chapter 3 we discuss in detail possible explanations for this outcome, and for the fact that for the other two compounds used here (humic acid and cysteine) the adsorption efficiency was very low (TOC<1% wt.). Below we list the main criteria which were not always ideal in this first attempt to form organo-mineral associations.

3.3.2.1. Selection of materials

The main requirement to create adsorption between organic matter and minerals is the formation of bonds depending on the attraction or repulsion forces and the physical barriers of the system. When selecting materials to use in a batch sorption experiment biochemicals with charged functional groups and smaller-sized molecules are better candidates. At the same time, minerals with high specific surface area and cation exchange capacity should be more prone to interact with organic compounds (Kleber et al., 2007; Sposito, 2008).

3.3.2.2. Solution pH and background electrolyte

Probably even more important than the selection of materials is the solution pH and the use of a background electrolyte when necessary, since they will define the type of bonds that can be created. According to the isoelectric point (IEP) of an organic compound and the synonymous point of zero charge (PZC) of a mineral, an ideal solution pH range should exist for a given pair of compounds in which one of the compounds is positively and the other one negatively charged. Setting the solution pH in this range would result in increased electrostatic attraction. In cases where this pH range cannot be reached, an alternative is to set the pH solution at high values, above the IEP of the mineral compound and the PZC of the mineral, causing them both to be negatively charged. Then, using a background electrolyte solution is recommended since it can favour formation of Ca^{2+} cation bridges. Although this is a tedious and time consuming step adjusting pH is crucial for the success of adsorption formation.

3.3.2.3. Solubility

A basic criterion but noteworthy due to its relevance to humic acid is the solubility of organic compounds in water-based solutions. For this specific example, although humic acid dissolves at pH>10 (Brigante et al., 2007) studies report maxima of adsorption formation at pH=4 (Chen et al., 2017). Moreover, the use of a background electrolyte is known to have a flocculating effect on humic acid even at very low Ca^{2+} concentrations (Brigante et al., 2007). Yet, references found in literature suggest using $CaCl_2$ to favour bridging the repulsive (negative) charges between humic acid and mineral surfaces (Feng et al., 2005). These contradictory results suggest that it might be necessary to measure and report dissolution efficiency or alternatively, use control treatments to ensure that incomplete dissolution or flocculation is not contributing to the final concentration of humic acid.

3.3.3. A strong mineral effect on the Rock-Eval® signal

As organo-mineral associations were successfully formed for bovine serum albumin, this compound was further considered to study the effect of adsorption on the Rock-Eval® signal. To distinguish the effect of adsorption from the effect of the presence of minerals in the pyrolysis oven we prepared and analysed simple dry mixtures with the same composition in terms of materials and carbon content as the produced adsorbed mixtures. We observed that goethite had the strongest influence on the Rock-Eval® signal, oxidizing pyrolysis effluents (in agreement with literature; Huizinga et al., 1987; Ma et al., 2018) and causing a carbon-yield deficit in the adsorbed mixture. Moreover, montmorillonite had a strong influence particularly on the hydrocarbon signal. The observed retention of hydrocarbons is a known effect of clay minerals, partially attributed to their large specific surface area (Spiro, 1991; Espitalié et al., 1984; Ma et al., 2018; Hayakawa et al., 2018). Soil mineral matrices had similar effects on the Rock-Eval® signal although less pronounced.

3.3.4. Inferring the effect of adsorption on thermal stability

Subtracting effect of the mineral presence is necessary to infer the effect of adsorption. Contrary to the effect of the mere presence of minerals, the adsorption had an unexpectedly low impact on the Rock-Eval®. We attribute this observation to the large size of the bovine serum albumin molecule.

Assuming that only specific sites of this large and complex protein are capable of building bonds with each mineral surface, this would mean that although a small part of the molecule is adsorbed, and might require more energy in order to be thermally degraded, the remaining larger part of the molecule can still react and be pyrolyzed or oxidized without delay. Testing this hypothesis would require repeating this experiment and successfully forming associations between smaller organic molecules and minerals.

3.3.5. The importance of differences in C content

Studies show that the less carbon there is in a sample the stronger the influence of the minerals becomes (Dahl et al., 2004; Zegouagh et al., 2004). Here although we adjusted the carbon content of dry mixtures to match the one of adsorbed mixtures to infer the effect of adsorption, the differences between the pairs are still important (e.g., bovine serum albumin-kaolinite mixture contains twice the amount of carbon present in the bovine serum albumin-goethite mixture). Moreover, the observed increase on thermal parameters of the pyrolysis step and persistent SOC proportion in adsorbed and dry mixtures is inversely proportional to the carbon content of the mixtures. This leads us to believe that for samples with low carbon concentrations (TOC<2 % wt.) even small differences (such as 1% wt.) might be important.

An interesting perspective would be to homogenize carbon content for all mixtures. At the moment, the easiest way to achieve this is to make the availability of carbon the limiting factor instead of the capacity of the system to support adsorption formation. In essence, that would require setting the initial carbon content in all solutions equal to the lowest adsorbed value observed (here 1% wt. for goethite). On the one hand it might be more difficult to achieve this objective in practice than in theory, as change in initial availability of organic compound will influence the system balance. Additionally, aiming for lower carbon values seems counterintuitive since mixture homogenization and detection issues become more important. On the other hand, in the optimistic case that optimal adsorption conditions can be reached through adaptations of the experimental protocol (mainly pH regulation) this might lead to overall higher final carbon concentrations in the produced organo-mineral associations.

3.3.6. Complementary methods

The interest of successfully produced organo-mineral associations extends further than the information that can be obtained with the Rock-Eval® method. Complementary methods could provide additional information at a smaller scale and help us understand better the formation of organo-mineral associations. Some examples include differential scanning calorimetry (DCS) to evaluate the change in energetic balance due to the existence of organo-mineral associations, precise elemental analysis of mixtures (e.g., C, H, N, S, and O for cysteine) with and without adsorption to detect potential stoichiometric shifts caused during the adsorption experiment. Additionally, imaging and chemical mapping techniques such as scanning and transmission electron microscopy could provide an idea of the spatial distribution of organic matter in the different mixtures.

3.3.7. Towards quantification

For the preliminary qualitative information obtained here to be useful for providing threshold values and guidelines and even for their incorporation into modelling approaches it is necessary to progress towards quantitative estimations and associated uncertainty. Regarding the limits of the Rock-Eval® method and its application in soil environments this study does not raise any red flags, since the materials and conditions of this experiment are far from any natural system.

Intentionally blank page

4. Conclusions and perspectives

The work presented in this thesis relies on an enormous amount of work conducted over the past decades by soil scientists, agronomists and technicians involved in the creation of the long-term field trials included in the LTBF, the AIAL and the SOEREPRO networks. Maintenance of these experimental sites, sample collection and analysis amounting to countless hours allow me to present this work today. Moreover, the work of my team of supervisors, advancing our understanding and quantification of SOC persistence for longer than a decade and creating the machine-learning model offered me a great and generous opportunity to use it and test it directly. Finally, the important parametrization work conducted on the AMG model in the last two decades and the access to the data necessary to run SOC simulations equipped me with a tool of immense value.

We present several advancements regarding the main objectives of this thesis— "*Understand and use the estimation of soil organic carbon persistence by Rock-Eval® thermal analysis*".

General conclusions

Our results show that the $PARTY_{SOC}$ model, a thermal-analysis-based machine-learning approach, can provide accurate information on SOC persistence in unknown samples. We highlight the ability of this approach to account for legacy effects of land-use history and we show how the information it provides can be used to initialize SOC simulations and improve their precision. The value of this thermal-analysis-based approach is clear, not only because of its predictive power but also due to practical reasons such as quick analysis time and low cost. The current need to estimate SOC persistence fulfilled by this approach supports the expansion of this method to a larger scale. Moreover, one of our main conclusions concerns the value of AMG, the simple SOC dynamics model used in this work. Due to the optimal precision of simulations provided with this model, when correctly initialized, we recommend that it should be used as an operational predictive tool for developing appropriate land management practices.

We believe that the simplicity and the efficiency of the AMG model is an important advantage over more recent and more complex SOC models that are very challenging to constrain and evaluate. The proposed approach combining simple SOC models with experimentally defined kinetic pools should result in more accurate predictions of SOC stock evolution on a national and even continental scale and it could support the implementation of efficient climate change mitigation policies.

In the context of applying this method on a larger scale we considered the question of applying this thermal analysis technique on soil samples already available but obtained at different depths as is often the case across different soil monitoring projects. We show that it is possible to infer the Rock-Eval® characterisation (including SOC persistence) of a soil layer from sublayer characteristics as the parameters obtained with this method are linearly additive for most temperate forest soils tested here. We underline that the calculation method has to be adapted to the type of parameter and we provide guidelines regarding the most appropriate procedures. We draw attention to a mineral effect hindering the prediction of hydrocarbon-related Rock-Eval® parameters in soils with pronounced clay content difference between soil sublayers (Δ_{CLAY}>20 wt.%). Our work can be considered as an empirical guideline paving the way towards harmonization of data obtained with this technique.

Finally, in an effort to open the black-box of Rock-Eval® we attempted to gain a better understanding regarding the interactions between organic matter and minerals influencing the Rock-Eval® signal and the associated estimations of thermal stability. We used a model system composed of pure organics and minerals as well as natural particulate organic matter and soil matrices. We provide insights regarding the deficiency of the Rock-Eval® method when used to analyse oxygenated compounds and we call for revision of former studies using similar compounds and ignoring this effect. We underline the strong effect of pure minerals such as goethite and montmorillonite on the Rock-Eval® signal, however this observation is mostly irrelevant to the soil environment due to the big differences in the nature of the used materials and the occurring conditions. The analysis of particulate organic matter confirms the ability of Rock-Eval® to detect SOM with high precision.

A slight increase in predictions of persistent SOC proportion is observed when adsorbed organo-mineral mixtures are analysed instead of simple dry mixtures of the same composition. However, there is no consistent trend amongst thermal parameters leading to the conclusion that the effect of the adsorption produced in this study on the Rock-Eval® signal is not significant. Moreover, this observation is overshadowed by the strong influence of the mineral matrix and the strong differences in carbon content across mixtures. We discuss possible options for favouring adsorption formation and we suggest that more work is necessary to decipher the link between thermal and biogeochemical stability.

General perspectives

An immediate perspective regarding the use of the PARTY$_{SOC}$ model on a national scale is its application on a wider sample set. An effort already underway is the use of PARTY$_{SOC}$ on samples from the RMQS network (Réseau de Mesures de la Qualité des Sols) comprising a total of 2240 sites, evenly distributed all over metropolitan France and Corsica according to a 16 km^2 grid. This would offer the opportunity of generating a map of SOC persistence that could contribute to the improvement of SOC stock evolution simulations at a national scale, with a clear significance as a decision making support tool (thesis project of Amicie Delahaie).

Moreover, a high priority is to include more sites in the learning set of the PARTY$_{SOC}$ model to increase its robustness and its genericity. Based on the good agreement between PARTY$_{SOC}$ predictions of SOC persistence and AMG-post-optimized pool partitioning observed in this study, we suggest that most agricultural LTEs with accurate AMG simulations could be used as reference sites for a future version of the PARTY$_{SOC}$ model. This could improve the precision of the PARTY$_{SOC}$ model in temperate pedoclimatic conditions and it could eventually lift an important technical limitation to its geographical expansion.

Before the continental or worldwide implementation of PARTY$_{SOC}$ in SOC modelling studies, its performance will have to be validated on a wider range of pedoclimates. Moreover, similar to this study, its potential to improve the precision of simulations will have to be demonstrated for the new application areas. To continue using the AMG model for these evaluation steps, LTEs from a variety of geographical locations and long-term monitoring data (climate, land cover, C inputs, management practices, soil characteristics) will be required to run simulations.

Another possibility would be to take advantage of literature on modelling work already conducted with other multi-compartmental models (e.g., Century, RothC) and test the potential of $PARTY_{SOC}$ to improve their simulations by initializing their "inert" or "passive" SOC pools.

The long-term the future of this approach could look like a constant adapting-expanding-and-improving cycle to make the most out of what SOC characterization has to offer. Eventually, other parameters (e.g., soil pH) could be added to the calibration set of the $PARTY_{SOC}$ model and its structure could evolve by testing other machine-learning algorithms.

Our work on linearity of Rock-Eval® parameters suggests that it is possible to infer the characterization of a soil layer based on its composing sublayers. Although this observation is in favour of the implementation of the $PARTY_{SOC}$ model on existing samples, it is important not to surpass the pedoclimatic limits for which this is true.

We suggest conducting similar but more extensive experiments covering a greater variability of soils with vertical textural and mineralogical heterogeneity. Better defined thresholds of mineral effect in natural soils could be helpful for predicting the limits of the method.

An interesting perspective related generally to the Rock Eval® method would be to estimate the uncertainty associated with each parameter. Similar indications exist but only for standard materials. Considering the strong heterogeneity of soil material, an estimation of the error of corresponding to different soil types or soil fractions would seem useful. This would allow uncertainty propagation in calculations such as the one presented in this work but also in others cases where Rock-Eval® acquisition parameters are used to infer more elaborate terms.

Regarding our understanding of the processes reflected in the Rock-Eval® signal that make it a suitable proxy for biogeochemical stability more work is necessary.

An important drawback of the results presented in this study that should be improved immediately is the variability in carbon content across mixtures. However, in case this perspective proves too challenging, an interesting approach would be to investigate the effect of carbon concentration on the intensity of the mineral effect, for different organic and mineral compounds.

More effort should be devoted to create stronger associations and further examine their influence on thermal stability. The solution pH should always be closely monitored and controlled since it can impact the type of bonds that can be generated. The influence of the molecular size of the used compounds on the efficiency of adsorption formation is an interesting aspect to be examined.

Although it seems trivial that complete dissolution of compounds in a stock solution should be confirmed, either by direct measurement or by use of a control, this is not always taken into account by batch sorption studies found in literature. Comparability of results obtained with these experiments would benefit from more standardized and commonly accepted procedures.

Also, other mechanisms at interplay that might influence the link between thermal and biogeochemical stability such as recalcitrance and the balance between energy gain and activation energy should also be studied. Use of complementary methods such as differential scanning calorimetry and transmission electron microscopy might provide valuable insights to help us understand better the formation of organo-mineral associations.

As a closing remark, it would be misleading at best to consider that soils are the only solution, or that they are an adequate solution on their own, for compensating global greenhouse gas emissions and fighting climate change. However, climate-smart management of global soils remains a practical win-win option that can significantly contribute to avoiding additional emissions and ensuring food security. It can be an ethical and socially responsible solution if we allow it.

5. References

References

Abramoff, R., Xu, X., Hartman, M., O'Brien, S., Feng, W., Davidson, E., Finzi, A., Moorhead, D., Schimel, J., Torn, M., and Mayes, M. A.: The Millennial model: in search of measurable pools and transformations for modeling soil carbon in the new century, Biogeochemistry, 137, 51–71, https://doi.org/10.1007/s10533-017-0409-7, 2018.

Ågren, G. and Bossata, E.: Theoretical Ecosystem Ecology: Understanding Element Cycles, Biology (Basel)., 1996.

Alcañiz, J., Romera, J., Comellas, L., Munne, R., and Puigbo, A.: Effects of some mineral matrices on flash pyrolysis-GC of soil humic substances, Sci. Total Environ., 81–82, 81–90, https://doi.org/10.1016/0048-9697(89)90113-7, 1989.

Alves, L., Medronho, B., Antunes, F. E., Topgaard, D., and Lindman, B.: Dissolution state of cellulose in aqueous systems. 2. Acidic solvents, Carbohydr. Polym., 151, 707–715, https://doi.org/10.1016/j.carbpol.2016.06.015, 2016.

Amelung, W., Bossio, D., de Vries, W., Kögel-Knabner, I., Lehmann, J., Amundson, R., Bol, R., Collins, C., Lal, R., Leifeld, J., Minasny, B., Pan, G., Paustian, K., Rumpel, C., Sanderman, J., van Groenigen, J. W., Mooney, S., van Wesemael, B., Wander, M., and Chabbi, A.: Towards a global-scale soil climate mitigation strategy, Nat. Commun., 11, 1–10, https://doi.org/10.1038/s41467-020-18887-7, 2020.

Amundson, R. and Biardeau, L.: Opinion: Soil carbon sequestration is an elusive climate mitigation tool, Proc. Natl. Acad. Sci. U. S. A., 115, 11652–11656, https://doi.org/10.1073/pnas.1815901115, 2018.

Amundson, R., Berhe, A. A., Hopmans, J. W., Olson, C., Sztein, A. E., and Sparks, D. L.: Soil and human security in the 21st century, Science (80-.)., 348, https://doi.org/10.1126/science.1261071, 2015.

Anderson, C. M., DeFries, R. S., Litterman, R., Matson, P. A., Nepstad, D. C., Pacala, S., Schlesinger, W. H., Rebecca Shaw, M., Smith, P., Weber, C., and Field, C. B.: Natural climate solutions are not enough, Science (80-.)., 363, 933–934, https://doi.org/10.1126/science.aaw2741, 2019.

Andrén, O. and Kätterer, T.: ICBM: The introductory carbon balance model for exploration of soil carbon balances, Ecol. Appl., 7, 1226–1236, https://doi.org/10.1890/1051-0761(1997)007[1226:ITICBM]2.0.CO;2, 1997.

Andriulo, A. E., Mary, B., and Guerif, J.: Modelling soil carbon dynamics with various cropping sequences on the rolling pampas, 19, 365–377, https://doi.org/10.1051/agro, 1999.

Global Carbon Project: https://www.globalcarbonproject.org/about/index.htm.

Aphalo, P. J.: Learn R ...as you learnt your mother tongue., 2016.

Arrouays, D., Balesdent, J., Germon, J. C., Payet, P. a., Soussana, J. F., and Stengel, P.: Stocker du carbone dans les sols agricoles de France, Expert. Sci. Collect., 36, 2002.

Arrouays, D., Jolivet, C., Boulonne, L., and Saby, N. P. A.: Le Réseau de Mesures de la Qualité des Sols de France (RMQS) Etat d ' avancement et premiers résultats, 2003.

Autret, B., Mary, B., Chenu, C., Balabane, M., Girardin, C., Bertrand, M., Grandeau, G., and

Beaudoin, N.: Alternative arable cropping systems: A key to increase soil organic carbon storage? Results from a 16 year field experiment, Agric. Ecosyst. Environ., https://doi.org/10.1016/j.agee.2016.07.008, 2016.

Bailey, V. L., Bond-Lamberty, B., DeAngelis, K., Grandy, A. S., Hawkes, C. V., Heckman, K., Lajtha, K., Phillips, R. P., Sulman, B. N., Todd-Brown, K. E. O., and Wallenstein, M. D.: Soil carbon cycling proxies: Understanding their critical role in predicting climate change feedbacks, Glob. Chang. Biol., 24, 895–905, https://doi.org/10.1111/gcb.13926, 2018.

Baldock, J. A. and Skjemstad, J. O.: Role of the soil matrix and minerals in protecting natural organic materials against biological attack, 31, 697–710, 2000.

Baldock, J. A., Hawke, B., Sanderman, J., and Macdonald, L. M.: Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra, Soil Res., 51, 577, https://doi.org/10.1071/SR13077, 2013.

Balesdent, J.: The turnover of soil organic fractions estimated by radiocarbon dating, Sci. Total Environ., 62, 405–408, https://doi.org/10.1016/0048-9697(87)90528-6, 1987.

Balesdent, J.: Estimation du renouvellement du carbone des sols par mesure isotopique ^{13}C Precision, risque de biais, Cah. l'Orstom, 26, 315–326, 1991.

Balesdent, J.: The significance of organic separates to carbon dynamics and its modelling in some cultivated soils, Eur. J. Soil Sci., 47, 485–493, https://doi.org/10.1111/j.1365-2389.1996.tb01848.x, 1996.

Balesdent, J. and Arrouays, D.: Usage des terres et stockage de carbone dans les sols du territoire français. Une estimation des flux nets pour la période 1900-1999, Comptes rendus l'Académie d'Agriculture Fr., 85, 265–277, 1999.

Balesdent, J. and Balabane, M.: Major contribution of roots to soil carbon storage inferred from maize cultivated soils, Soil Biol. Biochem., 28, 1261–1263, https://doi.org/10.1016/0038-0717(96)00112-5, 1996.

Balesdent, J. and Recous, S.: Les temps de résidence du carbone et le potentiel de stockage de carbone dans quelques sols cultivés français, Can. J. Soil Sci., 77, 187–193, https://doi.org/10.4141/s96-109, 1997.

Balesdent, J., Mariotti, A., and Guillet, B.: Natural 13C abundance as a tracer for studies of soil organic matter dynamics, Soil Biol. Biochem., 19, 25–30, https://doi.org/10.1016/0038-0717(87)90120-9, 1987.

Balesdent, J., Basile-Doelsch, I., Chadoeuf, J., Cornu, S., Derrien, D., Fekiacova, Z., and Hatté, C.: Atmosphere–soil carbon transfer as a function of soil depth, Nature, 559, 599–602, https://doi.org/10.1038/s41586-018-0328-3, 2018.

Barré, P.: Mémoire de synthèse de mes activités scientifiques, Ecole Normale Superieure Paris, 1–59 pp., 2020.

Barré, P., Montagnier, C., Chenu, C., Abbadie, L., and Velde, B.: Clay minerals as a soil potassium reservoir: Observation and quantification through X-ray diffraction, Plant Soil, 302, 213–220, https://doi.org/10.1007/s11104-007-9471-6, 2008.

Barré, P., Eglin, T., Christensen, B. T., Ciais, P., Houot, S., Kätterer, T., Van Oort, F., Peylin, P., Poulton, P. R., Romanenkov, V., and Chenu, C.: Quantifying and isolating stable soil organic carbon using long-term bare fallow experiments, 7, 3839–3850, https://doi.org/10.5194/bg-7-3839-2010, 2010.

Barré, P., Plante, A. F., Cécillon, L., Lutfalla, S., Baudin, F., Bernard, S., Christensen, B. T.,

Eglin, T., Fernandez, J. M., Houot, S., Kätterer, T., Le Guillou, C., Macdonald, A., van Oort, F., and Chenu, C.: The energetic and chemical signatures of persistent soil organic matter, Biogeochemistry, 130, https://doi.org/10.1007/s10533-016-0246-0, 2016.

Barthès, B. G., Brunet, D., Hien, E., Enjalric, F., Conche, S., Freschet, G. T., D'Annunzio, R., and Toucet-Louri, J.: Determining the distributions of soil carbon and nitrogen in particle size fractions using near-infrared reflectance spectrum of bulk soil samples, Soil Biol. Biochem., 40, 1533–1537, https://doi.org/10.1016/j.soilbio.2007.12.023, 2008.

Batjes, N. H.: Total carbon and nitrogen in the soils of the world, Eur. J. Soil Sci., 65, 10–21, https://doi.org/10.1111/ejss.12114_2, 1996.

Batjes, N. H.: Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks, Geoderma, 269, 61–68, https://doi.org/10.1016/j.geoderma.2016.01.034, 2016.

Batjes, N. H., Ribeiro, E., and van Oostrum, A.: Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019), Earth Syst. Sci. Data, 12, 299–320, https://doi.org/10.5194/essd-12-299-2020, 2020.

Baudin, F., Disnar, J. R., Aboussou, A., and Savignac, F.: Guidelines for Rock-Eval analysis of recent marine sediments, Org. Geochem., 86, 71–80, https://doi.org/10.1016/j.orggeochem.2015.06.009, 2015.

Behar, F., Beaumont, V., and De B. Penteado, H. L.: Rock-Eval 6 Technology: Performances and Developments, Oil Gas Sci. Technol., 56, 111–134, https://doi.org/10.2516/ogst:2001013, 2001.

Beleites, C. and Sergo, V.: hyperSpec: a package to handle hyperspectral data sets in R, R package version 0.99-20180627., http://hyperspec.r-forge.r-project.org, 2018.

Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M., and Kirk, G. J. D.: Carbon losses from all soils across England and Wales 1978-2003, Nature, 437, 245–248, https://doi.org/10.1038/nature04038, 2005.

Berhe, A. A.: A climate change solution that's right under our feet [Video]., https://www.ted.com/talks/asmeret_asefaw_berhe_a_climate_change_solution_that_s_right_under_our_feet, 2019.

FCBA network: https://www6.inrae.fr/in-sylva-france_eng/Services/In-Situ/FCBA-network, last access: 20 May 2021.

Blankinship, J. C., Berhe, A. A., Crow, S. E., Druhan, J. L., Heckman, K. A., Keiluweit, M., Lawrence, C. R., Marín-Spiotta, E., Plante, A. F., Rasmussen, C., Schädel, C., Schimel, J. P., Sierra, C. A., Thompson, A., Wagai, R., and Wieder, W. R.: Improving understanding of soil organic matter dynamics by triangulating theories, measurements, and models, Biogeochemistry, 140, 1–13, https://doi.org/10.1007/s10533-018-0478-2, 2018.

Bleam, W.: Soil and Environmental Chemistry, second., Elsevier, 573 pp., 2017.

Boiffin, J., Zagbahi, J. K., and Sebillotte, M.: Systèmes de culture et statut organique des sols dans le Noyonnais : application du modèle de Hénin-Dupuis, 6, 437–446, https://doi.org/10.1051/agro:19860503, 1986.

Bolinder, M. A., Janzen, H. H., Gregorich, E. G., Angers, D. A., and VandenBygaart, A. J.: An approach for estimating net primary productivity and annual carbon inputs to soil for common agricultural crops in Canada, Agric. Ecosyst. Environ., 118, 29–42,

https://doi.org/10.1016/j.agee.2006.05.013, 2007.

Borchers, H. W.: pracma: Practical Numerical Math Functions. R package version 2.2.2, https://cran.r-project.org/package=pracma, 2018.

Bossio, D. A., Cook-Patton, S. C., Ellis, P. W., Fargione, J., Sanderman, J., Smith, P., Wood, S., Zomer, R. J., von Unger, M., Emmer, I. M., and Griscom, B. W.: The role of soil carbon in natural climate solutions, Nat. Sustain., 3, 391–398, https://doi.org/10.1038/s41893-020-0491-z, 2020.

Bouthier, A., Duparque, A., Mary, B., Sagot, S., Trochard, R., Levert, M., Houot, S., Damay, N., Denoroy, P., Dinh, J.-L., Blin, B., and Ganteil, F.: Adaptation et mise en oeuvre du modèle de calcul de bilan humique à long terme AMG dans une large gamme de systèmes de grandes cultures et de polyculture-élevage, Innov. Agron., 34, 367–378, 2014.

Brigante, M., Zanini, G., and Avena, M.: On the dissolution kinetics of humic acid particles. Effects of pH, temperature and Ca2+ concentration, Colloids Surfaces A Physicochem. Eng. Asp., 294, 64–70, https://doi.org/10.1016/j.colsurfa.2006.07.045, 2007.

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y. M., Cellier, P., Debaeke, P., Gaudillère, J. P., Hénault, C., Maraux, F., Seguin, B., and Sinoquet, H.: An overview of the crop model STICS, in: European Journal of Agronomy, 309–332, https://doi.org/10.1016/S1161-0301(02)00110-7, 2003.

Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F. X., and Huard, F.: Why are wheat yields stagnating in Europe? A comprehensive data analysis for France, F. Crop. Res., 119, 201–212, https://doi.org/10.1016/j.fcr.2010.07.012, 2010.

Bruni, E., Guenet, B., and Chenu, C.: Increasing soil organic carbon storage: The 4 per 1000 objective through a modelling based approach (Oral presentation), in: EUROSOIL, 2021.

Bruun, S. and Jensen, L. S.: Initialisation of the soil organic matter pools of the Daisy model, Ecol. Modell., 153, 291–295, 2002.

Bruun, S., Ågren, G. I., Christensen, B. T., and Jensen, L. S.: Measuring and modeling continuous quality distributions of soil organic matter, 7, 27–41, https://doi.org/10.5194/bg-7-27-2010, 2010.

Bu, H., Yuan, P., Liu, H., Liu, D., and Liu, J.: ScienceDirect Effects of complexation between organic matter (OM) and clay mineral on OM pyrolysis, Geochim. Cosmochim. Acta, 212, 1–15, https://doi.org/10.1016/j.gca.2017.04.045, 2017.

Cagnarini, C., Renella, G., Mayer, J., Hirte, J., Schulin, R., Costerousse, B., Della Marta, A., Orlandini, S., and Menichetti, L.: Multi-objective calibration of RothC using measured carbon stocks and auxiliary data of a long-term experiment in Switzerland, Eur. J. Soil Sci., 70, 819–832, https://doi.org/10.1111/ejss.12802, 2019.

Calamai, L., Lozzi, I., Stotzky, G., Fusi, P., and Ristori, G. .: Interaction of catalase with montmorillonite homoionic to cations with different hydrophobicity: effect on enzymatic activity and microbial utilization, Soil Biol. Biochem., 32, 815–823, https://doi.org/10.1016/S0038-0717(99)00211-4, 2000.

Campbell, E. E. and Paustian, K.: Current developments in soil organic matter modeling and the expansion of model applications: A review, Environ. Res. Lett., 10, https://doi.org/10.1088/1748-9326/10/12/123004, 2015.

Carpenter, S. R.: Decay of heterogeneous detritus: A general model, J. Theor. Biol., 89, 539–

547, 1981.

Carrie, J., Sanei, H., and Stern, G.: Standardisation of Rock-Eval pyrolysis for the analysis of recent sediments and soils, Org. Geochem., 46, 38–53, https://doi.org/10.1016/j.orggeochem.2012.01.011, 2012a.

Carrie, J., Sanei, H., and Stern, G.: Standardisation of Rock–Eval pyrolysis for the analysis of recent sediments and soils, Org. Geochem., 46, 38–53, https://doi.org/10.1016/j.orggeochem.2012.01.011, 2012b.

Cécillon, L.: A dual response, Nat. Geosci., 14, 262–263, https://doi.org/10.1038/s41561-021-00749-6, 2021a.

Cécillon, L.: lauric-cecillon/PARTYsoc: Second version of the PARTYsoc statistical model (Version v2.0). Zenodo, http://doi.org/10.5281/zenodo.4446138, 2021b.

Cécillon, L., Soucemarianadin, L., Berthelot, A., Duverger, M., De Boisseson, J.-M., Gosselin, F., Guenet, B., Barthès, B., De Danieli, S., Barrier, R., Abiven, S., Chenu, C., Girardin, C., Baudin, F., Savignac, F., Nicolas, M., Mériguet, J., and Barré, P.: piCaSo : pilotage sylvicole et contrôle pédologique des stocks de carbone des sols forestiers, 103 pp., 2017.

Cécillon, L., Baudin, F., Chenu, C., Houot, S., Jolivet, R., Kätterer, T., Lutfalla, S., Macdonald, A., Van Oort, F., Plante, A. F., Savignac, F., Soucémarianadin, L. N., and Barré, P.: A model based on Rock-Eval thermal analysis to quantify the size of the centennially persistent organic carbon pool in temperate soils, 15, 2835–2849, https://doi.org/10.5194/bg-15-2835-2018, 2018.

Cécillon, L., Baudin, F., Chenu, C., Christensen, B. T., Franko, U., Houot, S., Kanari, E., Kätterer, T., Merbach, I., van Oort, F., Poeplau, C., Quezada, J. C., Savignac, F., Soucémarianadin, L. N., and Barré, P.: Partitioning soil organic carbon into its centennially stable and active fractions with machine-learning models based on Rock-Eval® thermal analysis (PARTY$_{SOC}$v2.0 and PARTY$_{SOC}$v2.0$_{EU}$), Geosci. Model Dev., 14, 3879–3898, https://doi.org/10.5194/gmd-14-3879-2021, 2021.

Chassé, M., Luftalla, S., Cécillon, L., Baudin, F., Abiven, S., Chenu, C., and Barré, P.: Long-Term bare-fallow soil fractions reveal thermo-chemical properties controlling soil organic carbon dynamics, 18, 1703–1718, https://doi.org/10.5194/bg-18-1703-2021, 2021.

Chen, H., Koopal, L. K., Xiong, J., Avena, M., and Tan, W.: Mechanisms of soil humic acid adsorption onto montmorillonite and kaolinite, J. Colloid Interface Sci., 504, 457–467, https://doi.org/10.1016/j.jcis.2017.05.078, 2017.

Chenu, C. and Stotzky, G.: Interactions between Microorganisms and Soil Particles: An Overview, in: Interactions between Soil Particles and Microorganisms: Impact on the Terrestrial Ecosystem, edited by: Huang, P. M., Bollag, J.-M., and Sensei, N., Wiley, 584, 2002.

Chevallier, T., Muchaonyerwa, P., and Chenu, C.: Microbial utilisation of two proteins adsorbed to a vertisol clay fraction : toxin from Bacillus thuringiensis subsp . tenebrionis and bovine serum albumin, 35, 1211–1218, https://doi.org/10.1016/S0038-0717(03)00182-2, 2003.

Christensen, B. T.: Physical Fractionation of Soil and Organic Matter in Primary Particle Size and Density Separates, Adv. Soil Sci., 20, 1–90, https://doi.org/10.1007/978-1-4612-2930-8_1, 1992.

Clivot, H., Mary, B., Valé, M., Cohan, J. P., Champolivier, L., Piraux, F., Laurent, F., and

Justes, E.: Quantifying in situ and modeling net nitrogen mineralization from soil organic matter in arable cropping systems, Soil Biol. Biochem., 111, 44–59, https://doi.org/10.1016/j.soilbio.2017.03.010, 2017.

Clivot, H., Mouny, J. C., Duparque, A., Dinh, J. L., Denoroy, P., Houot, S., Vertès, F., Trochard, R., Bouthier, A., Sagot, S., and Mary, B.: Modeling soil organic carbon evolution in long-term arable experiments with AMG model, Environ. Model. Softw., 118, 99–113, https://doi.org/10.1016/j.envsoft.2019.04.004, 2019.

Coleman, K. and Jenkinson, D. S.: RothC-26.3 - A Model for the turnover of carbon in soil, in: Evaluation of Soil Organic Matter Models, Springer Berlin Heidelberg, Berlin, Heidelberg, 237–246, https://doi.org/10.1007/978-3-642-61094-3_17, 1996.

Coleman, K., Jenkinson, D. S., Crocker, G. J., Grace, P. R., Klír, J., Körschens, M., Poulton, P. R., and Richter, D. D.: Simulating trends in soil organic carbon in long-term experiments using RothC-26.3, Geoderma, 81, 29–44, https://doi.org/10.1016/S0016-7061(97)00079-7, 1997.

Colomb, B., Debaeke, P., Jouany, C., and Nolot, J. M.: Phosphorus management in low input stockless cropping systems: Crop and soil responses to contrasting P regimes in a 36-year experiment in southern France, Eur. J. Agron., 26, 154–165, https://doi.org/10.1016/j.eja.2006.09.004, 2007.

Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K., and Paul, E.: The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: Do labile plant inputs form stable soil organic matter?, Glob. Chang. Biol., 19, 988–995, https://doi.org/10.1111/gcb.12113, 2013.

Cotrufo, M. F., Ranalli, M. G., Haddix, M. L., Six, J., and Lugato, E.: Soil carbon storage informed by particulate and mineral-associated organic matter, Nat. Geosci., 12, 989–994, https://doi.org/10.1038/s41561-019-0484-6, 2019.

Crowther, T. W., Riggs, C., Lind, E. M., Borer, E. T., Seabloom, E. W., Hobbie, S. E., Wubs, J., Adler, P. B., Firn, J., Gherardi, L., Hagenah, N., Hofmockel, K. S., Knops, J. M. H., McCulley, R. L., MacDougall, A. S., Peri, P. L., Prober, S. M., Stevens, C. J., and Routh, D.: Sensitivity of global soil carbon stocks to combined nutrient enrichment, Ecol. Lett., 22, 936–945, https://doi.org/10.1111/ele.13258, 2019.

Czirbus, N., Nyilas, T., Raucsik, B., and Hetényi, M.: Investigation of the effect of soil mineral composition on soil organic matter stability, Soil Water Res., 11, 147–154, https://doi.org/10.17221/47/2015-SWR, 2016.

Dangal, S. R. S., Schwalm, C., Cavigelli, M. A., Gollany, H. T., Jin, V. L., and Sanderman, J.: Improving soil carbon estimates by linking conceptual pools against measurable carbon fractions in the DAYCENT Model Version 4.5, J. Adv. Model. Earth Syst., 2021.

Davis, J. B. and Stanley, J. P.: Catalytic effect of smectite clays in hydrocarbon generation revealed by pyrolysis-gas chromatography, J. Anal. Appl. Pyrolysis, 4, 227–240, https://doi.org/10.1016/0165-2370(82)80033-8, 1982.

Derrick, M. R., Stulik, D., and Landry, J. M.: Infrared Spectroscopy in conservation science, edited by: Ball, T. and Tidwell, S., The Getty Conservation Institute, https://doi.org/10.1002/9781118162897.ch5, 1999.

Dewar, R. A.: The flame ionization detector a theoretical approach, J. Chromatogr. A, 6, 312–323, https://doi.org/10.1016/s0021-9673(61)80265-3, 1961.

Di-Giovanni, C., Disnar, J.-R., Bakyono, J.-P., Kéravis, D., Millet, F., and Olivier, J.-E.:

Application de l'étude de la matière organique à l'analyse de l'érosion : exemple du bassin versant du Moulin, dans les terres noires des Alpes-de-Haute-Provence (France), Comptes Rendus l'Académie des Sci. - Ser. IIA - Earth Planet. Sci., 331, 7–14, https://doi.org/10.1016/S1251-8050(00)01384-7, 2000.

Dietz, W. A.: Response Factors for Gas Chromatographic Analyses, 1967.

Dignac, M. F., Derrien, D., Barré, P., Barot, S., Cécillon, L., Chenu, C., Chevallier, T., Freschet, G. T., Garnier, P., Guenet, B., Hedde, M., Klumpp, K., Lashermes, G., Maron, P. A., Nunan, N., Roumet, C., and Basile-Doelsch, I.: Increasing soil carbon storage: mechanisms, effects of agricultural practices and proxies. A review, Agron. Sustain. Dev., 37, https://doi.org/10.1007/s13593-017-0421-2, 2017.

Dimassi, B., Mary, B., Wylleman, R., Labreuche, J. Ô., Couture, D., Piraux, F., and Cohan, J. P.: Long-term effect of contrasted tillage and crop management on soil carbon dynamics during 41 years, Agric. Ecosyst. Environ., 188, 134–146, https://doi.org/10.1016/j.agee.2014.02.014, 2014.

Disnar, J. R., Guillet, B., Keravis, D., Di-Giovanni, C., and Sebag, D.: Soil organic matter (SOM) characterization by Rock-Eval pyrolysis: Scope and limitations, Org. Geochem., 34, 327–343, https://doi.org/10.1016/S0146-6380(02)00239-5, 2003.

Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Casanova Pinto, M., Casanova-Katny, A., Muñoz, C., Boudin, M., Zagal Venegas, E., and Boeckx, P.: Soil carbon storage controlled by interactions between geochemistry and climate, Nat. Geosci., 8, 780–783, https://doi.org/10.1038/ngeo2516, 2015.

Dondini, M., Hastings, A., Saiz, G., Jones, M. B., and Smith, P.: The potential of Miscanthus to sequester carbon in soils: Comparing field measurements in Carlow, Ireland to model predictions, GCB Bioenergy, 1, 413–425, https://doi.org/10.1111/j.1757-1707.2010.01033.x, 2009.

Doran, J. W. and Parkin, T. B.: Defining and assessing soil quality, in: Defining Soil Quality for a Sustainable En- vironment, Soil Science Society of America, Madison, 3–21, 1994.

Dungait, J. A. J., Hopkins, D. W., Gregory, A. S., and Whitmore, A. P.: Soil organic matter turnover is governed by accessibility not recalcitrance, Glob. Chang. Biol., 18, 1781–1796, https://doi.org/10.1111/j.1365-2486.2012.02665.x, 2012.

Duparque, A., Dinh, J., Mary, B., Bouthier, A., Blin, B., Denoroy, P., Ganteil, F., Houot, S., Levert, M., Sagot, S., Trochard, R., Territoires, A. R., and Brunehaut, C.: AMG : a simple SOC balance model used in France for decision support AMG-Research Tool Postgre Database, 2013.

EEA: Soil, land and climate change, 2019.

Eglin, T., Ciais, P., Piao, S. L., Barre, P., Bellassen, V., Cadule, P., Chenu, C., Gasser, T., Koven, C., Reichstein, M., and Smith, P.: Historical and future perspectives of global soil carbon response to climate and land-use changes, Tellus, Ser. B Chem. Phys. Meteorol., 62, 700–718, https://doi.org/10.1111/j.1600-0889.2010.00499.x, 2010.

Ekschmitt, K., Kandeler, E., Poll, C., Brune, A., Buscot, F., Friedrich, M., Gleixner, G., Hartmann, A., Kästner, M., Marhan, S., Miltner, A., Scheu, S., and Wolters, V.: Soil-carbon preservation through habitat constraints and biological limitations on decomposer activity, J. Plant Nutr. Soil Sci., 171, 27–35, https://doi.org/10.1002/jpln.200700051, 2008.

Ellam, R.: Isotopes: A very short introduction, Oxford University Press,

https://doi.org/10.1093/actrade/9780198723622.001.0001, 2016.

Ellert, B. H. and Bettany, J. R.: Calculation of organic matter and nutrients stored in soils under contrasting management regimes, Can. J. Soil Sci., 75, 529–538, https://doi.org/10.4141/cjss95-075, 1995.

Elliott, E. T., Paustian, K., and Frey, S. D.: Modeling the Measurable or Measuring the Modelable: A Hierarchical Approach to Isolating Meaningful Soil Organic Matter Fractionations, in: Evaluation of Soil Organic Matter Models, Springer Berlin Heidelberg, 161–179, 1996.

Erb, K. H., Luyssaert, S., Meyfroidt, P., Pongratz, J., Don, A., Kloster, S., Kuemmerle, T., Fetzel, T., Fuchs, R., Herold, M., Haberl, H., Jones, C. D., Marín-Spiotta, E., McCallum, I., Robertson, E., Seufert, V., Fritz, S., Valade, A., Wiltshire, A., and Dolman, A. J.: Land management: data availability and process understanding for global change studies, Glob. Chang. Biol., 23, 512–533, https://doi.org/10.1111/gcb.13443, 2017.

Espitalie, J., Deroo, G., and Marquis, F.: La pyrolyse Rock-Eval et ses applications. Deuxième partie., Rev. l'Institut Français du Pétrole, 40, 755–784, https://doi.org/10.2516/ogst:1985045, 1985a.

Espitalie, J., Deroo, G., and Marquis, F.: La pyrolyse Rock-Eval et ses applications. Première partie., Rev. l'Institut Français du Pétrole, 40, 563–579, https://doi.org/10.2516/ogst:1985035, 1985b.

Espitalie, J., Deroo, G., and Marquis, F.: La pyrolyse Rock-Eval et ses applications. Troisième partie., Rev. l'Institut Français du Pétrole, 41, 73–89, https://doi.org/10.2516/ogst:1986003, 1986.

Espitalié, J., Laporte, J. L., Madec, M., Marquis, F., Leplat, P., Paulet, J., and Boutefeu, A.: Méthode rapide de caractérisation des roches mètres, de leur potentiel pétrolier et de leur degré d'évolution, Rev. l'Institut Français du Pétrole, 32, 23–42, https://doi.org/10.2516/ogst:1977002, 1977.

Espitalié, J., Madec, M., and Tissot, B.: Role of mineral matrix in kerogen pyrolysis: influence on petroleum generation and migration., Am. Assoc. Pet. Geol. Bull., 64, 59–66, https://doi.org/10.1306/2f918928-16ce-11d7-8645000102c1865d, 1980.

Espitalié, J., Senga Makadi, K., and Trichet, J.: Role of the mineral matrix during kerogen pyrolysis, Org. Geochem., 6, 365–382, 1984.

European Commision: Proposal for a directive of the European Parliament and the council establishing a framework for the protection of soil amending directive, https://doi.org/2004/35/EC, 2006.

European Commission: The European Green Deal, Eur. Comm., 53, 24, https://doi.org/10.1017/CBO9781107415324.004, 2019.

LUCAS: Land Use and Coverage Area frame Survey: https://esdac.jrc.ec.europa.eu/projects/lucas, last access: 20 May 2021.

Eusterhues, K., Rumpel, C., Kleber, M., and Kögel-Knabner, I.: Stabilisation of soil organic matter by interactions with minerals as revealed by mineral dissolution and oxidative degradation, Org. Geochem., 34, 1591–1600, https://doi.org/10.1016/j.orggeochem.2003.08.007, 2003.

Faghihian, H. and Nejati-Yazdinejad, M.: Equilibrium study of adsorption of L-cysteine by natural bentonite, Clay Miner., 44, 125–133,

https://doi.org/10.1180/claymin.2009.044.1.125, 2009.

Fan, J., McConkey, B., Wang, H., and Janzen, H.: Root distribution by depth for temperate agricultural crops, F. Crop. Res., 189, 68–74, https://doi.org/10.1016/j.fcr.2016.02.013, 2016.

FAO: The importance of soil organic matter. Key to drought-resistant soil and sustained food production., Food and Agriculture Organization of the United Nations, Rome, Italy, https://doi.org/10.1080/03650340214162, 2005.

International Year of Soils 2015: http://www.fao.org/soils-2015/about/en/, last access: 20 May 2021.

FAO: Soil Organic Carbon: the hidden potential., Food and Agriculture Organization of the United Nations, Rome, Italy, https://doi.org/10.1038/nrg2350, 2017.

FAO: Measuring and modelling soil carbon stocks and stock changes in livestock production systems – Guidelines for assessment (Draft for public review), Rome, Italy, 2018.

FAO: Recarbonization of global soils: a dynamic response to offset global emissions, 8 pp., 2019.

FAO: Technical specifications and country guidelines for Global Soil Organic Carbon Sequestration Potential Map (GSOCseq), Rome, 2020.

FAO & ITPS: Intergovernmental Technical Panel on Soils. Status of the World's Soil Resources., 100–146 pp., 2015.

Farina, R., Sándor, R., Abdalla, M., Álvaro-Fuentes, J., Bechini, L., Bolinder, M. A., Brilli, L., Chenu, C., Clivot, H., De Antoni Migliorati, M., Di Bene, C., Dorich, C. D., Ehrhardt, F., Ferchaud, F., Fitton, N., Francaviglia, R., Franko, U., Giltrap, D. L., Grant, B. B., Guenet, B., Harrison, M. T., Kirschbaum, M. U. F., Kuka, K., Kulmala, L., Liski, J., McGrath, M. J., Meier, E., Menichetti, L., Moyano, F., Nendel, C., Recous, S., Reibold, N., Shepherd, A., Smith, W. N., Smith, P., Soussana, J. F., Stella, T., Taghizadeh-Toosi, A., Tsutskikh, E., and Bellocchi, G.: Ensemble modelling, uncertainty and robust predictions of organic carbon in long-term bare-fallow soils, Glob. Chang. Biol., 27, 904–928, https://doi.org/10.1111/gcb.15441, 2021.

Feller, C. and Bernoux, M.: Historical advances in the study of global terrestrial soil organic carbon sequestration, Waste Manag., 28, 734–740, https://doi.org/10.1016/j.wasman.2007.09.022, 2008.

Feng, X., Simpson, A. J., and Simpson, M. J.: Chemical and mineralogical controls on humic acid sorption to clay mineral surfaces, Org. Geochem., 36, 1553–1566, https://doi.org/10.1016/j.orggeochem.2005.06.008, 2005.

Fernández-Ugalde, O., Barré, P., Hubert, F., Virto, I., Girardin, C., Ferrage, E., Caner, L., and Chenu, C.: Clay mineralogy differs qualitatively in aggregate-size classes: clay-mineral-based evidence for aggregate hierarchy in temperate soils, Eur. J. Soil Sci., 64, 410–422, https://doi.org/10.1111/ejss.12046, 2013.

Fernández-Ugalde, O., Barré, P., Virto, I., Hubert, F., Billiou, D., and Chenu, C.: Does phyllosilicate mineralogy explain organic matter stabilization in different particle-size fractions in a 19-year C3/C4 chronosequence in a temperate Cambisol?, Geoderma, 264, 171–178, https://doi.org/10.1016/j.geoderma.2015.10.017, 2016.

FEU-US: The economic case for climate action in the United States, https://feu-us.org/our-work/case-for-climate-action-us/, 2017.

Friedlingstein, P., Sullivan, M. O., Jones, M. W., Andrew, R. M., and Hauck, J.: Data description paper Global Carbon Budget 2020, Earth Syst. Sci. Data, 2020, 3269–3340, 2020.

Fuchs, R., Herold, M., Verburg, P. H., Clevers, J. G. P. W., and Eberle, J.: Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010, Glob. Chang. Biol., 21, 299–313, https://doi.org/10.1111/gcb.12714, 2015.

Gale, M. R. and Grigal, D. F.: Vertical root distributions of northern tree species in relation to successional status, Can. J. For. Res., 18, 829–834, 1987.

Gleixner, G., Poirier, N., Bol, R., and Balesdent, J.: Molecular dynamics of organic matter in a cultivated soil, Org. Geochem., 33, 357–366, https://doi.org/10.1016/S0146-6380(01)00166-8, 2002.

Gosselin, F., Brun, J.-J., Cécillon, L., De Danieli, S., Goïtre, T., Daumergue, N., Tardif, P., Heiniger, C., Wei, L., Bilger, I., Hullin, F., Villemey, A., Dumas, Y., Chevalier, R., Archaux, F., and Paillet, Y.: Gestion, naturalité, biodiversité - volet sol (GNB-sol), Ministère la Transit. écologique solidaire., 2015.

Gregorich, E. G., Beare, M. H., McKim, U. F., and Skjemstad, J. O.: Chemical and Biological Characteristics of Physically Uncomplexed Organic Matter, Soil Sci. Soc. Am. J., 70, 975–985, https://doi.org/10.2136/sssaj2005.0116, 2006.

Gregorich, E. G., Gillespie, A. W., Beare, M. H., Curtin, D., Sanei, H., and Yanni, S. F.: Evaluating biodegradability of soil organic matter by its thermal stability and chemical composition, Soil Biol. Biochem., 91, 182–191, https://doi.org/10.1016/j.soilbio.2015.08.032, 2015.

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamäki, J. V., Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campari, J., Conant, R. T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M. R., Herrero, M., Kiesecker, J., Landis, E., Laestadius, L., Leavitt, S. M., Minnemeyer, S., Polasky, S., Potapov, P., Putz, F. E., Sanderman, J., Silvius, M., Wollenberg, E., and Fargione, J.: Natural climate solutions, Proc. Natl. Acad. Sci. U. S. A., 114, 11645–11650, https://doi.org/10.1073/pnas.1710465114, 2017.

Hatfield, J. L., Sauer, T. J., and Cruse, R. M.: Soil: The Forgotten Piece of the Water, Food, Energy Nexus, 1st ed., Elsevier Inc., 1–46 pp., https://doi.org/10.1016/bs.agron.2017.02.001, 2017.

Hayakawa, C., Fujii, K., Funakawa, S., and Kosaki, T.: Effects of sorption on biodegradation of low-molecular-weight organic acids in highly-weathered tropical soils, Geoderma, 324, 109–118, https://doi.org/10.1016/j.geoderma.2018.03.014, 2018.

Hayes, M. H. B.: Soil organic matter extraction, fractionation, structure and effects on soil structure, in: The Role of Organic Matter in Modern Agriculture. Developments in Plant and Soil Sciences, edited by: Chen, Y. and Avnimelech, Y., Springer, https://doi.org/, vol 25. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-4426-8_9, 1986.

Haynes, W. M., Bell, I. H., Berger, L. I., Bradley, P. E., Bruno, T. J., Carraher, C. E., Cheng, J.-P., Chirico, R. D., Cibulka, I., Cramer, C. J., Vladimir, D., Frenkel, M., Fuhr, J. R., Gloldberg, R. N., ..., and Zwilliner, D.: CRC Handbook of chemistry and physics, 97th ed., edited by: Haynes, W. M., Taylor & Francis, 2017.

He, Y., Trumbore, S. E., Torn, M. S., Harden, J. W., Vaugh, L. J. S., Allison, S. D., and Randerson, J. T.: Radiocarbon constraints imply reduced carbon uptake by soils during

the 21st century, Science (80-.)., 353, 1419–1424, 2016.

Heimann, M. and Reichstein, M.: Terrestrial ecosystem carbon dynamics and climate feedbacks, Nature, 451, 289–292, https://doi.org/10.1038/nature06591, 2008.

Helmy, A. K., Ferreiro, E. A., and De Bussetti, S. G.: Cation exchange capacity and condition of zero charge of hydroxy-Al montmorillonite, Clays Clay Miner., 42, 444–450, https://doi.org/10.1346/CCMN.1994.0420410, 1994.

Hemingway, J. D., Rothman, D. H., Grant, K. E., Rosengard, S. Z., Eglinton, T. I., Derry, L. A., and Galy, V. V.: Mineral protection regulates long-term global preservation of natural organic carbon, Nature, 570, 228–231, https://doi.org/10.1038/s41586-019-1280-6, 2019.

Henin, S. and Dupuis, M.: Essai de bilan de la matière organique du sol, in: Ann. Agron. 11, Dunod (impr. de Chaix), Paris, 17–29, 1945.

Hénin, S. and Dupuis, M.: Essai de bilan de la matière organique du sol, Ann. Agron., 1, 19–29, 1945.

Herbst, M., Welp, G., Macdonald, A., Jate, M., Hädicke, A., Scherer, H., Gaiser, T., Herrmann, F., Amelung, W., and Vanderborght, J.: Correspondence of measured soil carbon fractions and RothC pools for equilibrium and non-equilibrium states, Geoderma, 314, 37–46, https://doi.org/10.1016/j.geoderma.2017.10.047, 2018.

Hetényi, M., Nyilas, T., and Tóth, T. M.: Stepwise Rock-Eval pyrolysis as a tool for typing heterogeneous organic matter in soils, J. Anal. Appl. Pyrolysis, 74, 45–54, https://doi.org/10.1016/j.jaap.2004.11.012, 2005.

The Flame Ionization Detector: https://www.chromatographyonline.com/view/flame-ionization-detector.

Houghton, R. A.: Land-use change and the carbon cycle, Glob. Chang. Biol., 1, 275–287, https://doi.org/10.1111/j.1365-2486.1995.tb00026.x, 1995.

Houghton, R. A.: Balancing the Global Carbon Budget, Annu. Rev. Earth Planet. Sci., 35, 313–347, https://doi.org/10.1146/annurev.earth.35.031306.140057, 2007.

Hu, C., Hu, H., Song, M., Tan, J., Huang, G., and Zuo, J.: Preparation, characterization, and Cd(II) sorption of/on cysteine-montmorillonite composites synthesized at various pH, Environ. Sci. Pollut. Res., 27, 10599–10606, https://doi.org/10.1007/s11356-019-07550-4, 2020.

Hudson, B. D.: Soil organic matter and available water capacity, J. Soil Water Conserv., 49, 189–194, 1994.

Huizinga, B. J., Tannenbaum, E., and Kaplan, I. R.: The role of minerals in the thermal alteration of organic matter-III. Generation of bitumen in laboratory experiments, Org. Geochem., 11, 591–604, https://doi.org/10.1016/0146-6380(87)90012-X, 1987.

Ijagbemi, C. O., Baek, M., and Kim, D.: Montmorillonite surface properties and sorption characteristics for heavy metal removal from aqueous solutions, 166, 538–546, https://doi.org/10.1016/j.jhazmat.2008.11.085, 2009.

IN-SYLVA France: https://www6.inrae.fr/in-sylva-france_eng/, last access: 20 May 2021.

IPBES: Summary for policymakers of the assessment report on land degradation and restoration of the Intergovernmental Science- Policy Platform on Biodiversity and Ecosystem Services., edited by: Scholes, R., Montanarella, L., Brainich, A., Barger, N., ten Brink, B., Cantele, M., Erasmus, B., Fisher, J., Gardner, T., Holland, T. G., Kohler, F., Kotiaho,

J. S., Von Maltitz, G., Nangendo, G., Pandit, R., Parrotta, J., Potts, M. D., Prince, S., Sankaran, M., and Willemen, L., 1–54 pp., 2018.

IPCC: Climate change 1995. The science of climate change, edited by: Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kettenberg, A., and Maskell, K., https://doi.org/10.1007/978-3-662-03925-0_1, 1995.

IPCC: 2006 IPCC Guidelines for National Greenhouse Gas Inventories – A primer, Prepared by the National Greenhouse Gas Inventories Programme, edited by: Eggleston, H. S., Miwa, K., Srivastava, N., and Tanabe, K., IPCC, Hayama,, Japan, 2006.

IPCC: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, edited by: Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J. B. R., Chen, Y., Zhou, X., Gomis, M. I., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T., Intergovernmental Panel on Climate Change, 2018.

IPCC: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems, 2019.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., P. Zhai, A., Pirani, S. L., Connors, C., Péan, S., Berger, N., Caud, Y., Chen, L., Goldfarb, M. I., Gomis, M., Huang, K., Leitzell, E., Lonnoy, J. B. R., Matthews, T. K., Maycock, T., Waterfield, O., Yelekçi, R. Y., and Zhou, B., Cambridge University Press, 2021.

WoSIS Soil Profile Database: https://www.isric.org/explore/wosis, last access: 20 May 2021.

Janssen, B. H.: A simple method for calculating decomposition and accumulation of "young" soil organic matter, Plant Soil, 76, 297–304, https://doi.org/10.1007/BF02205588, 1984.

Jenkinson, D. S.: The Turnover of Organic Carbon and Nitrogen in Soil, Phil. Trans. R. Soc. L. B, 329, 361–368, https://doi.org/10.1098/rstb.1990.0177, 1990.

Jenkinson, D. S. and Coleman, K.: The turnover of organic carbon in subsoils. Part 2. Modelling carbon turnover, Eur. J. Soil Sci., 59, 400–413, https://doi.org/10.1111/j.1365-2389.2008.01026.x, 2008.

Jenkinson, D. S. and Johnston, A. E.: Soil Organic Matter in the Hoosfield Continuous Barley Experiment, Rothamsted Exp. Stn. Reports 1976, 2, 87–101, 1977.

Jenkinson, D. S. and Rayner, J. H.: The turnover of soil organic matter in some of the Rothamsted classical experiments, Soil Sci., 123, 298–305, 1977.

Jenkinson, D. S., Hart, P. B., Rayner, J. H., and Parry, L. C.: Modelling the turnover of organic matter in long-term experiments at Rothamsted, INTECOL Bull., 1–8, 1987.

Jenkinson, D. S., Poulton, P. R., and Bryant, C.: The turnover of organic carbon in subsoils. Part 1. Natural and bomb radiocarbon in soil profiles from the Rothamsted long-term field experiments, Eur. J. Soil Sci., 59, 391–399, https://doi.org/10.1111/j.1365-2389.2008.01025.x, 2008.

Jenny, H.: Factors of Soil Formation: A System of Quantitative Pedology, McGraw- Hill Book Company, New York, 336 pp., https://doi.org/10.2307/211491, 1941.

Jobbagy, E. G. and Jackson, R. B.: The vertical distribution of Soil Organic Carbon and its relation to climate and vegetation, Ecol. Appl., 10, 423–436, 2000.

Jolivet, C., Boulonne, L., and Ratié, C.: Manuel du Réseau de Mesures de la Qualité des Sols (RMQS), 10, 187, 2006.

Jonard, P. M., Nicolas, M., Coomes, D., Caignet, I., and Saenger, A.: Les sols des forêts du réseau Renecofor séquestrent le carbone, 67–72, 2019.

JRC: Soil threats in Europe, edited by: Stolte, J., Tesfai, M., Øygarden, L., Kværnø, S., Keizer, J., Verheijen, F., Panagos, P., Ballabio, C., and Hessel, R., European Union, https://doi.org/10.2788/488054, 2016.

Just, C., Poeplau, C., Don, A., van Wesemael, B., Kögel-Knabner, I., and Wiesmeier, M.: A Simple Approach to Isolate Slow and Fast Cycling Organic Carbon Fractions in Central European Soils—Importance of Dispersion Method, Front. Soil Sci., 1, https://doi.org/10.3389/fsoil.2021.692583, 2021.

Justes, E., Mary, B., and Nicolardot, B.: Quantifying and modelling C and N mineralization kinetics of catch crop residues in soil: Parameterization of the residue decomposition module of STICS model for mature and non mature residues, Plant Soil, 325, 171–185, https://doi.org/10.1007/s11104-009-9966-4, 2009.

Kaiser, K. and Guggenberger, G.: Mineral surfaces and soil organic matter, Eur. J. Soil Sci., 54, 219–236, 2003.

Kanari, E., Barré, P., Baudin, F., Berthelot, A., Bouton, N., Gosselin, F., Soucémarianadin, L., Savignac, F., and Cécillon, L.: Predicting Rock-Eval® thermal analysis parameters of a soil layer based on samples from its sublayers; an experimental study on forest soils, Org. Geochem., 160, https://doi.org/10.1016/j.orggeochem.2021.104289, 2021.

Kanari, E., Cécillon, L., Baudin, F., Clivot, H., Ferchaud, F., Houot, S., Levavasseur, F., Mary, B., Soucémarianadin, L., Chenu, C., and Barré, P.: A robust initialization method for accurate soil organic carbon simulations, 19, 375–387, https://doi.org/10.5194/bg-19-375-2022, 2022.

Khedim, N., Cécillon, L., Poulenard, J., Barré, P., Baudin, F., Marta, S., Rabatel, A., Dentant, C., Fabien, S. C., Ludovic, A., Roberto, G., Franzetti, A., Sergio, R., Marco, A., and Caccianiga, S.: Topsoil organic matter build-up in glacier forelands around the world, 1–16, https://doi.org/10.1111/gcb.15496, 2021.

Kleber, M., Mertz, C., Zikeli, S., Knicker, H., and Jahn, R.: Changes in surface reactivity and organic matter composition of clay subfractions with duration of fertilizer deprivation, Eur. J. Soil Sci., 55, 381–391, https://doi.org/10.1111/j.1365-2389.2004.00610.x, 2004.

Kleber, M., Sollins, P., and Sutton, R.: A conceptual model of organo-mineral interactions in soils: Self-assembly of organic molecular fragments into zonal structures on mineral surfaces, Biogeochemistry, 85, 9–24, https://doi.org/10.1007/s10533-007-9103-5, 2007.

Kleber, M., Eusterhues, K., Keiluweit, M., Mikutta, C., Mikutta, R., and Nico, P. S.: Mineral – Organic Associations : Formation , Properties , and Relevance in Soil Environments, Elsevier Ltd, https://doi.org/10.1016/bs.agron.2014.10.005, 2015.

Kögel-Knabner, I.: The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter: Fourteen years on, Soil Biol. Biochem., 105, A3–A8, https://doi.org/10.1016/j.soilbio.2016.08.011, 2017.

Kopac, T., Bozgeyik, K., and Yener, J.: Effect of pH and temperature on the adsorption of

bovine serum albumin onto titanium dioxide, Colloids Surfaces A Physicochem. Eng. Asp., 322, 19–28, https://doi.org/10.1016/j.colsurfa.2008.02.010, 2008.

Kristiansen, S. M., Hansen, E. M., Jensen, L. S., and Christensen, B. T.: Natural 13C abundance and carbon storage in Danish soils under continuous silage maize, Eur. J. Agron., 22, 107–117, https://doi.org/10.1016/j.eja.2004.01.002, 2005.

Kudelski, A.: Influence of electrostatically bound proteins on the structure of linkage monolayers: Adsorption of bovine serum albumin on silver and gold substrates coated with monolayers of 2-mercaptoethanesulphonate, Vib. Spectrosc., 33, 197–204, https://doi.org/10.1016/j.vibspec.2003.09.003, 2003.

Lafargue, E., Espitalié, J., Marquis, F., and Pillot, D.: Rock-Eval 6 Applications in hydocarbon exploration, production and in soil contamination stdies, Rev. l'Institut Français du Pétrole, 53, 421–437, 1998.

Lal, R.: World cropland soils as a source or sink for atmospheric carbon, Adv. Agron., 71, 145–191, https://doi.org/10.1016/s0065-2113(01)71014-0, 2001.

Lal, R.: Soil carbon sequestration impacts on global climate change and food security, Science (80-.)., 304, 1623–1627, https://doi.org/10.1126/science.1097396, 2004a.

Lal, R.: Soil carbon sequestration to mitigate climate change, Geoderma, 123, 1–22, https://doi.org/10.1016/j.geoderma.2004.01.032, 2004b.

Lal, R.: Intensive Agriculture and the Soil Carbon Pool, J. Crop Improv., 27, 735–751, https://doi.org/10.1080/15427528.2013.845053, 2013.

Lavallee, J. M., Soong, J. L., and Cotrufo, M. F.: Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century, Glob. Chang. Biol., 26, 261–273, https://doi.org/10.1111/gcb.14859, 2019.

Lee, J., Viscarra Rossel, R. A., Zhang, M., Luo, Z., and Wang, Y. P.: Assessing the response of soil carbon in Australia to changing inputs and climate using a consistent modelling framework, 18, 5185–5202, https://doi.org/10.5194/bg-18-5185-2021, 2021.

Lehmann, J. and Kleber, M.: The contentious nature of soil organic matter, Nature, 528, 60–68, https://doi.org/10.1038/nature16069, 2015.

Lehmann, J., Hansel, C. M., Kaiser, C., Kleber, M., Maher, K., Manzoni, S., Nunan, N., Reichstein, M., Schimel, J. P., Torn, M. S., Wieder, W. R., and Kögel-Knabner, I.: Persistence of soil organic carbon caused by functional complexity, Nat. Geosci., 13, 529–534, https://doi.org/10.1038/s41561-020-0612-3, 2020.

Leifeld, J., Zimmermann, M., Fuhrer, J., and Conen, F.: Storage and turnover of carbon in grassland soils along an elevation gradient in the Swiss Alps, Glob. Chang. Biol., 15, 668–679, https://doi.org/10.1111/j.1365-2486.2008.01782.x, 2009a.

Leifeld, J., Reiser, R., and Oberholzer, H. R.: Consequences of conventional versus organic farming on soil carbon: Results from a 27-year field experiment, Agron. J., 101, 1204–1218, https://doi.org/10.2134/agronj2009.0002, 2009b.

Levavasseur, F., Mary, B., Christensen, B. T., Duparque, A., Ferchaud, F., Kätterer, T., Lagrange, H., Montenach, D., Resseguier, C., and Houot, S.: The simple AMG model accurately simulates organic carbon storage in soils after repeated application of exogenous organic matter, Nutr. Cycl. Agroecosystems, 117, 215–229, https://doi.org/10.1007/s10705-020-10065-x, 2020.

Liu, X., Lee Burras, C., Kravchenko, Y. S., Duran, A., Huffman, T., Morras, H., Studdert, G.,

Zhang, X., Cruse, R. M., and Yuan, X.: Overview of Mollisols in the world: Distribution, land use and management, Can. J. Soil Sci., 92, 383–402, https://doi.org/10.4141/CJSS2010-058, 2012.

Lubet, E., Plénet, D., and Juste, C.: Effet à long terme de la monoculture sur le rendement en grain du maïs (Zea mays L) en conditions non irriguées., 13, 673–683, https://doi.org/10.1051/agro:19930801, 1993.

Lugato, E., Lavallee, J. M., Haddix, M. L., Panagos, P., and Cotrufo, M. F.: Different climate sensitivity of particulate and mineral-associated soil organic matter, Nat. Geosci., 14, 295–300, https://doi.org/10.1038/s41561-021-00744-x, 2021.

Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J. W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, Robert, B., Jones, C. D., Lara, M. J., Liang, J., McGuire, A. D., Parton, W., Peng, C., Randerson, J. T., Salazar, A., Sierra, C. A., Smith, M. J., Tian, H., Todd-Brown, K. E. O., Torn, M., Van Groenigen, K. J., Wang, Y. P., West, T. O., Wei, Y., Wieder, W. R., Xu, X., Xu, X., and Zhou, T.: Towards more realistic projections of soil carbon dynamics by Earth system models, Global Biogeochem. Cycles, 30, 40–56, https://doi.org/10.1002/2015GB005239, 2016.

Luo, Z. and Viscarra-Rossel, R.: Soil properties override climate controls on global soil organic carbon stocks, Biogeosciences Discuss., 1–24, https://doi.org/10.5194/bg-2020-298, 2020.

Luo, Z., Wang, E., Fillery, I. R. P., Macdonald, L. M., Huth, N., and Baldock, J.: Modelling soil carbon and nitrogen dynamics using measurable and conceptual soil organic matter pools in APSIM, Agric. Ecosyst. Environ., 186, 94–104, https://doi.org/10.1016/j.agee.2014.01.019, 2014.

Luo, Z., Wang, G., and Wang, E.: Global subsoil organic carbon turnover times dominantly controlled by soil properties rather than climate, Nat. Commun., 10, 1–10, https://doi.org/10.1038/s41467-019-11597-9, 2019.

Lutfalla, S., Abiven, S., Barré, P., Wiedemeier, D. B., Christensen, B. T., Houot, S., Kätterer, T., Macdonald, A. J., van Oort, F., and Chenu, C.: Pyrogenic Carbon Lacks Long-Term Persistence in Temperate Arable Soils, Front. Earth Sci., 5, https://doi.org/10.3389/feart.2017.00096, 2017.

von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Matzner, E., Guggenberger, G., Marschner, B., and Flessa, H.: Stabilization of organic matter in temperate soils: Mechanisms and their relevance under different soil conditions - A review, Eur. J. Soil Sci., 57, 426–445, https://doi.org/10.1111/j.1365-2389.2006.00809.x, 2006.

von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E., and Marschner, B.: SOM fractionation methods: Relevance to functional pools and to stabilization mechanisms, Soil Biol. Biochem., 39, 2183–2207, https://doi.org/10.1016/j.soilbio.2007.03.007, 2007.

Ma, X., Zheng, G., Sajjad, W., Xu, W., Fan, Q., Zheng, J., and Xia, Y.: Influence of minerals and iron on natural gases generation during pyrolysis of type-III kerogen, Mar. Pet. Geol., 89, 216–224, https://doi.org/10.1016/j.marpetgeo.2017.01.012, 2018.

Machet, J. M., Dubrulle, P., Damay, N., Duval, R., Julien, J. L., and Recous, S.: A dynamic decision-making tool for calculating the optimal rates of N application for 40 annual crops while minimising the residual level of mineral N at harvest, 7,

https://doi.org/10.3390/agronomy7040073, 2017.

Maduskar, S., Teixeira, A. R., Paulsen, A. D., Krumm, C., Mountziaris, T. J., and Dauenhauer, P. J.: Lab on a Chip of complex mixtures †, 440–447, https://doi.org/10.1039/c4lc01180e, 2015.

Manzoni, S. and Porporato, A.: Soil carbon and nitrogen mineralization: Theory and models across scales, Soil Biol. Biochem., 41, 1355–1379, https://doi.org/10.1016/j.soilbio.2009.02.031, 2009.

Mao, Z., Derrien, D., Didion, M., Liski, J., Eglin, T., Nicolas, M., Jonard, M., and Saint-André, L.: Modeling soil organic carbon dynamics in temperate forests with Yasso07, 16, 1955–1973, https://doi.org/10.5194/bg-16-1955-2019, 2019.

Martin, M., Dimassi, B., Millet, F., Picaud, C., Bounoua, E.-M., Bardy, M., Bispo, A., Boulonne, L., Bouthier, A., Duparque, A., Eglin, T., Guenet, B., Huard, F., Mary, B., Mathias, E., Mignolet, C., Robert, C., Saby, N., Sagot, S., Schott, C., and Toutain, B. R.: Méthodes de comptabilisation du stockage de carbone organique des sols sous l'effet des pratiques culturales (CSopra)., 2019.

Mary, B. and Guérif, J.: Intérêts et limites des modèles de prévision de l'évolution des matières organiques et de l'azote dans le sol, CIRAD Recues - Cah. Agric., 3, 247–257, 1994.

Mary, B. and Wylleman, R.: Characterization and modelling of organic C and N in soil in different cropping systems, in: Proceedings of the 11th Nitrogen Workshop, 251–252, 2001.

Mary, B., Beaudoin, N., Justes, E., and Machet, J. M.: Calculation of nitrogen mineralization and leaching in fallow soil using a simple dynamic model, Eur. J. Soil Sci., 50, 549–566, https://doi.org/10.1046/j.1365-2389.1999.00264.x, 1999.

Mary, B., Clivot, H., Blaszczyk, N., Labreuche, J., and Ferchaud, F.: Soil carbon storage and mineralization rates are affected by carbon inputs rather than physical disturbance: Evidence from a 47-year tillage experiment, Agric. Ecosyst. Environ., 299, 106972, https://doi.org/10.1016/j.agee.2020.106972, 2020.

Mcgill, W. B.: Review and Classification of Ten Soil Organic Matter (SOM) Models, in: Evaluation of Soil Organic Matter Models, edited by: Powlson, D. S., Smith, P., and Smith, J. U., Springer Berlin Heidelberg, 111–132, 1996.

Mcgill, W. B., Hunt, W. H., Woodmansee, R. G., and Reuss, J. O.: PHOENIX , A model of the dynamics of Carbon and Nitrogen in Grassland Soils, Terr. Nitrogen Cycles, 33, 49–115, 1981.

Messiga, A. J., Ziadi, N., Plénet, D., Parent, L. E., and Morel, C.: Long-term changes in soil phosphorus status related to P budgets under maize monoculture and mineral P fertilization, Soil Use Manag., 26, 354–364, https://doi.org/10.1111/j.1475-2743.2010.00287.x, 2010.

Mikutta, R., Kleber, M., Torn, M. S., and Jahn, R.: Stabilization of soil organic matter: Association with minerals or chemical recalcitrance?, Biogeochemistry, 77, 25–56, https://doi.org/10.1007/s10533-005-0712-6, 2006.

Mikutta, R., Mikutta, C., Kalbitz, K., Scheel, T., Kaiser, K., and Jahn, R.: Biodegradation of forest floor organic matter bound to minerals via different binding mechanisms, Geochim. Cosmochim. Acta, 71, 2569–2590, https://doi.org/10.1016/j.gca.2007.03.002, 2007.

Mikutta, R., Turner, S., Schippers, A., Gentsch, N., Meyer-Stüve, S., Condron, L. M., Peltzer,

D. A., Richardson, S. J., Eger, A., Hempel, G., Kaiser, K., Klotzbücher, T., and Guggenberger, G.: Microbial and abiotic controls on mineral-associated organic matter in soil profiles along an ecosystem gradient, Sci. Rep., 9, 1–9, https://doi.org/10.1038/s41598-019-46501-4, 2019.

Mills, R. T. E., Tipping, E., Bryant, C. L., and Emmett, B. A.: Long-term organic carbon turnover rates in natural and semi-natural topsoils, Biogeochemistry, 118, 257–272, https://doi.org/10.1007/s10533-013-9928-z, 2014.

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z. S., Cheng, K., Das, B. S., Field, D. J., Gimona, A., Hedley, C. B., Hong, S. Y., Mandal, B., Marchant, B. P., Martin, M., McConkey, B. G., Mulder, V. L., O'Rourke, S., Richer-de-Forges, A. C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C. C., Vågen, T. G., van Wesemael, B., and Winowiecki, L.: Soil carbon 4 per mille, Geoderma, 292, 59–86, https://doi.org/10.1016/j.geoderma.2017.01.002, 2017.

Montanarella, L.: The Global Soil Partnership, in: IOP Conference Series: Earth and Environmental Science, https://doi.org/10.1088/1755-1315/25/1/012001, 2015.

Morel, C., Ziadi, N., Messiga, A., Bélanger, G., Denoroy, P., Jeangros, B., Jouany, C., Fardeau, J. C., Mollier, A., Parent, L. E., Proix, N., Rabeharisoa, L., and Sinaj, S.: Modeling of phosphorus dynamics in contrasting agroecosystems using long-term field experiments, Can. J. Soil Sci., 94, 377–387, https://doi.org/10.4141/CJSS2013-024, 2014.

Nave, L., Johnson, K., Ingen, C. van, Agarwal, D., Humphrey, M., and Beekwilder, N.: International Soil Carbon Network (ISCN) Database V3-1, https://doi.org/10.17040/ISCN/1305039, 2015.

Nemo, Klumpp, K., Coleman, K., Dondini, M., Goulding, K., Hastings, A., Jones, M. B., Leifeld, J., Osborne, B., Saunders, M., Scott, T., Teh, Y. A., and Smith, P.: Soil Organic Carbon (SOC) Equilibrium and Model Initialisation Methods: an Application to the Rothamsted Carbon (RothC) Model, Environ. Model. Assess., 22, 215–229, https://doi.org/10.1007/s10666-016-9536-0, 2016.

Newcomb, C. J., Qafoku, N. P., Grate, J. W., Bailey, V. L., and De Yoreo, J. J.: Developing a molecular picture of soil organic matter–mineral interactions by quantifying organo–mineral binding, Nat. Commun., 8, 396, https://doi.org/10.1038/s41467-017-00407-9, 2017.

Nicolas, M., Cecchini, S., Croisé, L., Lavalley, C., and Macé, S.: RENECOFOR - Bilan technique de l'année 2017, 32 pp., 2018.

Noirot-Cosson, P. E., Vaudour, E., Gilliot, J. M., Gabrielle, B., and Houot, S.: Modelling the long-term effect of urban waste compost applications on carbon and nitrogen dynamics in temperate cropland, Soil Biol. Biochem., 94, 138–153, https://doi.org/10.1016/j.soilbio.2015.11.014, 2016.

Nowak, B. and Marliac, G.: Optimization of carbon stock models to local conditions using farmers' soil tests: A case study with AMGv2 for a cereal plain in central France, Soil Use Manag., 36, 633–645, https://doi.org/10.1111/sum.12608, 2020.

Nunan, N., Schmidt, H., and Raynaud, X.: The ecology of heterogeneity: Soil bacterial communities and C dynamics, Philos. Trans. R. Soc. B Biol. Sci., 375, https://doi.org/10.1098/rstb.2019.0249, 2020.

Oades, J. M. and Waters, A. G.: Aggregate hierarchy in soils, Aust. J. Soil Res., 29, 815–825,

https://doi.org/10.1071/SR9910815, 1991.

Oberholzer, H. R., Leifeld, J., and Mayer, J.: Changes in soil carbon and crop yield over 60 years in the Zurich Organic Fertilization Experiment, following land-use change from grassland to cropland, J. Plant Nutr. Soil Sci., 177, 696–704, https://doi.org/10.1002/jpln.201300385, 2014.

Obriot, F.: Epandage de produits résiduaires organiques et fonctionnement biologique des sols : De la quantification des impacts sur les cycles carbone et azote à l'évaluation multicritère de la pratique à l'échelle de la parcelle., AgroParisTech, 194–195 pp., 2016.

Olsson, L., Barbosa, H., Bhadwal, S., Cowie, A., Delusca, K., Flores-Renteria, D., Hermans, K., Jobbagy, E., Kurz, W., Li, D., Sonwa, D. J., and Stringer, L.: Land Degredation, Clim. Chang. L. an IPCC Spec. Rep. Clim. Chang. Desertif. L. Degrad. Sustain. L. Manag. food Secur. Greenh. gas fluxes Terr. Ecosyst., 345–436, 2019.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., and Fernández-Ugalde, O.: LUCAS Soil, the largest expandable soil dataset for Europe: a review, Eur. J. Soil Sci., 69, 140–153, https://doi.org/10.1111/ejss.12499, 2018.

Otto, S. A., Kadin, M., Casini, M., Torres, M. A., and Blenckner, T.: A quantitative framework for selecting and validating food web indicators, Ecol. Indic., 84, 619–631, https://doi.org/10.1016/j.ecolind.2017.05.045, 2018.

Pan, C., Jiang, L., Liu, J., Zhang, S., and Zhu, G.: The effects of calcite and montmorillonite on oil cracking in confined pyrolysis experiments, Org. Geochem., 41, 611–626, https://doi.org/10.1016/j.orggeochem.2010.04.011, 2010.

Panagos, P., Ballabio, C., Himics, M., Scarpa, S., Matthews, F., Bogonos, M., Poesen, J., and Borrelli, P.: Projections of soil loss by water erosion in Europe by 2050, Environ. Sci. Policy, 124, 380–392, https://doi.org/10.1016/j.envsci.2021.07.012, 2021.

Parton, W. J., Schimel, D. S., Cole, C. V., and Ojima, D. S.: Analysis of Factors Controlling Soil Organic Matter Levels in Great Plains Grasslands, Soil Sci. Soc. Am. J., 51, 1173–1179, https://doi.org/10.2136/sssaj1987.03615995005100050015x, 1987.

Paul, E. A., Collins, H. P., and Leavitt, S. W.: Dynamics of resistant soil carbon of midwestern agricultural soils measured by naturally occurring 14C abundance, Geoderma, 104, 239–256, https://doi.org/10.1016/S0016-7061(01)00083-0, 2001.

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G. P., and Smith, P.: Climate-smart soils, Nature, 532, 49–57, https://doi.org/10.1038/nature17174, 2016.

Pellerin, S., Bamière, L., Savini, I., and Réchauchère, O.: Stocker du carbone dans les sols français, Stock. du carbone dans les sols français, https://doi.org/10.35690/978-2-7592-3149-2, 2019.

Pellerin, S., Bamière, L., Launay, C., Martin, R., Schiavo, M., Angers, D., Augusto, L., Balesdent, J., Basile-Doelsch, I., Bellassen, V., Cardinael, R., Cécillon, L., Ceschia, E., Chenu, C., Constantin, J., Darroussin, J., Delacote, P., Delame, N., Gastal, F., Gilbert, D., Graux, A.-I., Guenet, B., Houot, S., Klumpp, K., Letort, E., Litrico, I., Martin, M., Menasseri, S., Mézière, D., Morvan, T., Mosnier, C., Roger-Estrade, J., Saint-André, L., Sierra, J., Thérond, O., Viaud, V., Grateau, R., Le Perchec, S., Savini, I., and Réchauchère, O.: Stocker du carbone dans les sols français, Quel potentiel au regard de l'objectif 4 pour 1000 et à quel coût ? Rapport scientifique de l'étude, INRA (France), 540 p., 540, 2020.

Perdue, E. M. and Koprivnjak, J.-F.: Using the C / N ratio to estimate terrigenous inputs of

organic matter to aquatic environments, Estuar. Coast. Shelf Sci., 73, 65–72, https://doi.org/10.1016/j.ecss.2006.12.021, 2007.

Peters, G. P., Andrew, R. M., Canadell, J. G., Friedlingstein, P., Jackson, R. B., Korsbakken, J. I., Le Quéré, C., and Peregon, A.: Carbon dioxide emissions continue to grow amidst slowly emerging climate policies, Nat. Clim. Chang., 10, 3–6, https://doi.org/10.1038/s41558-019-0659-6, 2020.

Phan, H. T. M., Bartelt-Hunt, S., Rodenhausen, K. B., Schubert, M., and Bartz, J. C.: Investigation of Bovine Serum Albumin (BSA) Attachment onto Self-Assembled Monolayers (SAMs) Using Combinatorial Quartz Crystal Microbalance with Dissipation (QCM-D) and Spectroscopic Ellipsometry (SE), PLoS One, 10, e0141282, https://doi.org/10.1371/journal.pone.0141282, 2015.

Plante, A. F., Beaupré, S. R., Roberts, M. L., and Baisden, T.: Distribution of Radiocarbon Ages in Soil Organic Matter by Thermal Fractionation, Radiocarbon, 55, 1077–1083, https://doi.org/10.1017/s0033822200058215, 2013.

Poeplau, C., Don, A., Vesterdal, L., Leifeld, J., Van Wesemael, B., Schumacher, J., and Gensior, A.: Temporal dynamics of soil organic carbon after land-use change in the temperate zone - carbon response functions as a model approach, Glob. Chang. Biol., 17, 2415–2427, https://doi.org/10.1111/j.1365-2486.2011.02408.x, 2011.

Poeplau, C., Don, A., Dondini, M., Leifeld, J., Nemo, R., Schumacher, J., Senapati, N., and Wiesmeier, M.: Reproducibility of a soil organic carbon fractionation method to derive RothC carbon pools, Eur. J. Soil Sci., 64, 735–746, https://doi.org/10.1111/ejss.12088, 2013.

Poeplau, C., Don, A., Six, J., Kaiser, M., Benbi, D., Chenu, C., Cotrufo, M. F., Derrien, D., Gioacchini, P., Grand, S., Gregorich, E., Griepentrog, M., Gunina, A., Haddix, M., Kuzyakov, Y., Kühnel, A., Macdonald, L. M., Soong, J., Trigalet, S., Vermeire, M. L., Rovira, P., van Wesemael, B., Wiesmeier, M., Yeasmin, S., Yevdokimov, I., and Nieder, R.: Isolating organic carbon fractions with varying turnover rates in temperate agricultural soils – A comprehensive method comparison, Soil Biol. Biochem., 125, 10–26, https://doi.org/10.1016/j.soilbio.2018.06.025, 2018.

Poeplau, C., Barré, P., Cécillon, L., Baudin, F., and Sigurdsson, B. D.: Changes in the Rock-Eval signature of soil organic carbon upon extreme soil warming and chemical oxidation - A comparison, Geoderma, 337, 181–190, https://doi.org/10.1016/j.geoderma.2018.09.025, 2019.

R Core Team: R: A language and environment for statistical computing, https://www.r-project.org/, 2017.

Rahman, H. M., Kennedy, M., Löhr, S., and Dewhurst, D. N.: Clay-organic association as a control on hydrocarbon generation in shale, Org. Geochem., 105, 42–55, https://doi.org/10.1016/j.orggeochem.2017.01.011, 2017.

Rasmussen, C., Throckmorton, H., Liles, G., Heckman, K., Meding, S., and Horwath, W. R.: Controls on soil organic carbon partitioning and stabilization in the california sierra nevada, Soil Syst., 2, 1–18, https://doi.org/10.3390/soilsystems2030041, 2018.

Rasse, D. P., Mulder, J., Moni, C., and Chenu, C.: Carbon Turnover Kinetics with Depth in a French Loamy Soil, Soil Sci. Soc. Am. J., 70, 2097–2105, https://doi.org/10.2136/sssaj2006.0056, 2006.

Riebeek, H.: The carbon cycle, https://earthobservatory.nasa.gov/features/CarbonCycle, 2011.

Rillig, M. C. and Mummey, D. L.: Mycorrhizas and soil structure, New Phytol., 171, 41–53, https://doi.org/10.1111/j.1469-8137.2006.01750.x, 2006.

Robertson, A. D., Paustian, K., Ogle, S., Wallenstein, M. D., Lugato, E., and Francesca Cotrufo, M.: Unifying soil organic matter formation and persistence frameworks: The MEMS model, 16, 1225–1248, https://doi.org/10.5194/bg-16-1225-2019, 2019.

Rose, S., Khatri-Chhetri, A., Stier, M., Wiese-Rozanova, L., Shelton, S., and Wollenberg, E.: Ambition for soil organic carbon sequestration in the new and updated nationally determined contributions: 2020- 2021. Analysis of agricultural sub-sectors in national climate change strategies., 2021.

Rühlmann, J.: A new approach to estimating the pool of stable organic matter in soil using data from long-term field experiments, Plant Soil, 213, 149–160, https://doi.org/10.1023/a:1004552016182, 1999.

Rumpel, C., Rabia, N., Derenne, S., Quenea, K., Eusterhues, K., Kögel-Knabner, I., and Mariotti, A.: Alteration of soil organic matter following treatment with hydrofluoric acid (HF), Org. Geochem., 37, 1437–1451, 2006.

Rumpel, C., Amiraslani, F., Koutika, L. S., Smith, P., Whitehead, D., and Wollenberg, E.: Put more carbon in soils to meet Paris climate pledges, Nature, 564, 32–34, https://doi.org/10.1038/d41586-018-07587-4, 2018.

Saenger, A., Cécillon, L., Sebag, D., and Brun, J. J.: Soil organic carbon quantity, chemistry and thermal stability in a mountainous landscape: A Rock-Eval pyrolysis survey, Org. Geochem., 54, 101–114, https://doi.org/10.1016/j.orggeochem.2012.10.008, 2013.

Saenger, A., Cécillon, L., Poulenard, J., Bureau, F., De Danieli, S., Gonzalez, J.-M., and Brun, J. J.: Surveying the carbon pools of mountain soils : A comparison of physical fractionation and Rock- Eval pyrolysis, Geoderma, 279–288, https://doi.org/10.1016/j.geoderma.2014.12.001, 2015.

Saffih-Hdadi, K. and Mary, B.: Modeling consequences of straw residues export on soil organic carbon, Soil Biol. Biochem., 40, 594–607, https://doi.org/10.1016/j.soilbio.2007.08.022, 2008.

Salter, R. M. and Green, T. C.: Factors affecting the accumulation and loss of nitrogen and organic carbon in cropped soils, J. Am. Soc. Agron., 1933.

Sanderman, J. and Grandy, A. S.: Ramped thermal analysis for isolating biologically meaningful soil organic matter fractions with distinct residence times, 6, 131–144, https://doi.org/10.5194/soil-6-131-2020, 2020.

Sanderman, J., Maddern, T., and Baldock, J.: Similar composition but differential stability of mineral retained organic matter across four classes of clay minerals, Biogeochemistry, 121, 409–424, https://doi.org/10.1007/s10533-014-0009-8, 2014.

Sanderman, J., Hengl, T., and Fiske, G. J.: Soil carbon debt of 12,000 years of human land use, 115, 9575–9580, https://doi.org/10.1073/pnas.1800925115, 2017.

Sanderman, J., Baldock, J. A., Dangal, S. R. S., Ludwig, S., Potter, S., Rivard, C., and Savage, K.: Soil organic carbon fractions in the Great Plains of the United States: an application of mid-infrared spectroscopy, Biogeochemistry, 1, https://doi.org/10.1007/s10533-021-00755-1, 2021.

Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D. A. C., Nannipieri, P., Rasse,

D. P., Weiner, S., and Trumbore, S. E.: Persistence of soil organic matter as an ecosystem property, Nature, 478, 49–56, https://doi.org/10.1038/nature10386, 2011.

Schrumpf, M., Schulze, E. D., Kaiser, K., and Schumacher, J.: How accurately can soil organic carbon stocks and stock changes be quantified by soil inventories?, 8, 1193–1212, https://doi.org/10.5194/bg-8-1193-2011, 2011.

Schulten, H. R. and Leinweber, P.: New insights into organic-mineral particles: Composition, properties and models of molecular structure, Biol. Fertil. Soils, 30, 399–432, https://doi.org/10.1007/s003740050020, 2000.

Sebag, D., Verrecchia, E. P., Cécillon, L., Adatte, T., Albrecht, R., Aubert, M., Bureau, F., Cailleau, G., Copard, Y., Decaens, T., Disnar, J. R., Hetényi, M., Nyilas, T., and Trombino, L.: Dynamics of soil organic matter based on new Rock-Eval indices, Geoderma, 284, 185–203, https://doi.org/10.1016/j.geoderma.2016.08.025, 2016.

Servagent-Noinville, S., Revault, M., Quiquampoix, H., and Baron, M. H.: Conformational changes of bovine serum albumin induced by adsorption on different clay surfaces: FTIR analysis, J. Colloid Interface Sci., 221, 273–283, https://doi.org/10.1006/jcis.1999.6576, 2000.

Shi, Z., Crowell, S., Luo, Y., and Moore, B.: Model structures amplify uncertainty in predicted soil carbon responses to climate change, Nat. Commun., 9, 1–11, https://doi.org/10.1038/s41467-018-04526-9, 2018.

Shi, Z., Allison, S. D., He, Y., Levine, P. A., Hoyt, A. M., Beem-Miller, J., Zhu, Q., Wieder, W. R., Trumbore, S., and Randerson, J. T.: The age distribution of global soil carbon inferred from radiocarbon measurements, Nat. Geosci., 13, 555–559, https://doi.org/10.1038/s41561-020-0596-z, 2020.

Sierra, C. A., Müller, M., and Trumbore, S. E.: Modeling radiocarbon dynamics in soils: SoilR version 1.1, Geosci. Model Dev., 7, 1919–1931, https://doi.org/10.5194/gmd-7-1919-2014, 2014.

Sierra, C. A., Müller, M., Metzler, H., Manzoni, S., and Trumbore, S. E.: The muddle of ages, turnover, transit, and residence times in the carbon cycle, Glob. Chang. Biol., 23, 1763–1773, https://doi.org/10.1111/gcb.13556, 2017.

Sigma: ALBUMIN, BOVINE - Product Information. A4919 03/21/97 - MAC, 2021.

SigmaAldrich: D-(+)-Glucose monohydrate. SAFETY DATA SHEET, 2021a.

SigmaAldrich: Starch (from wheat) for biochemistry. SAFETY DATA SHEET, 2021b.

Six, J. and Jastrow, J.: Organic Matter Turnover, in: Encyclopedia of Soil Science, edited by: Lal, R., Marcel Dekker, 936–942, https://doi.org/10.1201/noe0849338304.ch252, 2002.

Six, J., Elliott, E. T., and Paustian, K.: Soil structure and soil organic matter: II. A normalized stability index and the effect of mineralogy, Soil Sci. Soc. Am. J., 64, 1042–1049, 2000.

Six, J., Conant, R. T., Paul, E. A., and Paustian, K.: Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils, Plant Soil, 241, 155–176, https://doi.org/10.1023/A:1016125726789, 2002.

Six, J., Bossuyt, H., Degryze, S., and Denef, K.: A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics, Soil Tillage Res., 79, 7–31, https://doi.org/10.1016/j.still.2004.03.008, 2004.

Skjemstad, J. O., Le Feuvre, R. P., and Prebble, R. E.: Turnover of soil organic matter under

pasture as determined by 13C natural abundance, Aust. J. Soil Res., 28, 267–276, https://doi.org/10.1071/SR9900267, 1990.

Skjemstad, J. O., Spouncer, L. R., Cowie, B., and Swift, R. S.: Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools, Aust. J. Soil Res., 42, 79–88, https://doi.org/10.1071/SR03013, 2004.

Smith, J., Smith, P., and Addiscott, T.: Quantitative methods to evaluate and compare Soil Organic Matter (SOM) Models, in: Evaluation of Soil Organic Matter Models, 181–199, 1996.

Smith, O. L.: An analytical model of the decomposition of soil organic matter, Soil Biol. Biochem., 11, 585–606, 1979.

Smith, P.: How long before a change in soil organic carbon can be detected?, Glob. Chang. Biol., 10, 1878–1883, https://doi.org/10.1111/j.1365-2486.2004.00854.x, 2004.

Smith, P. and Falloon, P. D.: Modelling refractory soil organic matter, Biol. Fertil. Soils, 30, 388–398, https://doi.org/10.1007/s003740050019, 2000.

Smith, P., Andrén, O., Karlsson, T., Perälä, P., Regina, K., Rounsevell, M., and Van Wesemael, B.: Carbon sequestration potential in European croplands has been overestimated, Glob. Chang. Biol., 11, 2153–2163, https://doi.org/10.1111/j.1365-2486.2005.01052.x, 2005.

Smith, P., Chapman, S. J., Scott, W. A., Black, H. I. J., Wattenbach, M., Milne, R., Campbell, C. D., Lilly, A., Ostle, N., Levy, P. E., Lumsdon, D. G., Millard, P., Towers, W., Zaehle, S., and Smith, J. U.: Climate change cannot be entirely responsible for soil carbon loss observed in England and Wales, 1978-2003, Glob. Chang. Biol., 13, 2605–2609, https://doi.org/10.1111/j.1365-2486.2007.01458.x, 2007.

Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., McCarl, B., Ogle, S., O'Mara, F., Rice, C., Scholes, B., Sirotenko, O., Howden, M., McAllister, T., Pan, G., Romanenkov, V., Schneider, U., Towprayoon, S., Wattenbach, M., and Smith, J.: Greenhouse gas mitigation in agriculture, Philos. Trans. R. Soc. B Biol. Sci., 363, 789–813, https://doi.org/10.1098/rstb.2007.2184, 2008.

Smith, P., Davies, C. A., Ogle, S., Zanchi, G., Bellarby, J., Bird, N., Boddey, R. M., McNamara, N. P., Powlson, D., Cowie, A., van Noordwijk, M., Davis, S. C., Richter, D. D. B., Kryzanowski, L., van Wijk, M. T., Stuart, J., Kirton, A., Eggar, D., Newton-Cross, G., Adhya, T. K., and Braimoh, A. K.: Towards an integrated global framework to assess the impacts of land use and management change on soil carbon: Current capability and future vision, Glob. Chang. Biol., 18, 2089–2101, https://doi.org/10.1111/j.1365-2486.2012.02689.x, 2012.

Smith, P., Lutfalla, S., Riley, W. J., Torn, M. S., Schmidt, M. W. I., and Soussana, J. F.: The changing faces of soil organic matter research, Eur. J. Soil Sci., 69, 23–30, https://doi.org/10.1111/ejss.12500, 2017.

Smith, P., Francois, J., Denis, S., Louis, A., Chenu, C., Rasse, D. P., Batjes, N. H., Egmond, F. Van, Olesen, J. E., Mcneill, S., Kuhnert, M., and Arias, C.: How to measure , report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal, 219–241, https://doi.org/10.1111/gcb.14815, 2020.

Sollins, P., Homann, P., and Caldwell, B. A.: Stabilization and destabilization of soil organic matter: Mechanisms and controls, Geoderma, 74, 65–105, https://doi.org/10.1016/S0016-7061(96)00036-5, 1996.

Soucémarianadin, L., Cécillon, L., Chenu, C., Baudin, F., Nicolas, M., Girardin, C., and Barré,

P.: Is Rock-Eval 6 thermal analysis a good indicator of soil organic carbon lability? – A method-comparison study in forest soils, Soil Biol. Biochem., 117, 108–116, https://doi.org/10.1016/j.soilbio.2017.10.025, 2018.

Soucémarianadin, L., Cécillon, L., Chenu, C., Baudin, F., Nicolas, M., Girardin, C., Delahaie, A., and Barré, P.: Heterogeneity of the chemical composition and thermal stability of particulate organic matter in French forest soils, Geoderma, 342, 65–74, https://doi.org/10.1016/j.geoderma.2019.02.008, 2019.

UK: CHE 103 - Chemistry for Allied Health: https://chem.libretexts.org/@go/page/58228 .

Spaccini, R., Song, X., Cozzolino, V., and Piccolo, A.: Molecular evaluation of soil organic matter characteristics in three agricultural soils by improved off-line thermochemolysis: The effect of hydrofluoric acid demineralisation treatment, Anal. Chim. Acta, 802, 46–55, 2013.

Spangenberg, J. H., Görg, C., Truong, D. T., Tekken, V., Bustamante, J. V., and Settele, J.: Provision of ecosystem services is determined by human agency, not ecosystem functions. Four case studies, Int. J. Biodivers. Sci. Ecosyst. Serv. Manag., 10, 40–53, https://doi.org/10.1080/21513732.2014.884166, 2014.

Spiro, B.: Effects of minerals on Rock Eval pyrolysis of kerogen, J. Therm. Anal., 37, 1513–1522, https://doi.org/10.1007/BF01913484, 1991.

Sposito, G.: The Chemistry of Soils, edited by: Sposito, G., Oxford University Press, 2008.

Sulman, B. N., Moore, J. A. M., Abramoff, R., Averill, C., Kivlin, S., Georgiou, K., Sridhar, B., Hartman, M. D., Wang, G., Wieder, W. R., Bradford, M. A., Luo, Y., Mayes, M. A., Morrison, E., Riley, W. J., Salazar, A., Schimel, J. P., Tang, J., and Classen, A. T.: Multiple models and experiments underscore large uncertainty in soil carbon dynamics, Biogeochemistry, 141, 109–123, https://doi.org/10.1007/s10533-018-0509-z, 2018.

Taghizadeh-Toosi, A., Christensen, B. T., Hutchings, N. J., Vejlin, J., Kätterer, T., Glendining, M., and Olesen, J. E.: C-TOOL: A simple model for simulating whole-profile carbon storage in temperate agricultural soils, Ecol. Modell., 292, 11–25, https://doi.org/10.1016/j.ecolmodel.2014.08.016, 2014.

Taghizadeh-Toosi, A., Cong, W. F., Eriksen, J., Mayer, J., Olesen, J. E., Keel, S. G., Glendining, M., Kätterer, T., and Christensen, B. T.: Visiting dark sides of model simulation of carbon stocks in European temperate agricultural soils: allometric function and model initialization, Plant Soil, 450, 255–272, https://doi.org/10.1007/s11104-020-04500-9, 2020.

Thakur, V. K., Thakur, M. K., Raghavan, P., and Kessler, M. R.: Progress in green polymer composites from lignin for multifunctional applications: A review, ACS Sustain. Chem. Eng., 2, 1072–1092, https://doi.org/10.1021/sc500087z, 2014.

Tiessen, H., Cuevas, E., and Chacon, P.: The role of soil organic matter in sustaining soil fertility, Nature, 371, 783–785, https://doi.org/10.2134/agronj1940.00021962003200080001x, 1994.

Tisdall, J. M. and Oades, J. M.: Organic matter and water-stable aggregates in soils - TISDALL - 2006 - Journal of Soil Science - Wiley Online Library, Eur. J. Soil Sci., 141–163, 1982.

Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A. G., and Allison, S. D.: Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, 10, 1717–1736, https://doi.org/10.5194/bg-10-1717-2013, 2013.

Todd-Brown, K. E. O., Randerson, J. T., Hopkins, F., Arora, V., Hajima, T., Jones, C., Shevliakova, E., Tjiputra, J., Volodin, E., Wu, T., Zhang, Q., and Allison, S. D.: Changes in soil organic carbon storage predicted by Earth system models during the 21st century, 11, 2341–2356, https://doi.org/10.5194/bg-11-2341-2014, 2014.

Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M., and Hendricks, D. M.: Mineral control of soil organic carbon storage and turnover content were measured by horizon down to the depth at which, Nature, 389, 3601–3603, 1997.

Trumbore, S.: Potential responses of soil organic carbon to global environmental change, Proc. Natl. Acad. Sci. U. S. A., 94, 8284–8291, 1997.

Trumbore, S.: Age of soil organic matter and soil respiration: Radiocarbon constraints on belowground C dynamics, Ecol. Appl., 10, 399–411, https://doi.org/10.1890/1051-0761(2000)010[0399:AOSOMA]2.0.CO;2, 2000.

Trumbore, S.: Radiocarbon and soil carbon dynamics, Annu. Rev. Earth Planet. Sci., 37, 47–66, https://doi.org/10.1146/annurev.earth.36.031207.124300, 2009.

UN: Department of Economic and Social Affairs, Population Division. World Population Prospects 2019, 2019.

UN General Assembly: UN General Assembly, Transforming our world : the 2030 Agenda for Sustainable Development, 21 October 2015, A/RES/70/1, 2015.

UNEP: Frontiers 2018/19 Emerging Issues of Environmental Concern, Nairobi, 2019.

UNFCCC: Adoption of the Paris Agreement, in: 21st Conference of the Parties, 1–16, https://doi.org/10.1201/9781351116589-2, 2015.

Vermeulen, S., Bossio, D., Lehmann, J., Luu, P., Paustian, K., Webb, C., Augé, F., Bacudo, I., Baedeker, T., Havemann, T., Jones, C., King, R., Reddy, M., Sunga, I., Von Unger, M., and Warnken, M.: A global agenda for collective action on soil carbon, Nat. Sustain., 2, 2–4, https://doi.org/10.1038/s41893-018-0212-z, 2019.

Vertès, F., Simon, J.-C., Laurent, F., and Besnard, A.: Prairies et qualité de l'eau. Evaluation des risques de lixiviation d'azote et optimisation des pratiques, Fourrag., 192, 423–440, 2007.

Vieira, A. P., Berndt, G., De Souza Junior, I. G., Di Mauro, E., Paesano, A., De Santana, H., Da Costa, A. C. S., Zaia, C. T. B. V., and Zaia, D. A. M.: Adsorption of cysteine on hematite, magnetite and ferrihydrite: FT-IR, Mössbauer, EPR spectroscopy and X-ray diffractometry studies, Amino Acids, 40, 205–214, https://doi.org/10.1007/s00726-010-0635-y, 2011.

Virto, I., Barré, P., and Chenu, C.: Microaggregation and organic matter storage at the silt-size scale, Geoderma, 146, 326–335, https://doi.org/10.1016/j.geoderma.2008.05.021, 2008.

Virto, I., Moni, C., Swanston, C., and Chenu, C.: Turnover of intra- and extra-aggregate organic matter at the silt-size scale, Geoderma, 156, 1–10, https://doi.org/10.1016/j.geoderma.2009.12.028, 2010.

Viscarra Rossel, R. A., Lee, J., Behrens, T., Luo, Z., Baldock, J., and Richards, A.: Continental-scale soil carbon composition and vulnerability modulated by regional environmental controls, Nat. Geosci., 12, 547–552, https://doi.org/10.1038/s41561-019-0373-z, 2019.

Vogel, C., Mueller, C. W., Höschen, C., Buegger, F., Heister, K., Schulz, S., Schloter, M., and Kögel-Knabner, I.: Submicron structures provide preferential spots for carbon and nitrogen sequestration in soils, Nat. Commun., 5, 1–7,

https://doi.org/10.1038/ncomms3947, 2014.

Vogel, H. J., Eberhardt, E., Franko, U., Lang, B., Ließ, M., Weller, U., Wiesmeier, M., and Wollschläger, U.: Quantitative Evaluation of Soil Functions: Potential and State, Front. Environ. Sci., 7, https://doi.org/10.3389/fenvs.2019.00164, 2019.

Wagai, R. and Mayer, L. M.: Sorptive stabilization of organic matter in soils by hydrous iron oxides, Geochim. Cosmochim. Acta, 71, 25–35, https://doi.org/10.1016/j.gca.2006.08.047, 2007.

Wallach, D.: Evaluating crop models, Work. with Dyn. Crop Model. Elsevier, Amsterdam, Netherlands, 11–53, 2006.

Waring, B. G., Sulman, B. N., Reed, S., Smith, A. P., Averill, C., Creamer, C. A., Cusack, D. F., Hall, S. J., Jastrow, J. D., Jilling, A., Kemner, K. M., Kleber, M., Liu, X. J. A., Pett-Ridge, J., and Schulz, M.: From pools to flow: The PROMISE framework for new insights on soil carbon cycling in a changing world, Glob. Chang. Biol., 26, 6631–6643, https://doi.org/10.1111/gcb.15365, 2020.

Weihermüller, L., Graf, A., Herbst, M., and Vereecken, H.: Simple pedotransfer functions to initialize reactive carbon pools of the RothC model, Eur. J. Soil Sci., 64, 567–575, https://doi.org/10.1111/ejss.12036, 2013.

Weil, R. and Brady, N. C.: The Nature and Properties of Soils, Fifteenth., Pearson, 2016.

West, T. O., Marland, G., King, A. W., Post, W. M., Jain, A. K., and Andrasko, K.: Carbon management response curves: Estimates of temporal soil carbon dynamics, Environ. Manage., 33, 507–518, https://doi.org/10.1007/s00267-003-9108-3, 2004.

Whisler, F. D., Acock, B., Baker, D. N., Fye, R. E., Hodges, H. F., Lambert, J. R., Lemmon, H. E., McKinion, J. M., and Reddy, V. R.: Crop simulation models in agronomic systems, Adv. Agron., 40, 141–208, 1986.

Wickham, H.: Reshaping Data with the reshape Package, J. Stat. Softw., 21, 1–20, https://doi.org/10.18637/jss.v021.i12, 2007.

Wickham, H.: ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016.

Wieder, W. R., Grandy, A. S., Kallenbach, C. M., Taylor, P. G., and Bonan, G. B.: Representing life in the Earth system with soil microbial functional traits in the MIMICS model, Geosci. Model Dev., 8, 1789–1808, https://doi.org/10.5194/gmd-8-1789-2015, 2015.

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H. J., and Kögel-Knabner, I.: Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales, Geoderma, 333, 149–162, https://doi.org/10.1016/j.geoderma.2018.07.026, 2019.

WMO: State of the Global Climate 2020, 56 pp., 2021.

Woodruff, C. M.: Estimating the Nitrogen Delivery of Soil from the Organic Matter Determination as Reflected by Sanborn Field, Soil Sci. Soc. Proceednigs, 14, 208–212, https://doi.org/10.2136/sssaj1950.036159950014000c0047x, 1949.

Wutzler, T. and Reichstein, M.: Soils apart from equilibrium - Consequences for soil carbon balance modelling, 4, 125–136, https://doi.org/10.5194/bg-4-125-2007, 2007.

Xu, X., Liu, W., and Kiely, G.: Modeling the change in soil organic carbon of grassland in response to climate change: Effects of measured versus modelled carbon pools for

initializing the Rothamsted Carbon model, Agric. Ecosyst. Environ., 140, 372–381, https://doi.org/10.1016/j.agee.2010.12.018, 2011.

Zegouagh, Y., Derenne, S., Dignac, M. F., and Baruiso, E.: Demineralisation of a crop soil by mild hydrofluoric acid treatment Influence on organic matter composition and pyrolysis, 71, 119–135, https://doi.org/10.1016/S0165-2370(03)00059-7, 2004.

Zhang, Y., Lavallee, J. M., Robertson, A. D., Even, R., Ogle, S. M., Paustian, K., and Cotrufo, M. F.: Simulating measurable ecosystem carbon and nitrogen dynamics with the mechanistically defined MEMS 2.0 model, 18, 3147–3171, https://doi.org/10.5194/bg-18-3147-2021, 2021.

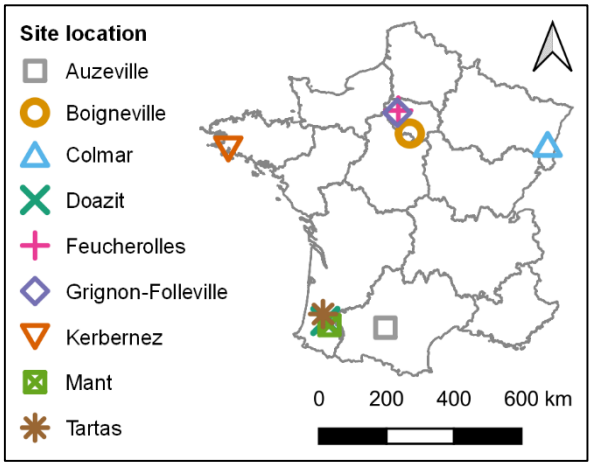
Zimmermann, M., Leifeld, J., Schmidt, M. W. I., Smith, P., and Fuhrer, J.: Measured soil organic matter fractions can be related to pools in the RothC model, Eur. J. Soil Sci., 58, 658–667, https://doi.org/10.1111/j.1365-2389.2006.00855.x, 2007a.

Zimmermann, M., Leifeld, J., and Fuhrer, J.: Quantifying soil organic carbon fractions by infrared-spectroscopy, Soil Biol. Biochem., 39, 224–231, https://doi.org/https://doi.org/10.1016/j.soilbio.2006.07.010, 2007b.

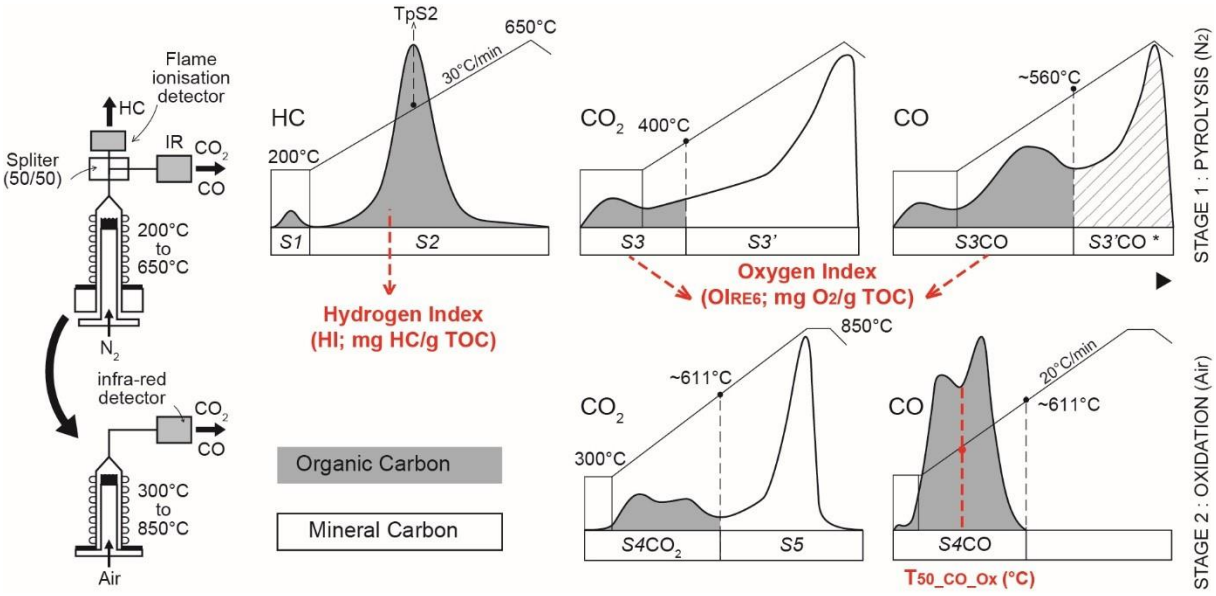
6. SUPPLEMENTARY MATERIAL

Supplementary material to Chapter 1

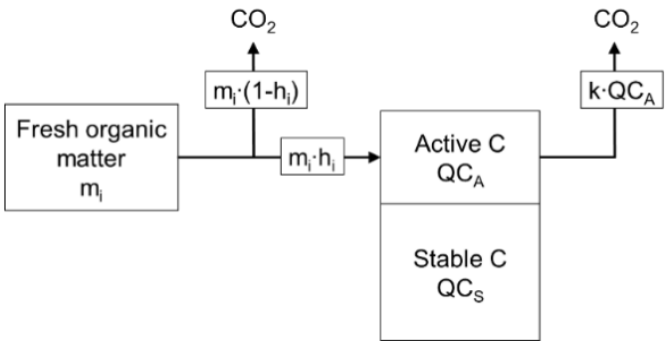
A robust initialization method for accurate soil organic carbon simulations



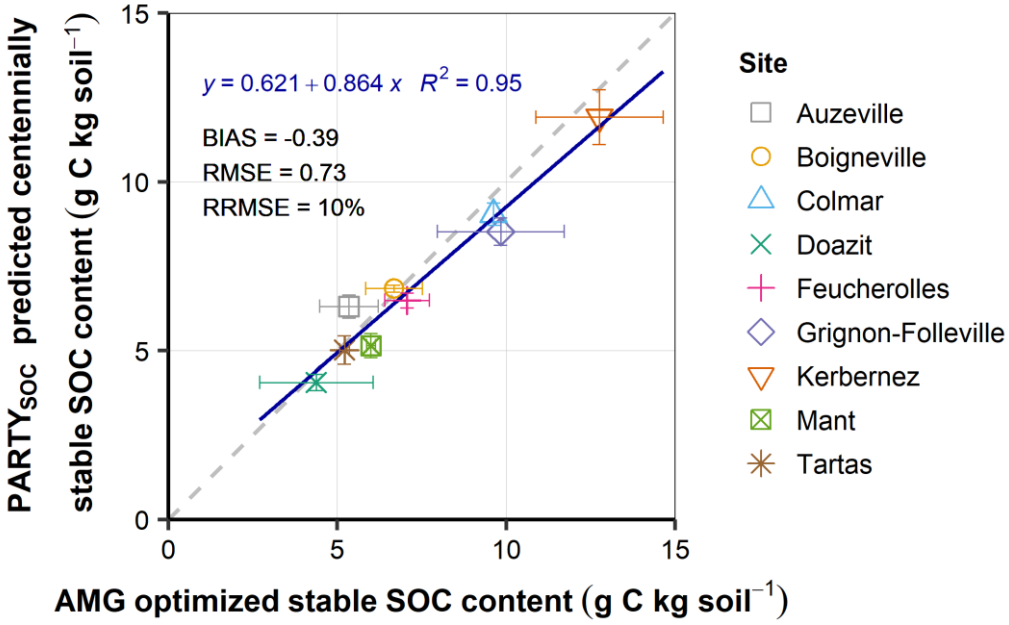
Supplementary Material Figure 1: Location of the nine French long-term agricultural experiments used in this study.



Supplementary Material Figure 2: Schematic representation of the two stages and output of the Rock-Eval® thermal analysis method. The sequential pyrolysis and oxidation produce five thermograms, from which thermal parameters are calculated. The temperature ramps shown here represent the analysis routine used in this study. The grey thermogram areas are used for the calculation of the organic carbon content and the white areas for the mineral carbon content.

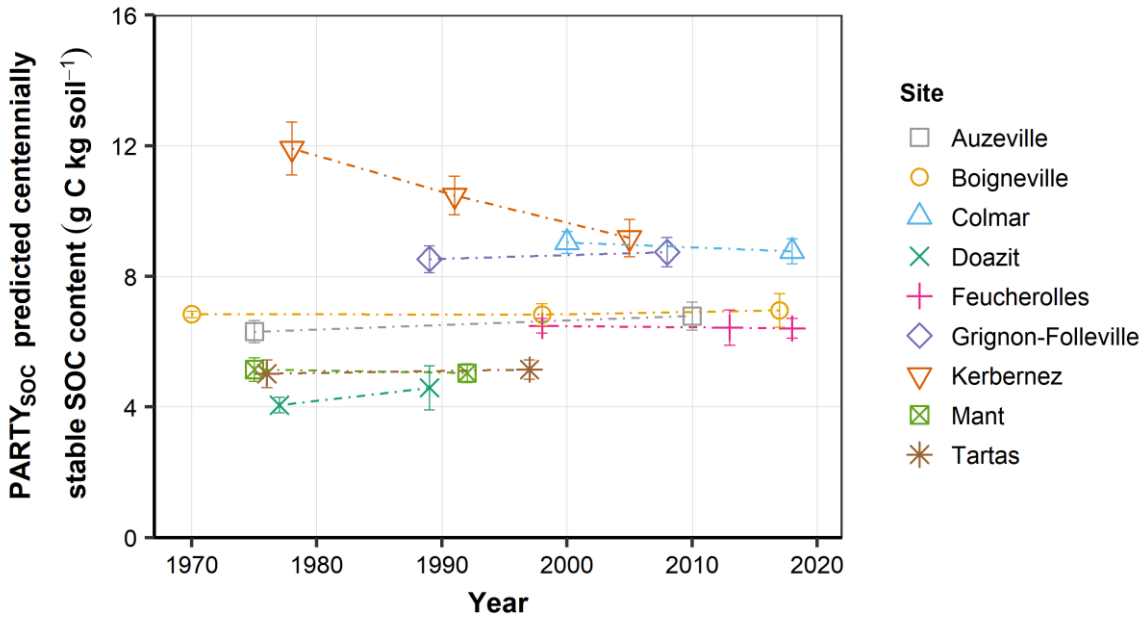


Supplementary Material Figure 3: Conceptual schematic diagram of the AMG model of SOC dynamics (modified from Levavasseur et al., 2020 and Duparque et al., 2013), showing SOC pools, fluxes and transport rates. A fraction (1-h) of fresh organic matter (m) is yearly mineralized and released in the atmosphere, whereas a fraction (h) is incorporated into the active SOC pool (C_A). The coefficient of mineralization (k) controls carbon discharge from C_A into the atmosphere. There is no exchange with the stable SOC pool (C_S).

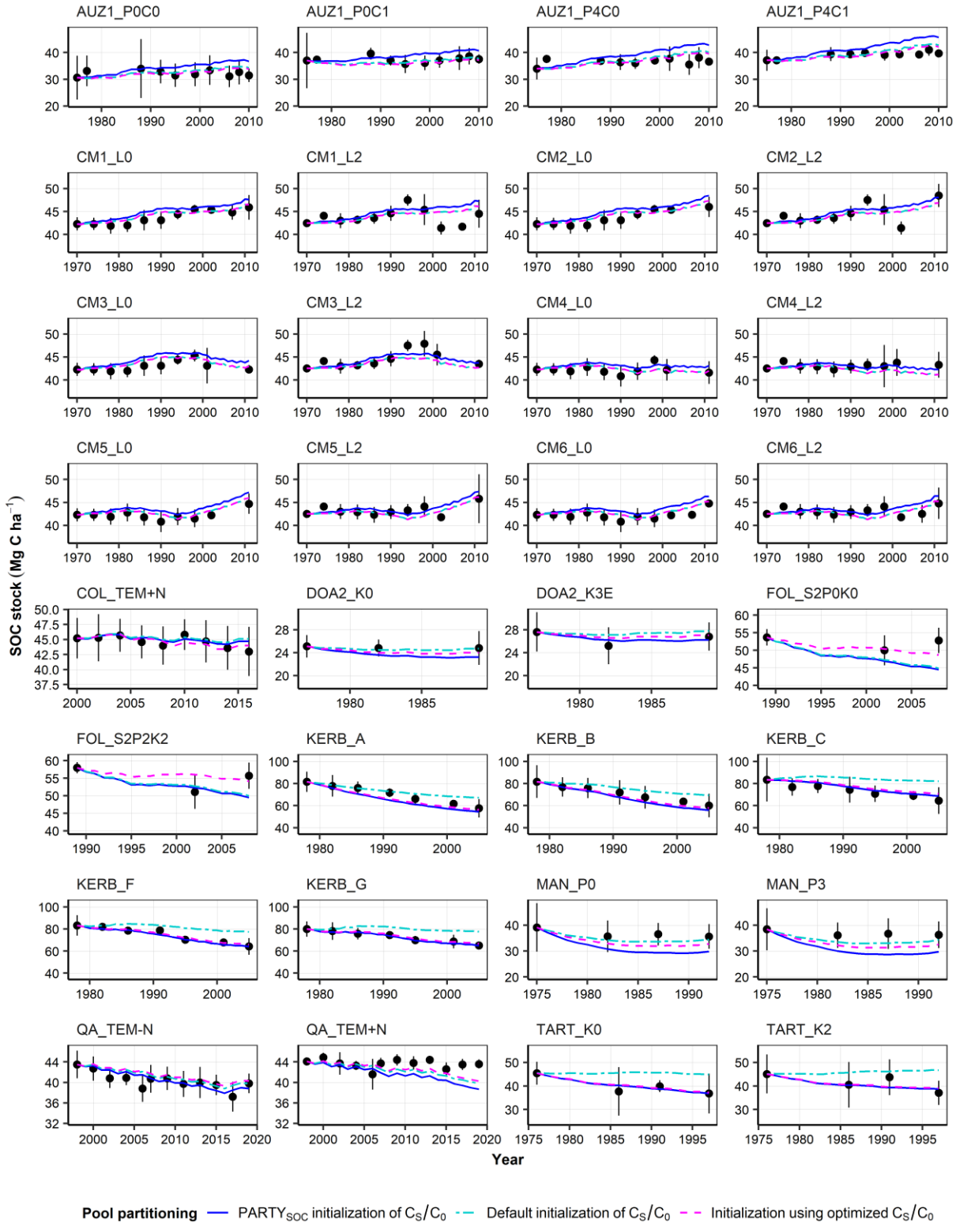


Supplementary Material Figure 4: Centennially stable SOC content predicted by the Rock-Eval®-based $PARTY_{SOC}$ machine-learning model compared to the AMG *ex-post* optimized stable SOC content. Points represent site-mean values based on initial topsoil samples. Statistics refer to the linear regression between x and y values (blue solid line). Horizontal error bars show the uncertainty associated with the AMG optimal stable SOC content, calculated as the standard deviation of treatment-wise optimized stable SOC content. Vertical error bars represent the prediction error of the centennially stable SOC content values, calculated from the standard deviation of the $PARTY_{SOC}$ model predictions on initial topsoil samples.

The apparent decrease in centennially stable SOC content for the site of Kerbernez could be explained by changes in soil bulk density, caused by the change in land-use (from grassland to cropland) in 1958. The subsequent soil compaction may have led to inclusion of deeper soil during standard sampling of the 0–25cm layer, causing a false effect of SOC content decrease. Lack of regular soil bulk density measurements during the experiment (1978–2005) hinders explicit analysis of this hypothesis.



Supplementary Material Figure 5: Centennially stable SOC content predicted by PARTY$_{SOC}$ as a function of time of the experiment. The points on the plot represent site mean values for the shown dates and the vertical error bars represent the standard deviation of the sample set used for averaging.



Supplementary Material Figure 6: AMG simulations of observed SOC dynamics for the 32 treatments used in this study. The black points represent observed SOC stocks in topsoils, the vertical error bars indicate the confidence interval of the measurements, and each coloured line corresponds to a simulation resulting from a different initialization method, namely initial pool partitioning according to: PARTY$_{SOC}$-based C_S/C_0 initialization in blue, AMG default C_S/C_0 in cyan, and AMG ex-post optimized C_S/C_0 in magenta. Note the different y-axis range across sites. The treatment names and their corresponding sites are presented in Supplementary Material Table 1.

Supplementary Material Table 1: Information on site location, long-term land cover history, climate and soil characteristics. Note that the arable land cover class may include temporary grassland in crop rotations, while the grassland land cover class does not include cultivated crops.

	Auzeville	**Boigneville**	**Colmar**	**Doazit**	**Feucherolles**	**Grignon-Folleville**	**Kerbernez**	**Mant**	**Tartas**
Latitude ° N	43.527479	48.327843	48.059271	43.700824	48.896501	48.841722	47.946698	43.5917	43.865475
Longitude ° E	1.506059	2.382406	7.328160	−0.629406	1.972125	1.936675	−4.127084	−0.5028	−0.729405
⁎Historical land cover 1820–1866	arable land	arable land	arable land	arable land	arable land	arable land	arable land	arable land	grassland
†Historical land cover 1950–1965	arable land	arable land	arable land	arable land	arable land	arable land	grassland	arable land	arable land
‡Treatment	AUZ1_P0C0 AUZ1_P0C1 **AUZ1_P4C0** **AUZ1_P4C1**	**CM1_L0** **CM1_L2** CM2_L0 CM2_L2 CM3_L0 CM3_L2 CM4_L0 CM4_L2 CM5_L0 CM5_L2 CM6_L0 CM6_L2	**TEM+N**	**DOA2_K0** DOA2_K3E	**QU_TEM-N** **QU_TEM+N**	**FOL_S2P0K0** **FOL_S2P2K2**	**KERB_A** **KERB_B** **KERB_C** KERB_F KERB_G	**MAN_P0** MAN_P3	**TART_K0** TART_K2
§MAT (°C)	13.5	10.9	11.2	13.1	10.8	11.0	11.9	13.1	13.4
||MAP-PET (mm)	−290	−87	−222	384	5	−69	489	364	383
¶Bulk density (g · cm^{-3})	1.40	**1.44**	**1.30**	1.40	**1.38**	1.40	1.30	1.40	1.40
Considered soil mass (Mg ha^{-1})	4200	4103	3640	3500	3864	4200	3023	3920	3920
Clay (g kg soil $^{-1}$)	275	248	180	72	170	244	163	94	43
Silt (g kg soil $^{-1}$)	339	672	628	403	779	601	391	554	166
Sand (g kg soil $^{-1}$)	372	80	76	525	51	97	446	349	791
$CaCO_3$ (g kg soil $^{-1}$)	15	0	115	0	0	58	0	3	0
C:N ratio	8.0	9.0	9.2	10.6	9.3	9.8	11.4	9.4	13.0
pH	7.6	6.8	8.3	6.4	6.9	8.1	5.7	7.6	6.0
Reference	(Colomb et al., 2007)	(Dimassi et al., 2014)	(Obriot, 2016)	(Lubet et al., 1993)	(Noirot-Cosson et al., 2016)	(Barré et al., 2008)	(Vertès et al., 2007)	(Messiga et al., 2010)	(Morel et al., 2014)

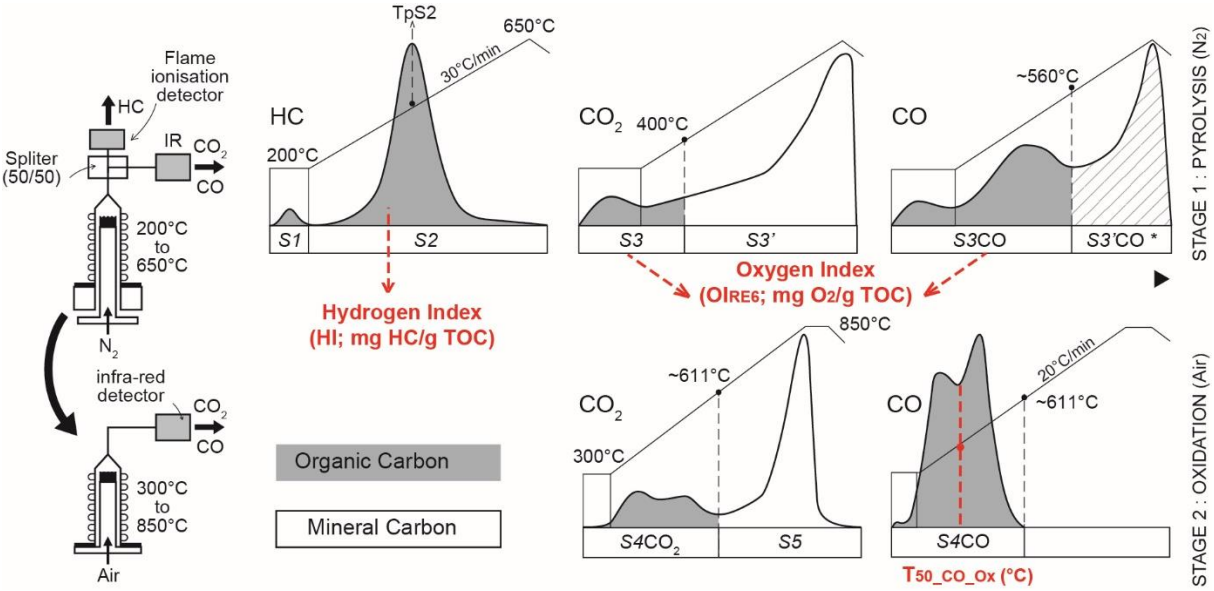
⁎ French "Carte de l'Etat Major", IGN
† aerial photography, IGN
‡ Treatments from which samples were available for Rock-Eval® analysis are in bold
§ Mean annual temperature
|| Mean annual precipitation-potential evapotranspiration
¶ LTEs for which changes in bulk density with time were measured and considered for the calculation of SOC stocks are in bold

Supplementary Material Table 2: Measurement error and variation of initial SOC stock values, and variation of initial centennially stable SOC proportion amongst sites. Left part: Comparison of the variation (standard deviation) and uncertainty (confidence interval) associated with initial SOC stock measurements. Right part: variation of initial centennially stable SOC proportions predicted by the $PARTY_{SOC}$ machine-learning model for each site.

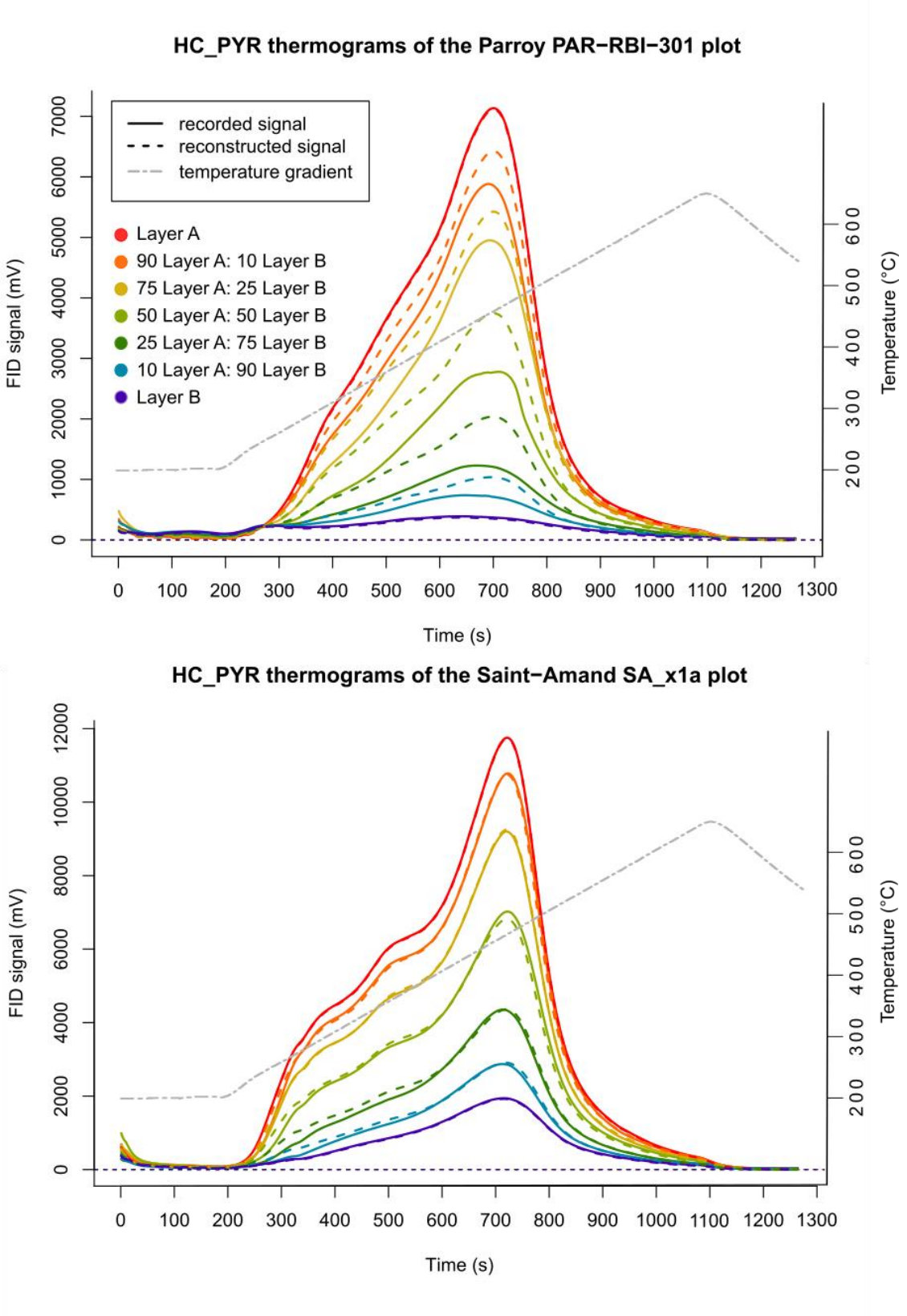
Site	**Initial SOC stock ($tC \cdot ha^{-1}$)**			**Initial centennially stable SOC proportion predicted using the $PARTY_{SOC}v2.0_{EU}$ statistical model**	
	Mean	**Standard deviation**	**Confidence interval**	**Mean**	**Standard deviation**
Auzeville	34.68	2.66	13.30	0.74	0.01
Boigneville	42.40	0.10	2.30	0.68	0.05
Colmar	45.20	-	6.74	0.64	0.02
Doazit	26.35	1.25	5.38	0.57	0.01
Grignon-Folleville	55.85	2.15	3.93	0.64	0.04
Feucherolles	43.80	0.42	3.49	0.62	0.02
Kerbernez	81.98	1.29	24.01	0.44	0.02
Mant	38.75	0.35	17.55	0.52	0.05
Tartas	45.25	0.15	13.14	0.44	0.05

Supplementary material to Chapter 2

Predicting Rock-Eval® thermal analysis parameters of a soil layer based on samples from its sublayers; an experimental study on forest soils.



Supplementary Figure 1: Schematic representation of the Rock-Eval® thermal analysis process. Pyrolysis is followed by oxidation, resulting in 3 and 2 thermograms respectively. The grey areas correspond to detection of carbon effluents of organic origin. For the calculation of temperature parameters, the integration is restricted to temperature limits of 560 °C for pyrolysis and 611 °C for oxidation to separate effluents emitted from an organic source from those caused by thermal decomposition of carbonates.



Supplementary Figure 2: Example of recorded and reconstructed HC_PYR thermograms of a plot with a heterogeneous clay content profile (Parroy) and a plot with little change of clay content with depth (Saint-Amand).

Supplementary Table 1: List of abbreviations and definitions of classic and extended Rock-Eval® parameters that were both measured and calculated in this study. The 18 Rock-Eval® parameters used as predictors by the latest version of the $PARTY_{SOC}$ model (Cécillon et al., 2021) are indicated with an asterisk. The S2 parameter, typically given in mg HC · g^{-1}soil is converted to g C · kg^{-1}soil in the $PARTY_{SOC}$ model.

#	Rock-Eval® Parameters	Definition
1	S1 (mg HC · g^{-1}soil)	Free hydrocarbons; HC_PYR thermogram area corresponding to the 200 °C isotherm
2*	S2 (mg HC · g^{-1}soil)	HC_PYR thermogram area corresponding to the temperature ramp (200–650 °C)
3	S3 (mg CO_2 · g^{-1}soil)	CO2_PYR thermogram area corresponding to the temperature ramp (200–400 °C)
4	S3' (mg CO_2 · g^{-1}soil)	CO2_PYR thermogram area corresponding to the temperature ramp (400–650 °C)
5	S3CO (mg CO · g^{-1}soil)	CO_PYR thermogram area corresponding to the temperature ramp (200–560 °C)
6	S3'CO (mg CO · g^{-1}soil)	CO_PYR thermogram area corresponding to the temperature ramp (560–650 °C)
7*	PC (g C · kg^{-1}soil)	Pyrolyzed Carbon $PC = [(S1 + S2) \times 0.83] + \left[S3 \times \frac{12}{44}\right] + \left[(S3CO + \frac{S3'CO}{2}) \times \frac{12}{28}\right]$
8*	TOC (g C · kg^{-1}soil)	Total Organic Carbon $TOC = PC + \left[S4CO \times \frac{12}{28}\right] + \left[S4CO_2 \times \frac{12}{44}\right]$
9	MinC (g C · kg^{-1}soil)	Mineral Carbon $MinC = PC + \left[S4CO \times \frac{12}{28}\right] + \left[S4CO_2 \times \frac{12}{44}\right]$
10*	$T70_{HC\_PYR}$ (°C)	Temperature at which 70% of HC of organic origin was pyrolyzed
11*	$T90_{HC\_PYR}$ (°C)	_____ 90% _____
12*	$T30_{CO2\_PYR}$ (°C)	Temperature at which 30% of CO_2 of organic origin was pyrolyzed
13*	$T50_{CO2\_PYR}$ (°C)	_____ 50% _____
14*	$T70_{CO2\_PYR}$ (°C)	_____ 70% _____
15*	$T90_{CO2\_PYR}$ (°C)	_____ 90% _____
16*	$T70_{CO\_OX}$ (°C)	Temperature at which 70% of residual CO of organic origin was oxidized
17*	$T50_{CO2\_OX}$ (°C)	Temperature at which 50% of residual CO_2 of organic origin was oxidized
18*	$T70_{CO2\_OX}$ (°C)	_____ 70% _____
19*	$T90_{CO2\_OX}$ (°C)	_____ 90% _____
20*	Pseudo_S1 (g C · kg^{-1}soil)	Peak area 0–200 seconds of the sum of pyrolysis thermograms (SUMPYR=HC_PYR+CO_PYR+CO2_PYR)

21*	HI (mg HC · g^{-1}C)	Hydrogen Index; amount of HC released relatively to TOC $HI = \frac{S2 \times 100}{TOC}$
22	OIRE6 (mg O_2 · g^{-1}C)	Oxygen Index; amount of O_2 released relatively to TOC $OI_{RE6} = \frac{\left[S3 \times \frac{32}{44}\right] + \left[S3CO \times \frac{16}{28}\right]}{TOC}$
23*	S2/PC	Ratio of S2 peak area (200–1263s or $200 - 650\ °C$, in g C · kg^{-1}soil) to Pyrolyzed Carbon
24*	PC/TOC	Ratio of Pyrolyzed Carbon to Total Organic Carbon
25*	HI/OI_{RE6}	Ratio of Hydrogen Index to Oxygen Index
26	I-index	Thermally labile hydrocarbons $I - index = \log_{10}\left(\frac{proportion\ of\ hydrocarbons\ released\ between\ 200 - 400\ °C}{proportion\ of\ hydrocarbons\ released\ between\ 400 - 460\ °C}\right)$
27	R-index	Refractory hydrocarbon fraction $R - index = \frac{HC\_pyr\ thermogram\ area\ between\ 400 - 650\ °C}{HC\_pyr\ thermogram\ area\ corresponding\ to\ the\ temperature\ ramp\ (200 - 650\ °C\ or\ 200 - 1263\ s)}$
28	TLHC-index	Thermo-labile hydrocarbon index $TLHC - index = \frac{HC\_pyr\ thermogram\ area\ between\ 200 - 450\ °C}{HC\_pyr\ thermogram\ area\ corresponding\ to\ the\ temperature\ ramp\ (200 - 650\ °C\ or\ 200 - 1263\ s)}$
29	Stable SOC proportion	Proportion of centennially persistent soil organic carbon, predicted with the $PARTY_{SOC}v2.0_{EU}$ model

Supplementary Table 2: Statistical evaluation of the two calculation methods (see Sections 3.1 and 3.2). Rows show calculated metrics for each method averaged for all plots. The last row provides the median value of the R^2 and NRMSE statistical parameters.

#	Parameter	Acquisition parameters weighted mean					Reconstructed thermograms calculation				
		R^2	p-value	BIAS	RMSE	NRMSE	R^2	p-value	BIAS	RMSE	NRMSE
1	S1	0.22	0.00061	0.01	0.05	40.16					
2	S2	1.00	3.95E-60	0.12	0.38	1.51					
3	S3	1.00	4.88E-72	0.00	0.14	0.88					
4	S3'	0.98	4.12E-41	-0.13	0.52	3.95					
5	S3CO	0.99	7.38E-56	-0.03	0.08	2.11					
6	S3'CO	0.94	6.44E-31	-0.04	0.14	6.28					
7	PC	1.00	4.13E-65	0.07	0.30	1.11					
8	TOC	1.00	7.43E-68	-0.47	0.86	1.14					
9	MinC	0.99	2.11E-49	-0.03	0.08	2.24					
10	$T70_{HC\_PYR}$	0.80	3.63E-18	-1.46	3.12	12	0.75	5.50E-16	-2.22	3.57	13.73
11	$T90_{HC\_PYR}$	0.91	4.88E-27	1.66	3.87	8.79	0.78	1.94E-17	-3.65	6.21	14.11
12	$T30_{CO2\_PYR}$	0.85	4.36E-21	-0.11	2.64	8.53	0.87	1.25E-22	-1.97	2.99	9.66
13	$T50_{CO2\_PYR}$	0.85	9.82E-22	0.83	3.81	8.86	0.88	2.25E-23	-2.03	3.71	8.62
14	$T70_{CO2\_PYR}$	0.86	2.04E-22	1.02	4.28	8.74	0.89	5.89E-25	-1.44	3.70	7.56
15	$T90_{CO2\_PYR}$	0.74	1.32E-15	0.43	4.73	12.12	0.76	1.81E-16	-0.90	4.55	11.65
16	$T70_{CO\_OX}$	0.83	7.29E-20	0.59	16.31	10.95	0.77	4.33E-17	-6.60	19.57	13.13
17	$T50_{CO2\_OX}$	0.93	6.63E-30	4.30	6.27	11.2	0.93	1.06E-29	-1.69	3.98	7.1
18	$T70_{CO2\_OX}$	0.96	1.11E-34	2.72	5.55	6.68	0.95	4.48E-32	-3.66	6.25	7.53
19	$T90_{CO2\_OX}$	0.93	3.33E-29	-0.25	5.40	7.1	0.93	1.14E-28	-4.04	6.79	8.94
20	Pseudo S1	0.94	2.71E-31	0.01	0.05	9.18					
21	HI	0.90	2.02E-25	17.82	26.56	9.8					
22	OI_{RE6}	0.94	4.73E-31	3.33	8.23	5.56					
23	S2/PC	0.88	1.51E-23	0.02	0.04	10.22					
24	PC/TOC	0.85	4.29E-21	0.01	0.02	12.32					
25	HI/OI_{RE6}	0.95	1.06E-32	0.07	0.12	5.95					
26	I-index	0.84	1.28E-20	0.15	0.18	58.91					
27	R-index	0.80	1.56E-18	-0.02	0.03	14.71					
28	TLHC-index	0.76	1.08E-16	0.01	0.02	13.53					
29	[++]Stable SOC proportion	0.94	0	-0.01	0.03	7.23	0.94	0	-0.02	0.04	8.59
	MEDIAN	**0.93**	**0**			**8.74**					

[++]Additional to the table: Results of alternative way of obtaining the stable SOC proportion on soil mixtures as weighted average of $PARTY_{SOC}v2.0_{EU}$ predictions obtained for layer A and layer B: R^2 = 0.85, p-value = 9.28E-22, BIAS = 0.03, RMSE = 0.06, NRMSE = 12.94%

Supplementary Table 3: Summary statistics of Rock-Eval® parameters measured on the 20 “natural” soil samples.

#	Parameter	Minimum	Maximum	Mean	Median	Standard Deviation
1	S1	0.01	0.3	0.05	0.04	0.06
2	S2	0.47	26.17	5.01	3.84	5.9
3	S3	0.68	17.8	4.99	3.25	4.31
4	S3'	0.7	13.7	4.71	3.3	3.55
5	S3CO	0.15	4.1	1.32	0.98	1.08
6	S3'CO	0.2	2.3	0.84	0.7	0.6
7	PC	0.8	29.1	6.3	4.8	6.61
8	TOC	2.9	83.2	21.52	15.7	20.28
9	MinC	0.03	4.09	0.4	0.16	0.88
10	$T70_{HC\_PYR}$	448	482	462.3	462	6.94
11	$T90_{HC\_PYR}$	488	544	511.5	506	14.56
12	$T30_{CO2\_PYR}$	315	353	332.75	332	8.73
13	$T50_{CO2\_PYR}$	356	410	381.15	380	12.9
14	$T70_{CO2\_PYR}$	407	467	433.7	430.5	14.64
15	$T90_{CO2\_PYR}$	481	527	503.45	500.5	11.57
16	$T70_{CO\_OX}$	394	522	460.55	450.5	44.53
17	$T50_{CO2\_OX}$	392	459	421.45	417	20.14
18	$T70_{CO2\_OX}$	422	512	465.85	456.5	27.95
19	$T90_{CO2\_OX}$	485	565	530.8	532.5	23.91
20	Pseudo S1	0.05	0.83	0.2	0.14	0.17
21	HI	102	387	206.65	206.5	71.49
22	OI_{RE6}	150	333	225.45	210	53.54
23	S2/PC	0.35	0.8	0.59	0.63	0.13
24	PC/TOC	0.22	0.4	0.28	0.28	0.04
25	HI/OI_{RE6}	0.32	2.58	1.02	0.99	0.55
26	I-index	-0.1	0.3	0.1	0.12	0.1
27	R-index	0.53	0.72	0.62	0.6	0.05
28	TLHC-index	0.51	0.71	0.63	0.64	0.04
29	Stable SOC proportion	0.26	0.78	0.47	0.37	0.19

Supplementary material to Chapter 3

Understanding the influence of organo-mineral interactions on organic matter thermal stability expressed by Rock-Eval® parameters.

Table S1: Measurements of pH in the initial solution containing the organic compound and mineral mixture. For sand and montmorillonite, which exhibited the strongest positive and negative influence on pH for the BSA mixtures a measurement was taken also during the last rinsing step to give an idea of the range of variation in pH during the experiment.

	Bovine serum albumin		Cysteine		Humic acid	
	Initial solution	**Last rinsing step**	**Initial solution**	**Last rinsing step**	**Initial solution**	**Last rinsing step**
Sand	6.53	5.90	5.60	6.91	12.47	7.10
Montmorillonite	3.61	3.83	3.33	3.87	10.84	8.93
Kaolinite	5.51		4.27		12.49	
Goethite	5.95		7.62		12.50	
Soil 1	4.55		4.19		12.50	
Soil 2	4.60		4.15		12.50	
Soil 3	5.20		4.53		12.52	

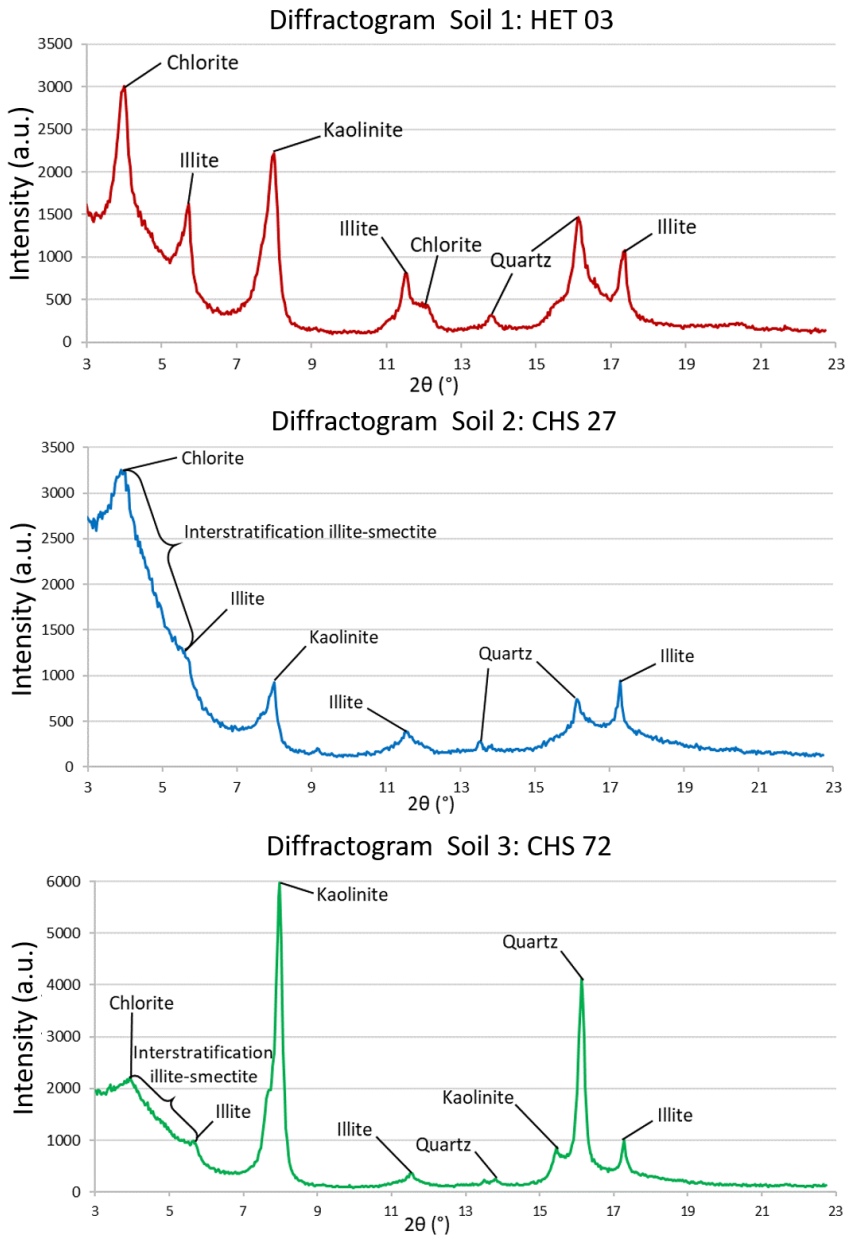


Figure S1: X-Ray diffraction spectra of the selected soil matrices, obtained with a Rigaku XMAX 2500. Peaks represent crystalline phases. Their position on the diffractogram given here as the scattering angle (2θ) in ° may be translated into distance between two crystallographic planes according to Bragg's law (Bragg, 1934) and it allows the identification of the mineral phases present.

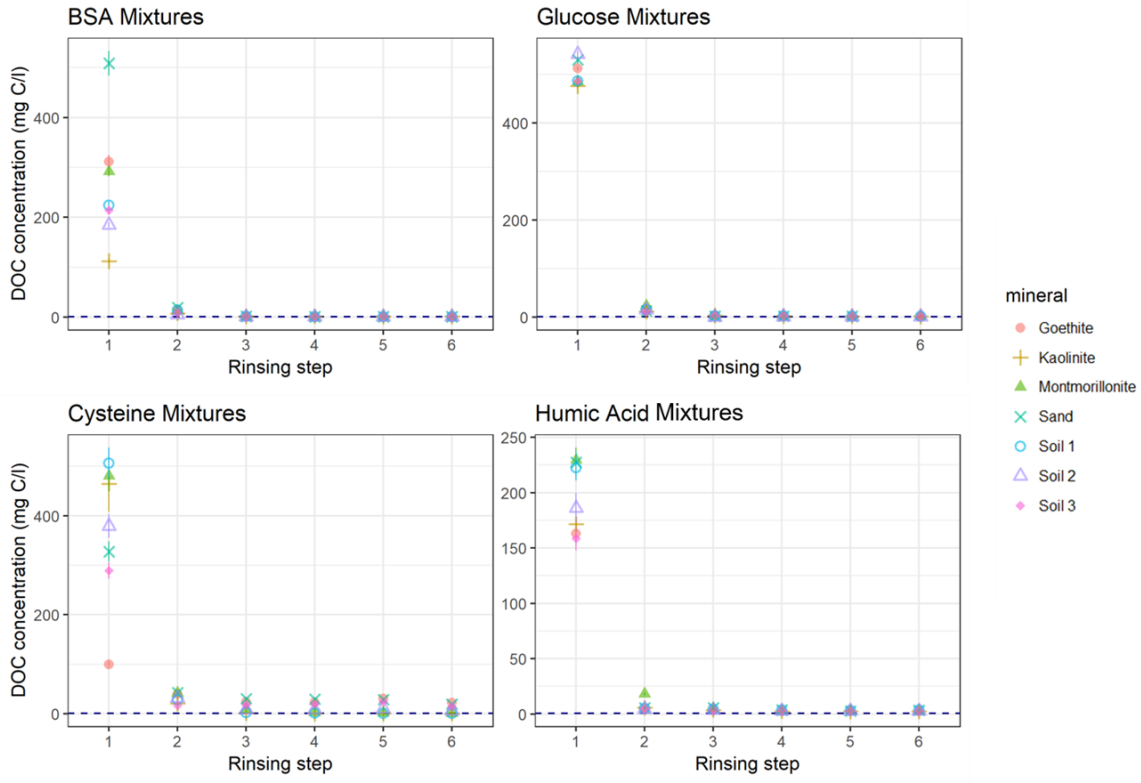
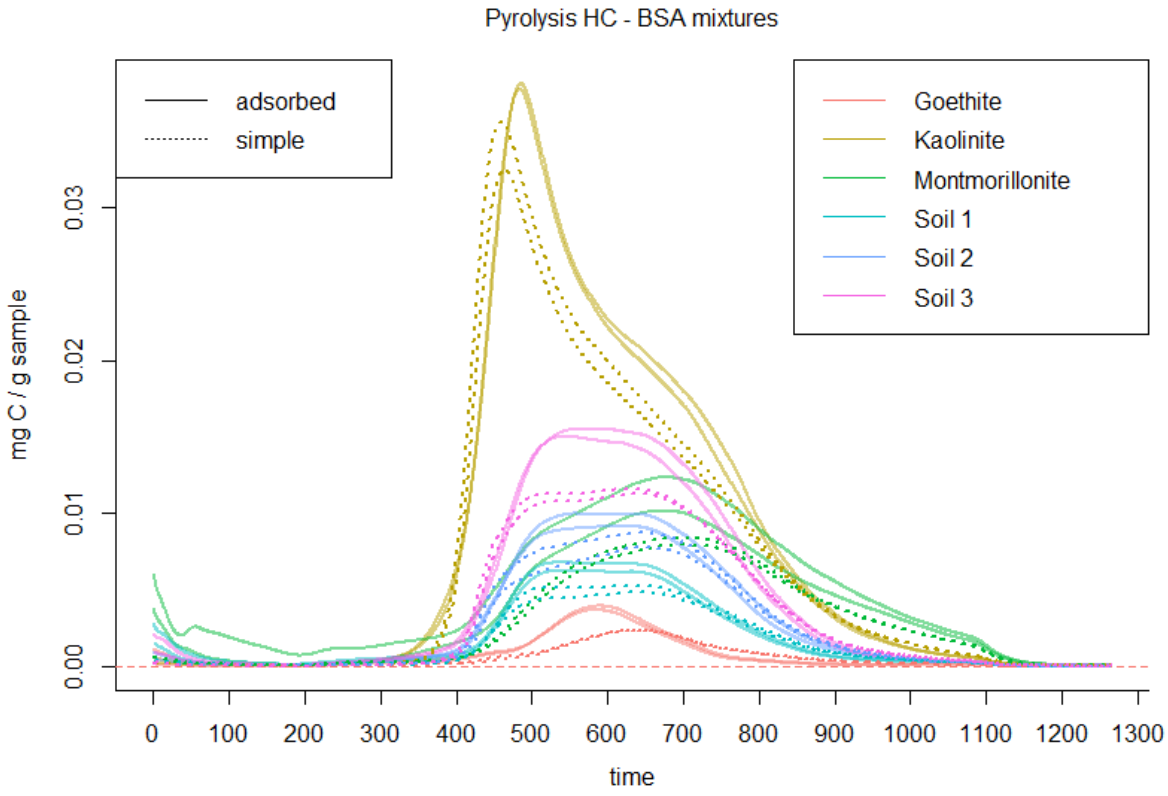
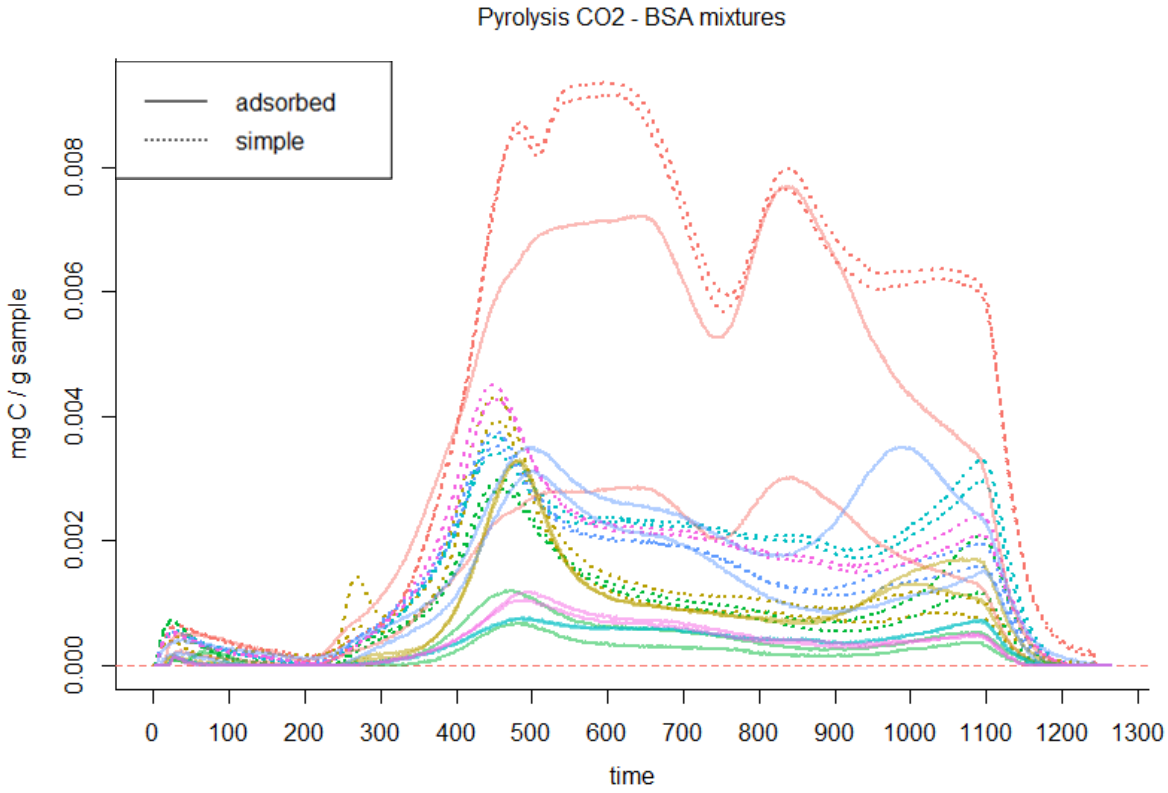
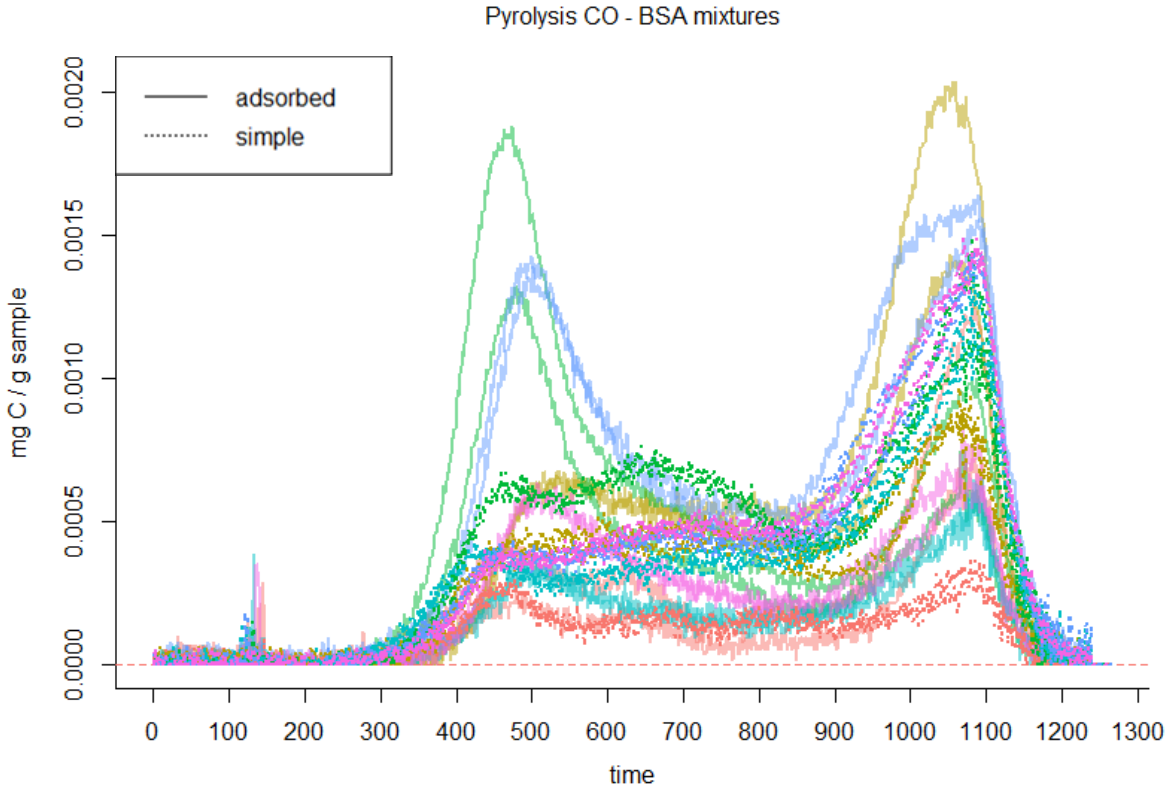
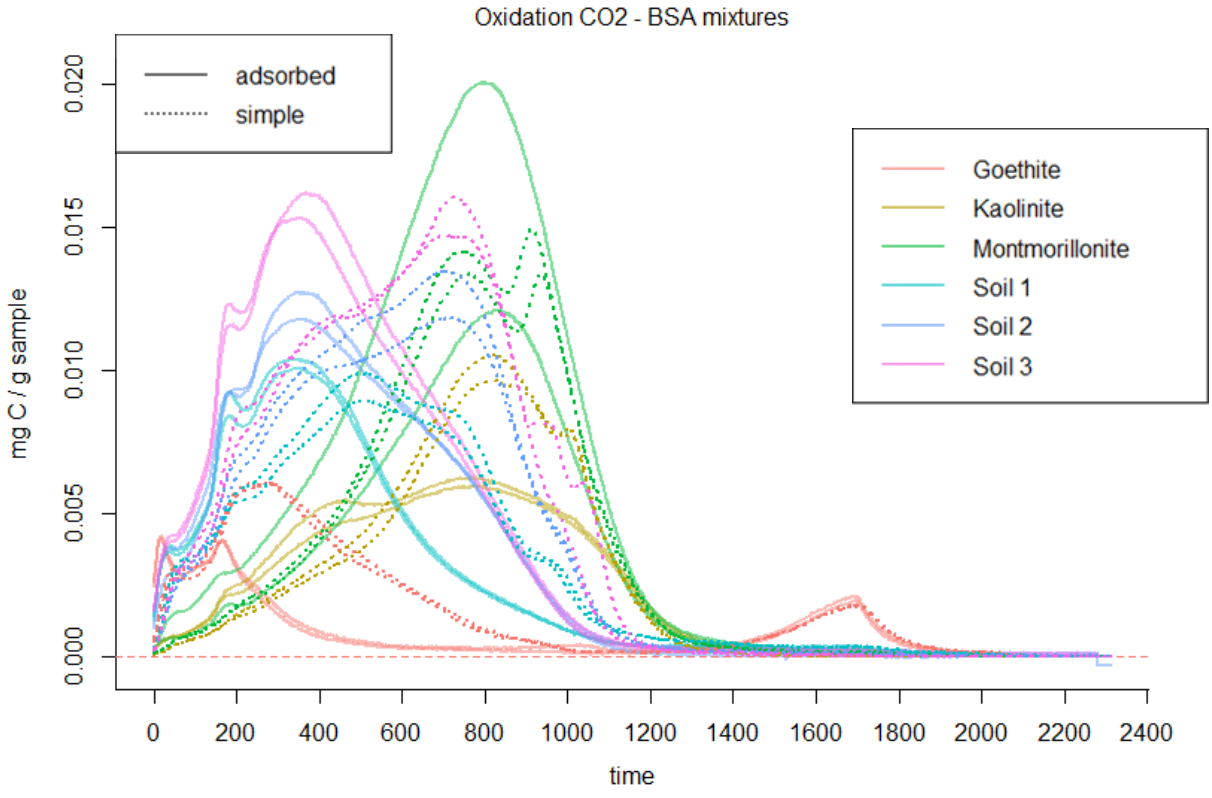


Figure S2: Measurements of dissolved organic carbon in rinsing solutions. Points represent mean values of triplicate analyses. Error bars (although not visible after the first step because they are too small) are given as the standard deviation calculated using the three replicates.


Pyrolysis HC - BSA mixtures
adsorbed
simple
Goethite
Kaolinite
Montmorillonite
Soil 1
Soil 2
Soil 3
mg C / g sample
0.00
0.01
0.02
0.03
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
time


Pyrolysis CO2 - BSA mixtures
adsorbed
simple
mg C / g sample
0.000
0.002
0.004
0.006
0.008
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
time


Pyrolysis CO - BSA mixtures
adsorbed
simple
mg C / g sample
0.0000
0.0005
0.0010
0.0015
0.0020
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
time


Oxidation CO2 - BSA mixtures
adsorbed
simple
Goethite
Kaolinite
Montmorillonite
Soil 1
Soil 2
Soil 3
mg C / g sample
0.000
0.005
0.010
0.015
0.020
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
time

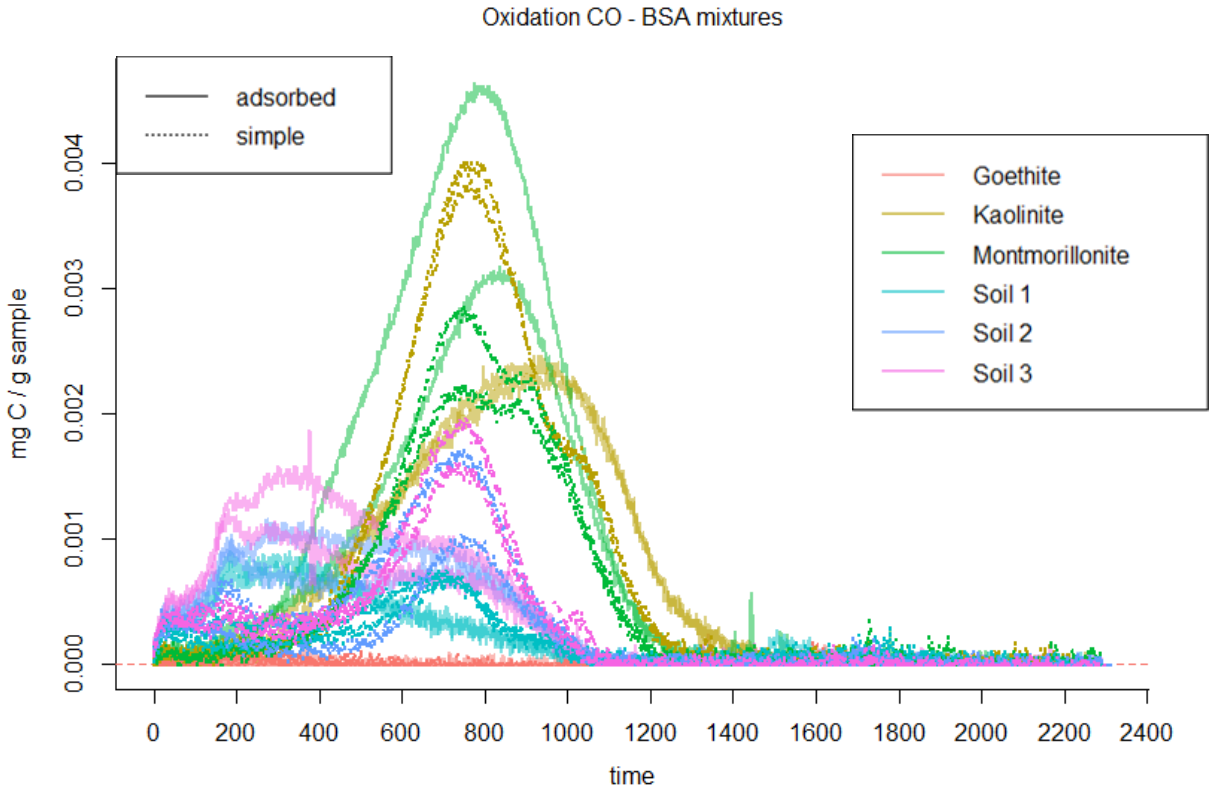


Figure S3: Changes on the shape of the five thermograms of bovine serum albumin (BSA) caused by the presence of dry and adsorbed minerals. The x-axis shows the time of analysis in seconds and corresponds to an increase in temperature.

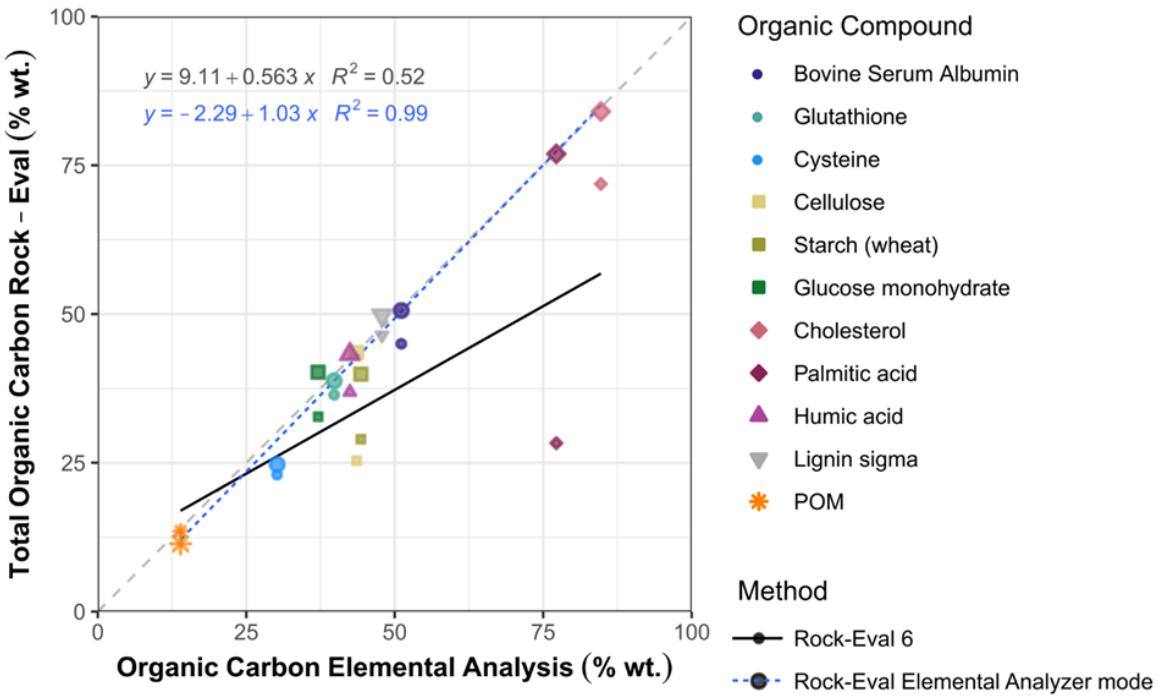


Figure S4: Improvement of carbon yield when using a Rock-Eval® apparatus in Elemental Analyser mode. Smaller points represent the detection of compounds by Rock-Eval® 6, whereas bigger ones show the detection of the same compounds by the specialized experimental Elemental Analyser mode configuration of Rock-Eval®. Black and blue equations and trend lines correspond to the linear models of organic compound detection with classic Rock-Eval® 6 and Elemental Analyser mode configuration of Rock-Eval® respectively in relation to classic Elemental Analyser.

References

Bragg, W.L. *The Crystalline State: Volume I.* New York: The Macmillan Company, 1934.

Short non-technical abstract:

One of the most important solutions to climate change lies literally right under our feet. Soils store twice the amount of carbon that is found in atmosphere and vegetation combined. They act as a buffer between solid earth and atmosphere and exercise a major control on the atmospheric concentration of CO_2 through the release or sink of greenhouse gases. Moreover, organic carbon in soils in the form of organic matter is essential to soil health and fertility, to nutrient availability and water quality. My work is centred around the most valuable tool at our disposal for understanding and predicting the evolution of this reservoir in the future: soil organic carbon (SOC) dynamics models. A missing key influencing the accuracy of SOC model projections and a major challenge in soil science is our ability to estimate the proportion of SOC that will remain unchanged over projection-relevant timescales. This important amount of carbon that has been present in soils for centuries or millennia, and is therefore considered to be “stable”, can vary greatly from one location to another. The goal of my thesis project was to explore a new approach based on thermal analysis of SOC and machine learning, to characterise SOC, estimate the proportion of “stable” carbon in soil samples, and eventually use this information to improve the accuracy of SOC dynamics models. In a second step, I focused on the thermal analysis technique in the heart of this approach to understand better the important information it offers, based on model laboratory experiments. Finally, the main results of my thesis consist, on the one hand, of a complete and validated operational approach improving the accuracy of SOC models with a clear and significant value for “climate-smart” soil management. On the other hand, an experimental part offers new insights into the working principle, limitations and possibilities of the thermal analysis technique at the heart of this approach.

Keywords: [carbon cycle; soil organic matter; SOC modelling; pyrolysis; organo-mineral associations]

Comprendre et utiliser l'estimation de la stabilité du carbone organique du sol par l'analyse thermique Rock-Eval®

Résumé populaire en langue française :

L'une des solutions les plus importantes au réchauffement climatique se trouve sous nos pieds. Les sols forment le plus grand réservoir terrestre de carbone organique. A la croisée de la terre solide et de l'atmosphère, ils constituent un contrôle majeur sur le flux des gaz à effet de serre. En outre, l'augmentation de la quantité du carbone organique dans les sols favorise leur santé et leur fertilité, ainsi que la qualité de l'eau. J'ai développé mon travail autour de l'outil le plus précieux dont nous disposons pour prédire l'évolution de ce réservoir : les modèles de dynamique du carbone organique des sols (COS). Une information clé pour leur précision et une problématique majeure en sciences des sols est notre capacité à estimer la proportion du COS qui persistera sur le long terme. Cette quantité importante de carbone présente dans les sols depuis des siècles ou des millénaires, et donc considérée comme « stable », peut varier fortement d'un endroit à l'autre. L'ambition de mon projet de thèse était d’explorer une nouvelle approche basée sur l'analyse thermique du COS et l’apprentissage automatique, pour caractériser le COS, estimer la proportion du carbone « stable » dans les échantillons de sol et ensuite utiliser cette nouvelle information pour améliorer la précision des modèles de dynamique du COS. Cette nouvelle approche est une clé essentielle pour développer une gestion « intelligente » des sols face au changement climatique. Dans un deuxième temps, je me suis concentrée sur la technique de l'analyse thermique pour déchiffrer et comprendre les informations importantes qu'elle offre de manière plus approfondie, à la base des expériences modèles en laboratoire. Enfin, les résultats principaux de cette thèse consistent, d'une part, en une approche opérationnelle complète et validée améliorant la précision des modèles du COS avec une valeur claire et significative et, d'autre part, en une partie expérimentale offrant de nouveaux aperçus sur le principe de fonctionnement, les limites et les possibilités de la technique d'analyse thermique au cœur de cette approche.

Mots clés : [cycle du carbone ; matière organique du sol ; modélisation du SOC ; pyrolyse ; associations organo-minérales]

ANNEX 1

Geosci. Model Dev., 14, 3879–3898, 2021
https://doi.org/10.5194/gmd-14-3879-2021

Geoscientific Model Development
Open Access
EGU

Partitioning soil organic carbon into its centennially stable and active fractions with machine-learning models based on Rock-Eval® thermal analysis (PARTY$_{SOC}$v2.0 and PARTY$_{SOC}$v2.0$_{EU}$)

Lauric Cécillon[1,2], François Baudin[3], Claire Chenu[4], Bent T. Christensen[5], Uwe Franko[6], Sabine Houot[4],Eva Kanari[2,3], Thomas Kätterer[7], Ines Merbach[8], Folkert van Oort[4], Christopher Poeplau[9], Juan Carlos Quezada[10,11,12], Florence Savignac[3], Laure N. Soucémarianadin[13], and Pierre Barré[2]

[1]Normandie Univ., UNIROUEN, INRAE, ECODIV, Rouen, France
[2]Laboratoire de Géologie, École normale supérieure, CNRS, PSL Univ., IPSL, Paris, France
[3]Institut des Sciences de la Terre de Paris, Sorbonne Université, CNRS, 75005 Paris, France
[4]UMR 1402 ECOSYS, INRAE, AgroParisTech, Univ. Paris Saclay, 78850 Thiverval-Grignon, France
[5]Department of Agroecology, Aarhus University, AU Foulum, 8830 Tjele, Denmark
[6]Department of soil system science, Helmholtz Centre for Environmental Research, UFZ, 06120 Halle, Germany
[7]Department of Ecology, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden
[8]Department Community Ecology, Helmholtz Centre for Environmental Research, UFZ, 06246 Bad Lauchstädt, Germany
[9]Thünen Institute of Climate-Smart Agriculture, 38116 Braunschweig, Germany
[10]Laboratory of Ecological Systems ECOS and Laboratory of Plant Ecology Research PERL, School of Architecture, Civil and Environmental Engineering ENAC, École Polytechnique Fédérale de Lausanne EPFL, 1015 Lausanne, Switzerland
[11]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, 1015 Lausanne, Switzerland
[12]Ecosystem Management, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETHZ, 8092 Zürich, Switzerland
[13]ACTA – les instituts techniques agricoles, 75595 Paris, France

Correspondence: Lauric Cécillon (lauric.cecillon@inrae.fr)

Received: 20 January 2021 – Discussion started: 16 February 2021
Revised: 17 April 2021 – Accepted: 28 May 2021 – Published: 24 June 2021

Abstract. Partitioning soil organic carbon (SOC) into two kinetically different fractions that are stable or active on a century scale is key for an improved monitoring of soil health and for more accurate models of the carbon cycle. However, all existing SOC fractionation methods isolate SOC fractions that are mixtures of centennially stable and active SOC. If the stable SOC fraction cannot be isolated, it has specific chemical and thermal characteristics that are quickly (ca. 1 h per sample) measurable using Rock-Eval® thermal analysis.An alternative would thus be to (1) train a machine-learning model on the Rock-Eval® thermal analysis data for soil samples from long-term experiments for which the size of the centennially stable and active SOC fractions can be estimated and (2) apply this model to the Rock-Eval® data for unknown soils to partition SOC into its centennially stable and active fractions. Here, we significantly extend the validity range of a previously published machine-learning model (Cécillon et al., 2018) that is built upon this strategy. The second version of this model, which we propose to name PARTY$_{SOC}$, uses six European long-term agricultural sites including a bare fallow treatment and one South American vegetation change (C_4 to C_3 plants) site as reference sites. The European version of the model (PARTY$_{SOC}$v2.0$_{EU}$) predicts the proportion of the centennially stable SOC fraction with a root mean square error of 0.15 (relative root mean square error of 0.27) at six independent validation sites. More specifically, our results show that PARTY$_{SOC}$v2.0$_{EU}$ reliably partitions SOC kinetic fractions at its northwestern European validation sites

on Cambisols and Luvisols, which are the two dominant soil groups in this region. We plan future developments of the $PARTY_{SOC}$ global model using additional reference soils developed under diverse pedoclimates and ecosystems to further expand its domain of application while reducing its prediction error.

1 Introduction

Soil organic carbon (SOC) is identified as a key element contributing to soil functions such as primary productivity, water purification and regulation, carbon sequestration and climate regulation, habitat for biodiversity, and recycling of nutrients (Keesstra et al., 2016; Koch et al., 2013; Schulte et al., 2014; Wiesmeier et al., 2019). While the magnitude and the historical dimension of the decrease in SOC at the global level are progressively being unveiled (IPBES, 2018; Sanderman et al., 2017; Stoorvogel et al., 2017), SOC stock preservation and even increase is a major challenge for human societies in the 21st century (Amundson et al., 2015). With widespread beneficial effects on soil functioning at the local level (Pellerin et al., 2020), increasing the size of the global SOC reservoir contributes directly to the Sustainable Development Goal related to life on land (https://www.globalgoals.org/15-life-on-land, last access: 17 June 2020). It is also one of the few land-management-based intervention options that has a broad and positive impact on food security and climate change mitigation and adaptation, two other Sustainable Development Goals set by the United Nations (IPCC, 2019; Lal, 2004).

There is experimental evidence showing that in all soils, SOC is made of carbon atoms with highly contrasting residence times ranging from hours to millennia (Balesdent et al., 1987; Trumbore et al., 1989). This continuum in SOC persistence is often simplified by considering SOC as a mixture formed of several fractions, also called kinetic pools by modellers (Hénin and Dupuis, 1945; Jenkinson, 1990; Nikiforoff, 1936). The most drastic conceptual simplification of SOC persistence considers only two pools: (1) one made of young SOC with a short turnover rate (typically 3 decades on average; the active SOC pool) and (2) one made of older SOC that persists much longer in the soil (more than a century; the stable, passive, or persistent SOC pool). This dualistic representation of SOC persistence was considered "a necessary simplification, but certainly not a utopian one" 4 decades ago (Balesdent and Guillet, 1982) and is still considered meaningful (e.g. Lavallee et al., 2020). The active and stable soil organic matter pools contribute differently to the various soil functions (Hsieh, 1992). The active organic matter pool efficiently fuels soil biological activity (with carbon, nutrients, and energy) and plant growth (with nutrients) through its rapid decay, and it sustains soil structure development (Abiven et al., 2009; Janzen, 2006). Conversely, the potential contribution of a soil to climate regulation would be most dependent on its stable organic matter pool size (He et al., 2016; Shi et al., 2020).

A myriad of methods has been developed and tested to partition SOC into active and stable fractions that would match kinetic pools for the assessment of SOC dynamics and related soil functions since the second half of the 20th century (Balesdent, 1996; Hénin and Turc, 1949; Monnier et al., 1962; Poeplau et al., 2018). Some of these methods based on chemical or physical (size, density, or thermal) fractionation schemes can separate SOC fractions with, on average, different turnover rates (Balesdent, 1996; Plante et al., 2013; Poeplau et al., 2018; Trumbore et al., 1989). Of these methods, only a few are reasonably reproducible and easy to implement such as the ones based on rapid thermal analysis and chemical extractions (Gregorich et al., 2015; Poeplauet al., 2013, 2018; Soucémarianadin et al., 2018a). Other methods, such as size and density SOC fractionation, need to be inferred from machine-learning models or infrared spectroscopy to be implemented for large soil sample sets (Baldock et al., 2013; Cotrufo et al., 2019; Jaconi et al., 2019; Viscarra Rossel et al., 2019; Viscarra Rossel and Hicks, 2015; Vos et al., 2018; Zimmermann et al., 2007b). However, all SOC fractionation methods fail to achieve a proper separation of stable from active SOC, and the isolated SOC fractions are thus mixtures of centennially stable and active SOC (Fig. 1; Balesdent, 1996; Hsieh, 1992; von Lützow et al., 2007; Sanderman and Grandy, 2020). This limitation is common to all existing SOC fractionation methods and compromises the results of any work using them directly to quantify soil functions specifically related to SOC fractions or to parametrize SOC partitioning in multi-compartmental models of SOC dynamics (Luo et al., 2016). Simulations of SOC stocks changes by multi-compartmental models are very sensitive to the initial proportion of the centennially stable SOC fraction, underlining the importance of its accurate estimation (Clivot et al., 2019; Falloon and Smith, 2000; Jenkinsonet al., 1991; Taghizadeh-Toosi et al., 2020).

If the stable SOC fraction cannot be isolated, it has specific chemical and thermal characteristics: stable SOC is depleted in hydrogen and thermally stable (Barré et al., 2016; Gregorich et al., 2015). These characteristics are measurable quickly (ca. 1 h per sample) and at a reasonable cost (less than USD 60 per sample in private laboratories) using Rock-Eval® thermal analysis, and they could be of use to identify the quantitative contribution of stable SOC to total SOC. An alternative to the elusive proper separation of stable and active SOC pools could thus be to directly predict their sizes by training a machine-learning model based on Rock-Eval® data to estimate the size of the stable and active SOC fractions without isolating them from each other (Fig. 1). This model would need a training set of soil samples for which SOC partitioning into its active and stable pools can be fairly estimated. Such soil samples are available in long-term (i.e. at least longer than 3 decades) bare fallow

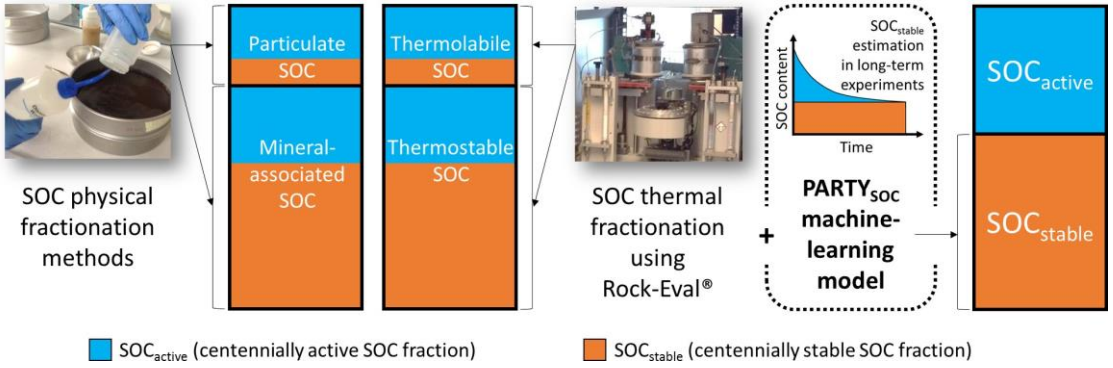


Figure 1. Conceptual representation of soil organic carbon fractionation methods vs. the $PARTY_{SOC}$ approach to quantify the size of the centennially stable and active soil organic carbon fractions. All existing soil organic carbon fractionation methods isolate fractions that are mixtures of centennially stable and active soil organic carbon. $PARTY_{SOC}$ is a machine-learning model trained on the Rock-Eval® thermal analysis data for soil samples from long-term experiments in which the size of the centennially stable SOC fraction can be estimated. When applied to the Rock-Eval® data for unknown topsoils, $PARTY_{SOC}$ partitions soil organic carbon into its active and stable fractions (i.e. without isolating soil organic carbon fractions from each other). SOC: soil organic carbon. Credits for photos: SOC physical fractionation methods, Mathilde Bryant; SOC thermal fractionation using Rock-Eval®, Lauric Cécillon.

experiments (LTBF; soils kept free of vegetation and thus with negligible SOC inputs) and long-term vegetation change (C_3 plants to C_4 plants or vice versa) experiments, as described by Balesdent et al. (1987, 2018), Barré et al. (2010), Cerri et al. (1985), and Rühlmann (1999). Cécillon et al. (2018) used this strategy to develop a machine-learning random forest regression model for topsoil samples obtained from the archives of four European long-term agricultural sites including an LTBF treatment. This model, which we propose to name $PARTY_{SOC}$, related thermal analysis parameters of topsoils measured with Rock-Eval® to their estimated proportion of the centennially stable SOC fraction (Fig. 1). This previous work positioned $PARTY_{SOC}$ as the first operational method quantifying the centennially stable and active SOC fractions in agricultural topsoils from northwestern Europe. However, the ability of this machine-learning model to fairly partition the centennially stable and the active SOC fractions of soil samples from new sites in and outside northwestern Europe is largely unknown because its training set is (1) rather limited with a low number of reference sites and (2) based on centennially stable SOC contents that are exclusively inferred from plant-free LTBF treatments.

In this study, we aimed to improve the accuracy and the genericity of the $PARTY_{SOC}$ machine-learning model that partitions SOC into its centennially stable and active fractions developed by Cécillon et al. (2018). (1) We increased the range of soil groups, soil texture classes, climates, and types of long-term experiments through the addition to the training set of topsoils from three new reference sites (two additional European long-term agricultural sites with an LTBF treatment and one South American long-term vegetation change site). (2) We integrated new predictor variables derived from Rock-Eval® thermal analysis. (3) In this second version of the model, we also changed the following series of technical details. We added a new criterion based on observed SOC content to estimate of the size of the centennially stable SOC fraction at reference sites to reduce the risk of overestimating this site-specific parameter. We calculated the proportion of the centennially stable SOC fraction differently in reference topsoil samples using SOC content estimated by Rock-Eval® rather than by dry combustion. We changed some criteria regarding the selection of reference topsoils in the training set of the model: we removed samples from agronomical treatments with compost or manure amendments, and preference was given to samples with good organic carbon yield in their Rock-Eval® thermal analysis. We better balanced the contribution of each reference site to $PARTY_{SOC}$v2.0. (4) We also aimed to build a regional version of the model restricted to the reference sites available in Europe (named $PARTY_{SOC}v2.0_{EU}$). (5) Finally, we carefully evaluated the performance of the models on unknown soils, and we further investigated the sensitivity of model performance to the training and test sets. For clarity, the main changes between the first version of $PARTY_{SOC}$ (Cécillon et al., 2018) and this second version of the model are summarized in Supplement Table S1.

2 Methods

2.1 Reference sites and estimation of the centennially stable SOC fraction content at each site

This second version of $PARTY_{SOC}$ uses seven long-term study sites as reference sites (i.e. sites where the size of the centennially stable SOC fraction can be estimated). The main characteristics of these seven reference sites, their respective soil group, and basic topsoil properties are presented in Supplement Table S2 and more thoroughly in the references cited

below. Six reference sites for $PARTY_{SOC}v2.0$ are long-term agricultural experiments located in northwestern Europe that include at least one LTBF treatment. (1) The long-term experiment on animal manure and mineral fertilizers (B3 and B4 fields) and its adjacent LTBF experiment started in 1956 and terminated in 1985 at the Lermarken site of Askov in Denmark (Christensen et al., 2019; Christensen and Johnston, 1997). (2) The static fertilization experiment (V120) started in 1902, and the fallow experiment (V505a) started in 1988 at Bad Lauchstädt in Germany (Franko and Merbach, 2017; Körschens et al., 1998; Ludwig et al., 2007). (3) The "36 parcelles" experiment started in 1959 at Grignon in France (Cardinael et al., 2015; Houot et al., 1989). (4) The "42 parcelles" experiment started in 1928 at Versailles in France (van Oort et al., 2018). (5) The Highfield bare fallow experiment started in 1959 at Rothamsted in England (Johnston et al., 2009). (6) The Ultuna continuous soil organic matter field experiment started in 1956 in Sweden (Kätterer et al., 2011). These six reference sites are used in the European version of the machine-learning model, $PARTY_{SOC}v2.0_{EU}$. One additional long-term vegetation change site completes the reference site list for the $PARTY_{SOC}v2.0$ global model. This site is a 56-year chronosequence of oil palm plantations (with C_3 plants) established on former pastures (with C_4 plants) located in South America (La Cabaña in Colombia) and sampled as a space-for-time substitution (Quezada et al., 2019).

For each reference site, data on total SOC content in topsoil (0–10 to 0–30 cm depending on the site; Supplement Table S2) were obtained from previously published studies (Barré et al., 2010; Cécillon et al., 2018; Franko and Merbach, 2017; Körschens et al., 1998; Quezada et al., 2019).

Total SOC content was measured by dry combustion with an elemental analyser (SOC_{EA}, g C kg^{-1}) according to ISO 10694 (1995) after the removal of soil carbonates using an HCl treatment for the topsoils of Grignon. For the site of La Cabaña, data on ^{13}C content (measured using an isotope ratio mass spectrometer coupled to the elemental analyser, the results being expressed in $\delta^{13}C$ abundance ratio, which is ‰ relative to the international standard) were obtained from Quezada et al. (2019), and the relative contributions of new (C_3-plant-derived) and old (C_4-plant-derived) carbon to total SOC in topsoils (0–10 cm) were calculated using Eq. (3) of the paper published by Balesdent and Mariotti (1996), as done in Quezada et al. (2019).

Based on these published data, the content of the centennially stable SOC fraction (g C kg^{-1}) at each reference site was estimated by modelling the decline of total SOC present at the onset of the experiment with time (sites with an LTBF treatment; SOC inputs are negligible in bare fallow systems) or by modelling the decline of C_4-plant-derived SOC present at the time of vegetation change with time (La Cabaña site; SOC inputs from C_4 plants are negligible after pasture conversion to oil palm plantation). For the seven reference sites, the decline in total SOC or C_4-plant-derived SOC over time had a similar shape, as shown in Barré et al. (2010), Cécillon et al. (2018), Franko and Merbach (2017), and Quezada et al. (2019), and it could be modelled using a first-order exponential decay with a constant term following Eq. (1):

$$\gamma\,(t) = ae^{-bt} + c, \quad (1)$$

where $\gamma\,(t)$ (g C kg^{-1}) is the total (sites with an LTBF treatment) or C_4-plant-derived (La Cabaña site) SOC content at time t, t (year) is the time under bare fallow (sites with an LTBF treatment) or since pasture conversion to oil palm plantation (La Cabaña site), and a, b, and c are fitting parameters. Parameter a (g C kg^{-1}) corresponds to the content of the active SOC fraction and b (yr^{-1}) is the characteristic decay rate. The parameter c (g C kg^{-1}) represents the content of theoretically inert SOC. Following Barré et al. (2010), Cécillon et al. (2018), and Franko and Merbach (2017), we considered this parameter c to be a site-specific metric of the centennially stable SOC fraction content. As already stated in Cécillon et al. (2018), in our view, the centennially stable SOC fraction is not biogeochemically inert; its mean age and mean residence time in soil are both assumed to be high(centuries) though not precisely defined here. As a result, its decline with time is negligible at the timescale of the long-term agricultural experiments and the long-term vegetation change site. We thus considered the centennially stable SOC fraction content at each experimental site to be constant. In this study, we used the centennially stable SOC fraction content already estimated by Franko and Merbach (2017) for the site of Bad Lauchstädt (on the LTBF experiment started in 1988) and by Cécillon et al. (2018) for the sites of Versailles, Grignon, Rothamsted, and Ultuna. We estimated the content of the centennially stable SOC fraction for the Askov and La Cabaña sites using the same Bayesian curve-fitting method described by Cécillon et al. (2018). The Bayesian inference method was performed using Python 2.7 and the PyMC library (Patil et al., 2010).

For the second version of $PARTY_{SOC}$, we aimed at reducing the potential bias towards an overestimation of the centennially stable SOC fraction content at reference sites using Eq. (1) (Supplement Table S1). This overestimation is possible at reference sites with an LTBF treatment, as SOC inputs to bare fallow topsoils are low but not null (e.g. Jenkinson and Coleman, 1994; Petersen et al., 2005). Similarly, C_4-plant-derived SOC inputs are possible after conversion to C_3 plants at the site of La Cabaña. We thus used the lowest observed total (sites with an LTBF treatment) or C_4-plant- derived (La Cabaña site) topsoil SOC content value as the best estimate of the centennially stable SOC fraction content at reference sites where this measured value was lower than the fitted value of the site-specific parameter c in Eq. (1).

2.2 Rock-Eval® thermal analysis of topsoil samples available from reference sites

Surface soil samples (0–10 to 0–30 cm depending on the site; see Supplement Table S2) were obtained from the seven reference sites described in Sect. 2.1. As described in Cécillon et al. (2018), the first version of the PARTY$_{SOC}$ model was based on a set of 118 topsoil samples corresponding to time series obtained from the soil archives of the sites of Rothamsted (12 samples from the LTBF treatment and 8 samples from the adjacent long-term grassland treatment), Ultuna (23 samples from the LTBF treatment and 11 samples from the associated long-term cropland treatments), Grignon (12 samples from the LTBF treatment, 6 samples from the LTBF plus straw amendment treatment, and 6 samples from the LTBF plus composted straw amendment treatment), and Versailles (20 samples from the LTBF treatment and 20 samples from the LTBF plus manure amendment treatment). All 118 topsoil samples were previously analysed using Rock-Eval® thermal analysis (Cécillon et al., 2018).

For the second version of the machine-learning model, 78 additional topsoil samples were provided by managers of the three new reference sites. A total of 35 topsoil samples were obtained from the soil archives of the Askov site (19 samples corresponding to different dates of the LTBF treatment and 16 samples corresponding to different dates of the associated long-term cropland treatments). A total of 27 topsoil samples were obtained from the soil archives of the Bad Lauchstädt site (8 samples from two dates of the mechanical LTBF treatment, 8 samples from two dates of the chemical LTBF treatment, and 11 samples from two dates of several long-term cropland treatments of the static fertilization experiment, with 8 of the latter coming from treatments with manure applications). A total of 16 topsoil samples were obtained from the site of La Cabaña (13 samples from different C_3-plant oil palm fields planted at different dates and three samples from different long-term C_4-plant pastures).

The 78 additional topsoil samples from Askov, Bad Lauchstädt, and La Cabaña were analysed using the same Rock-Eval® 6 Turbo device (Vinci Technologies, France; see Behar et al., 2001, for a description of the apparatus) and the same setup as the one used for the sample set in the first version of PARTY$_{SOC}$, described by Cécillon et al. (2018). Briefly, ca. 60 mg of ground (< 250 µm) topsoil samples were subjected to sequential pyrolysis and oxidation phases. The Rock-Eval® pyrolysis phase was carried out in an N_2 atmosphere (3 min isotherm at 200 °C followed by a temperature ramp from 200 to 650 °C at a heating rate of 30 °C min^{-1}). The Rock-Eval® oxidation phase was carried out in a laboratory air atmosphere (1 min isotherm at 300 °C followed by a temperature ramp from 300 to 850 °C at a heating rate of 20 °C min^{-1} and a final 5 min isotherm at 850 °C). Each Rock-Eval® analysis generated five thermograms corresponding to volatile hydrocarbon effluent (HC_PYR thermogram), CO (CO_PYR thermogram), and CO_2 (CO2_PYR thermogram) measured each second during the pyrolysis phase and to CO (CO_OX thermogram) and CO_2 (CO2_OX thermogram) measured each second during the oxidation phase (Behar et al., 2001).

A series of Rock-Eval® parameters was calculated from these five thermograms. For each thermogram, five temperature parameters (all in °C) were retained: T10, T30, T50, T70, and T90, which respectively represent the temperatures corresponding to the evolution of 10 %, 30 %, 50 %, 70 %, and 90 % of the total amount of evolved gas. The calculation of Rock-Eval® temperature parameters was performed using different intervals of integration depending on the maximum oven temperatures of 650 °C (HC_PYR thermogram), 560 °C (CO_PYR and CO2_PYR thermograms), 850 °C (CO_OX thermogram), and 611 °C (CO2_OX thermogram). These intervals of integration prevented any interference by inorganic carbon from most soil carbonates, and they ensured comparability with previous studies (Barré et al., 2016; Cécillon et al., 2018; Poeplau et al., 2019; Soucémarianadin et al., 2018b). Automatic baseline correction (as calculated by the software of the Rock-Eval® apparatus; Vinci Technologies, France) was performed for all thermograms but the CO_PYR and the CO2_PYR thermograms. This correction can yield some negative values for theCO_PYR and CO2_PYR thermograms of soil samples with very low SOC content (data not shown). For the HC_PYR thermogram we also determined three parameters reflecting a proportion of thermally resistant or labile hydrocarbons: a parameter representing the proportion of hydrocarbons evolved between 200 and 450 °C (thermolabile hydrocarbons, TLHC index, unitless; modified from Saenger et al., 2013, 2015), as described by Cécillon et al. (2018); a parameter representing the preservation of thermally labile hydrocarbons (*I* index, unitless; after Sebag et al., 2016); and a parameter representing the proportion of hydrocarbons thermally stable at 400 °C (*R* index, unitless; after Sebag et al., 2016). We also considered the hydrogen index (HI, mg HC g^{-1} C) and oxygen index (OI$_{RE6}$, mg O_2 g^{-1} C) that respectively describe the relative elemental hydrogen and oxygen enrichment of soil organic matter (see e.g. Barré et al., 2016). These 30 Rock-Eval® parameters are not directly related to total SOC content and were all included in the first version of the PARTY$_{SOC}$ model developed by Cécillon et al. (2018).

In this second version of PARTY$_{SOC}$, we considered 10 additional Rock-Eval® parameters as possible predictors, some of these being directly linked to SOC content (Supplement Table S1). These 10 parameters were calculated for all 196 topsoil samples available from the seven reference sites. They included the content of SOC as determined by Rock-Eval® (TOC$_{RE6}$, g C kg^{-1}); the content of soil inorganic carbon as determined by Rock-Eval® (MinC, g C kg^{-1}); the content of SOC evolved as HC, CO, or CO_2 during the pyrolysis phase of Rock-Eval® (PC, g C kg^{-1});

the content of SOC evolved as HC during the temperature ramp (200–650 °C) of the pyrolysis phase of Rock-Eval® (S2, g C kg^{-1}); the content of SOC that evolved as HC, CO, or CO2 during the first 200 s of the pyrolysis phase (at ca. 200 °C) of Rock-Eval® (PseudoS1, g C kg^{-1}; after Khedim et al., 2021); the ratio of PseudoS1 to PC (PseudoS1 / PC, unitless); the ratio of PseudoS1 to TOC_{RE6} (PseudoS1 / TOC_{RE6}, unitless); the ratio of S2 to PC (S2 / PC, unitless; after Poeplau et al., 2019); the ratio of PC to TOC_{RE6} (PC / TOC_{RE6}, unitless); and the ratio of HI to OI_{RE6} (HI / OI_{RE6}, mg HC mg^{-1} O_2). TOCRE6, MinC, PC, HI, and OI_{RE6} were obtained as default parameters from the software of the Rock-Eval® apparatus (Vinci Technologies, France). All other Rock-Eval® parameters were calculated from the integration of the five thermograms using R version 4.0.0 (R Core Team, 2020; RStudio Team, 2020) and functions from the R packages hyperSpec (Beleites and Sergo, 2020), pracma (Borchers, 2019), and stringr (Wickham, 2019).

2.3 Determination of the centennially stable SOC fraction proportion in topsoil samples from the reference sites

Following the first version of $PARTY_{SOC}$ (Cécillon et al., 2018), the proportion of the centennially stable SOC fraction in a topsoil sample of a reference site was calculatedas the ratio of the site-specific centennially stable SOC fraction content (see Sect. 2.1) to the SOC content of this particular sample. We thus assume that the centennially stable SOC fraction content in topsoils is the same in the various agronomical treatments of a reference site and that it remains constant within the time period studied at each site.

While for the first version of $PARTY_{SOC}$, the proportion of the centennially stable SOC fraction in reference topsoils was inferred using SOC contents determined by elemental analysis (SOC_{EA}), in this second version, we preferred the SOC content determined by Rock-Eval® (Table S1). The reason behind this choice was to link the Rock-Eval® parameters measured in a reference topsoil sample to an inferred proportion of the centennially stable SOC fraction that better reflected the organic carbon that actually evolved during its Rock-Eval® analysis. This choice was possible for reference topsoil samples for which Rock-Eval® analyses showed a good organic carbon yield (TOC_{RE6} divided by SOC_{EA} and multiplied by 100). This is generally the case for most soils, with typical organic carbon yields from Rock-Eval® ranging from 90 to 100 % SOC_{EA} (Disnar et al., 2003). For the topsoils of the sites of Grignon, Rothamsted, Ultuna, and Versailles used in the first version of $PARTY_{SOC}$, the organic carbon yield from Rock-Eval® was greater than 96 % (linear regression model, $R^2 = 0.97$, $n = 118$; Cécillon et al., 2018). Similarly, Rock-Eval® analyses of topsoil samples from the site of La Cabaña showed very good organic carbon yields (95 % on average, linear regression model $R^2 = 0.95$, $n = 16$). For these five reference sites (corresponding to 134 reference topsoil samples), we thus used the Rock-Eval® parameter TOC_{RE6} as a measure of the SOC content of topsoil samples to calculate their respective proportion of the centennially stable SOC fraction. Conversely, Rock-Eval® analyses of topsoil samples from the sites of Askov and Bad Lauchstädt showed moderate organic carbon yields (90 % on average for topsoils of Askov, with a noisy linear regression model, $R^2 = 0.68$, $n = 30$; 92 % on average for topsoils of Bad Lauchstädt, yet with a very good linear regression model, $R^2 = 0.96$, $n = 11$). Using the total carbon measured by Rock-Eval® (i.e. the sum of TOC_{RE6} plus MinC Rock-Eval® parameters) as an estimate of the SOC content of top-soil samples for these two sites – that are not carbonated – increased the organic carbon yield of Rock-Eval® analyses (96 % on average at Askov, still with a noisy linear regression model, $R^2 = 0.66$, $n = 30$; 101 % on average at Bad Lauchstädt, with a very good linear regression model, $R^2 = 0.95$, $n = 11$). For the two reference sites of Askov and Bad Lauchstädt (corresponding to 62 topsoil samples), we thus used the sum of Rock-Eval® parameters TOC_{RE6} plus MinC as a measure of the SOC content of topsoil samples to calculate their proportion of the centennially stable SOC fraction.

The uncertainty in the proportion of the centennially stable SOC fraction was calculated using Eq. (6) in the paper published by Cécillon et al. (2018), propagating the uncertaintiesin SOC content data (using a standard error of 0.5 g C kg^{-1} following Barré et al., 2010) and in the site-specific contents of the centennially stable SOC fraction (see above and Table 1).

2.4 Selection of the training set and of meaningful Rock-Eval® predictor variables for $PARTY_{SOC}$v2.0

In machine learning, the selection of the model training and test sets influences the performance of the model, just like the selection of the predictor variables: here, the Rock-Eval® parameters (e.g. Cécillon et al., 2008; Wehrens, 2020).

For this second version of $PARTY_{SOC}$, we changed some criteria regarding the inclusion of the available reference topsoil samples in the training set of the model (Supplement Table S1). We excluded from the training set all the topsoil samples experiencing agronomical treatments that may have changed the site-specific content of the centennially stable SOC fraction, in contradiction to our hypothesis of a constant content of this fraction at each reference site (see Sect. 2.3). These agronomical treatments concern the repeated application of some types of exogenous organic matter such as compost or manure, which we suspect may increase the content of the centennially stable SOC fraction after several decades. Therefore, we excluded all reference topsoil samples from plots that experienced repeated applications of composted straw (six samples from Grignon) or manure (20 samples

Table 1. Main statistics for soil organic carbon contents, site-specific contents of the centennially stable SOC fraction, and resulting proportions of centennially stable SOC fraction in topsoils of the seven reference sites used as the training sets for $PARTY_{SOC}v2.0$ and $PARTY_{SOC}v2.0_{EU}$. More details on agronomical treatments and sampling year of reference topsoil samples are provided in Supplement Table S3. Abbreviations are as follows. SOC: soil organic carbon; LTBF: long-term bare fallow; min: minimum; max: maximum; SD: standard deviation.

Reference site (country)	Treatments (number of samples)	SOC content of the reference soil samples ($g\,C\,kg^{-1}$) mean (min, max, SD) measurement method	Centennially stable SOC fraction content ($g\,C\,kg^{-1}$) mean (SD) estimation method	Proportion of the centennially stable SOC fraction (unitless) mean (min, max, SD)
Versailles (France)	LTBF ($n = 15$)	10.4 (5.6, 17.9, 3.9) TOC_{RE6}	5.50 (0.50) Lowest SOC_{EA} measured on-site	0.60 (0.31, 0.98, 0.20)
Rothamsted (England)	Grassland ($n = 7$)	28.3 (12.2, 41.5, 10.1)	9.72 (0.50) Lowest SOC_{EA} measured on-site	0.40 (0.23, 0.80, 0.18)
	LTBF ($n = 8$)	TOC_{RE6}		
Ultuna (Sweden)	Cropland ($n = 3$; + straw $n = 8$)	15.2 (10.0, 20.3, 2.8)	6.95 (0.88) Bayesian curve fitting	0.47 (0.34, 0.70, 0.09)
	LTBF ($n = 4$)	TOC_{RE6}		
Grignon (France)	LTBF ($n = 12$, + straw $n = 3$)	11.5 (8, 14.3, 1.7) TOC_{RE6}	7.12 (1.00) Bayesian curve fitting	0.63 (0.50, 0.89, 0.10)
Askov (Denmark)	Cropland ($n = 7$) LTBF ($n = 8$)	13.8 (11.1, 16.8, 1.9) TOC_{RE6}+MinC	5.10 (0.88) Bayesian curve fitting	0.38 (0.30, 0.46, 0.05)
Bad Lauchstädt (Germany)	Cropland ($n = 1$) LTBF ($n = 14$)	18.0 (16.8, 19.4, 0.6) TOC_{RE6}+MinC	15.00 (0.50) Lowest SOC_{EA} measured on-site	0.84 (0.77, 0.89, 0.03)
La Cabaña (Colombia)	Pasture ($n = 3$) Oil palm plantation ($n = 12$)	17.8 (10.2, 31.8, 5.7) TOC_{RE6}	4.75 (0.50) Lowest SOC_{EA} measured on-site	0.29 (0.15, 0.47, 0.10)
Reference soil sample set of $PARTY_{SOC}v2.0$ ($n = 105$)		16.4 (5.6, 41.5, 7.3)		0.52 (0.15, 0.98, 0.21)
Reference soil sample set of $PARTY_{SOC}v2.0_{EU}$ ($n = 90$)		16.2 (5.6, 41.5, 7.5)		0.55 (0.23, 0.98, 0.20)

from Versailles and 8 samples from Bad Lauchstädt) from the training set of the model. Yet, we kept some reference topsoil samples from Grignon and Ultuna experiencing repeated applications of straw.

We also excluded from the training set of the model the reference topsoil samples for which the organic carbon yield from Rock-Eval® is below 86 % or above 116 %. For the site of Askov, with a noisy relationship between SOC_{EA} and the sum TOC_{RE6} plus MinC (see Sect. 2.3), we excluded the five samples without an SOC_{EA} measurement preventing the calculation of the organic carbon yield from their Rock-Eval® analysis. Conversely, for the site of Bad Lauchstädt we kept topsoil samples without available SOC_{EA} measurements, as the linear relationship between SOC_{EA} and the sum TOC_{RE6} plus MinC was very good for this site (see Sect. 2.3). Thesecriteria regarding the organic carbon yield from Rock-Eval®lead to the exclusion of nine samples from the site of Askov,four additional samples from the site of Versailles, and two from the site of Ultuna.

Contrary to the first version of $PARTY_{SOC}$, this second version is based on a balanced contribution of each reference site (Supplement Table S1). Each reference site contributes to the model with 15 samples so that the reference sample set of $PARTY_{SOC}v2.0$ is composed of 105 topsoil samples (90 for the European version of the model $PARTY_{SOC}v2.0_{EU}$). Besides the above-mentioned exclusion criteria (that excluded 49 of the 196 topsoil samples available from the seven reference sites), the 15 topsoil samples retained for each referencesite were selected (1) to have a range of the proportion of the centennially stable SOC fraction as wide as possible and (2) to have the best organic carbon yield from Rock-Eval® analysis. On average, the organic carbon yield of the Rock-Eval® analyses for the retained training set of reference top- soil samples (calculated as described above) was greater than 98 % SOC_{EA} ($SOC_{DETERMINED\_BY\_ROCK\text{-}EVAL®} = 0.9924\ SOC_{EA} - 0.1051$, $R^2 = 0.99$, $n = 91$ topsoil samples with available SOC_{EA} measurements). The list of the 105

reference topsoil samples retained as the training set for PARTY$_{SOC}$v2.0 is provided in Table S3. This list includes, for each reference topsoil sample, information on its reference site, land cover, agronomical treatment, sampling year, and values for the 40 Rock-Eval® parameters.

The 40 Rock-Eval® parameters calculated (see Sect. 2.2) captured most of the information related to SOC thermal stability, elemental stoichiometry, and content that is contained in the five Rock-Eval® thermograms. However, not all Rock-Eval® parameters necessarily carry meaningful information for partitioning SOC into its centennially stable and active fractions (Cécillon et al., 2018). PARTY$_{SOC}$v2.0 and its European version PARTY$_{SOC}$v2.0$_{EU}$ incorporate as predictor variables only the Rock-Eval® parameters showing a strong relationship with the proportion of the centennially stable SOC fraction (Supplement Table S1). The absolute value of 0.50 for the Spearman's ρ (nonparametric and nonlinear correlation test) was used as a threshold to select meaningful Rock-Eval® predictor variables (calculated from the reference topsoil sample set for the PARTY$_{SOC}$v2.0 model, n = 105). Basic statistics of all Rock-Eval® parameters (training set for PARTY$_{SOC}$v2.0) are reported in Supplement Table S4.

2.5 Random forest regression models to predict the proportion of the centennially stable SOC fraction from Rock-Eval® parameters, performance assessment, and error propagation in the models

The PARTY$_{SOC}$v2.0 machine-learning model consists of a nonparametric and nonlinear multivariate regression model relating the proportion of the centennially stable SOC fraction (response vector or dependent variable y) of the reference soil sample set (n = 105 topsoil samples from the sevenreference sites; see Sect. 2.4) to their Rock-Eval® parameters summarized by a matrix of predictor variables (X) made up of the selected centered and scaled Rock-Eval® parameters. As stated above, we also built a regional (European) version of the model based on the six European reference sites only (PARTY$_{SOC}$v2.0$_{EU}$, using the 90 reference topsoil samples from Askov, Bad Lauchstädt, Grignon, Rothamsted, Ultuna, and Versailles).

Like the first version of PARTY$_{SOC}$, this second version uses the machine-learning algorithm of random forests-random inputs (hereafter termed random forests) proposed by Breiman (2001). This algorithm aggregates a collection of random regression trees (Breiman, 2001; Genuerand Poggi, 2020). PARTY$_{SOC}$v2.0 and its European version PARTY$_{SOC}$v2.0$_{EU}$ are based on a forest of 1000 different regression trees made of splits and nodes. The algorithm of random forests combines bootstrap resampling and random variable selection. Each of the 1000 regression trees was grown on a bootstrapped subset of the reference topsoil sample set (i.e. containing ca. two-thirds of "in-bag" samples). The algorithm randomly sampled one-third of the selected Rock- Eval® parameters (see Sect. 2.4) as candidates at each split of the regression tree, and it used a minimum size of terminal tree nodes of five topsoil samples. The relative importance (i.e. ranking) of each selected Rock-Eval® parameter in the regression models was computed as the unscaled permutation accuracy (Strobl et al., 2009).

The performance of PARTY$_{SOC}$v2.0 and PARTY$_{SOC}$v2.0$_{EU}$ was assessed by statistical metrics comparing the predicted vs. the estimated values of their reference topsoil sample set using three complementary validation procedures. First, the predictive ability of both models was assessed by an "internal" procedure that used their respective whole reference topsoil sample sets (n = 105 samples for PARTY$_{SOC}$v2.0, n = 90 samples for PARTY$_{SOC}$v2.0$_{EU}$). For this procedure, performance statistics were calculated only for the "out-of-bag" topsoil samples of the whole reference sets using a random seed of 1 to initialize the pseudorandom number generator of the R software. Out-of-bag samples are observations from the training set not used for a specific regression tree that can be used as a "built-in" test set for calculating its prediction accuracy (Strobl et al., 2009). Second, the predictive ability of the models was assessed by a "random splitting" procedure that randomly split their respective reference topsoil samplesets into a test set (made of n = 30 samples) and a training set (n = 75 samples for PARTY$_{SOC}$v2.0, n = 60 samples for PARTY$_{SOC}$v2.0$_{EU}$). This procedure was repeated 15 times using random seeds from 1 to 15 in the R software. Third, a fully independent "leave-one-site-out" procedure was used to assess the predictive ability of the models. This procedure successively excludes topsoil samples of one reference site from the training set and uses them as a test set (n = 15) for the models. It used the random seed of 1 in the R software. For the second and third procedures, performance statistics were calculated (1) for the out-of-bag topsoil samples of the training sets and (2) for the topsoil samples of the test sets. The leave-one-site-out validation should be seen as the procedure giving the most accurate estimation of the uncertainty of both regression models for unknown topsoil samples.

Finally, we assessed the sensitivity of model performance to the training and the test sets. For both sensitivity analyses, only the leave-one-site-out validation procedure was used (based exclusively on independent training and test sets). First, model sensitivity to the training set was assessed as its sensitivity to the independent reference sites included in the training set. It was performed successively using, as examples, two different test sets consisting of independent soils from the reference sites of Grignon and Versailles. Several random forest regression models were built using, as training sets, combinations of topsoil samples from a decreasing number of the remaining reference sites on the basis of their potential proximity to the topsoil samples of the test sets regarding their pedological or climatic conditions. The size of the various training sets ranged from n = 90 samples (six reference sites) to

n = 30 samples (only two reference sites). Second, model sensitivity to the test set was assessed as its sensitivity to independent test samples (1) from a reference soil group (FAO, 2014) not existing in the training set (i.e. excluding Chernozem soil samples from the test set) (2) that are unlikely to be encountered in agricultural soils (i.e. excluding from the test set soils sampled at late dates of bare fallow treatments more than 25 years after the experiment onset, which cannot represent soils with regular carbon input). Model sensitivity to the test set was performed only for $PARTY_{SOC}v2.0_{EU}$ to further investigate its predictive abilityfor soil samples from independent Cambisols and Luvisols of northwestern Europe.

Several statistics were used to assess the predictive ability of the regression models. The coefficient of determination, R^2_{OBB}, was calculated for the out-of-bag samples of the training set and R^2 was calculated for the samples of the test set. The root mean square error of prediction, $RMSEP_{OOB}$, was calculated for the out-of-bag samples of the training set, and RMSEP was calculated for the samples of the test set. The relative RMSEP, $_{R}$RMSEP, was calculated as the ratio of the RMSEP to the mean value of the test set. The ratio of performance to interquartile range (RPIQ) was calculated as the ratio of the interquartile range of the test set (Q3–Q1, which gives the range accounting for 50 % of the test set around its median value) to the RMSEP (Bellon-Maurel et al., 2010). The bias of the random forest regression models was calculated as the mean of the model predictions for the test set minus the actual mean of the test set. Additionally, site-specificRMSEP and $_{R}$RMSEP were calculated for the leave-one-site-out procedure (with the 15 independent test samples from each site). The uncertainty in the model predictions for new topsoils was determined using a methodology that was fullydescribed by Cécillon et al. (2018). This methodology was adapted after the work of Coulston et al. (2016) to explicitly take into account the uncertainty in the reference values of the proportion of the centennially stable SOC fraction (see Sect. 2.3) that were used to build the models (Cécillon et al.,2018).

$PARTY_{SOC}v2.0$ and $PARTY_{SOC}v2.0_{EU}$ were programmed as R scripts in the RStudio environment software (RStudio Team, 2020) and were run using the R version 4.0.0 (R Core Team, 2020). The R scripts use the random forest algorithm of the randomForest R package (Liaw and Wiener, 2002) and the boot R package for bootstrapping (Canty and Ripley, 2020; Davison and Hinkley, 1997).

3 Results

3.1 Content of the centennially stable SOC fraction at the reference sites

The two newly fitted values of the centennially stable SOC fraction content (i.e. parameter c in Eq. 1; see Sect. 2.1) were 5.10 g C kg^{-1} at the site of Askov (SD = 0.88 g C kg^{-1}) and 5.12 g C kg^{-1} at the site of La Cabaña (SD = 0.35 g C kg^{-1}). The fitted values of parameter c in Eq. (1) for all reference sites and their standard errors are provided in Supplement Table S2. A total (reference sites with an LTBF treatment) or a C_4-plant-derived (La Cabaña site) SOC content value lower than the fitted value of the site-specific parameter c in Eq. (1) was measured at four of the seven reference sites for the $PARTY_{SOC}v2.0$ model. At Bad Lauchstädt, an SOC_{EA} value of 15.0 g C kg^{-1} was reported by Körschens et al. (1998) for topsoils of the well ring experiment (Ansorge, 1966). At Rothamsted, an SOC_{EA} measurement of 9.72 g C kg^{-1} was reported for topsoils of the Highfield LTBF experiment by Cécillon et al. (2018). At Versailles, an SOCEA measurement of 5.50 g C kg^{-1} was reported after 80 years of bare fallow by Barré et al. (2010). At La Cabaña, a C_4-plant-derived SOC content of 4.75 g C kg kg^{-1} was calculated using data from Quezada et al. (2019). These values did not differ strongly from the values of the centennially stable SOC contents calculated from the Bayesian curve-fitting method (Tables 1, S2). In particular, the hierarchy in the centennially stable SOC content of the seven reference sites was unchanged whatever the calculation method. These values were retained as the best estimates of the site-specific content of the centennially stable SOC fraction in topsoils of the four sites to reduce the risk of overestimating the actual value of the centennially stable SOC content compared to the first published version of the model (see Sect. 2.1; Tables 1 and S1). As these site-specific values of the centennially stable SOC fraction content were derived from SOC_{EA} measurements, we attributed a standard deviation of 0.50 g C kg^{-1} to each of them following Barré et al. (2010). The final estimates of the content of the centennially stable SOC fraction at the seven reference sites that were used in $PARTY_{SOC}v2.0$ are provided in Table 1. They varied by a factor of 3 across the reference sites, ranging from 4.75 g C kg^{-1} at La Cabaña to 15.00 g C kg^{-1} at Bad Lauchstädt. The lowest value of the topsoil content of the centennially stable SOC fraction used in $PARTY_{SOC}v2.0_{EU}$ differed only slightly from the one of $PARTY_{SOC}v2.0$ (5.10 g C kg^{-1} at the site of Askov).

3.2 Content and biogeochemical stability of SOC in the training sets and selection of meaningful Rock-Eval® parameters as predictor variables for the $PARTY_{SOC}v2.0$ and $PARTY_{SOC}v2.0_{EU}$ models

The SOC content in the topsoil samples of the seven reference sites ranged from 5.6 to 41.5 g C kg^{-1} in the training sets for the $PARTY_{SOC}v2.0$ (n = 105) and $PARTY_{SOC}v2.0_{EU}$ (n = 90) models (Table 1). As shown in Table 1, this resulted in proportions of the centennially stable SOC fraction ranging from 0.15 to 0.98 ($PARTY_{SOC}v2.0$ training set) and from 0.23 to 0.98 ($PARTY_{SOC}v2.0_{EU}$ training set). All 25 calculated Rock-Eval® temperature parameters showed positive values of Spearman's ρ coefficient with the response variable of the

$PARTY_{SOC}$v2.0 model (n = 105; with Spearman's ρ values up to 0.81 for $T90_{HC\_PYR}$; Table 2). While the inorganic carbon content was not correlated with the proportion of the centennially stable SOC fraction, TOC_{RE6} was significantly and negatively correlated with the response variable of the $PARTY_{SOC}$v2.0 model (Spearman's ρ = 0.55; Table 2). Other Rock-Eval® parameters linked to soil carbon content showed a stronger relationship than TOC_{RE6} with the proportion of the centennially stable SOC fraction. This was the case for S2 and PC that showed the highest absolute Spearman's ρ coefficients, with a highly significant negative relationship (Spearman's ρ = 0.85; Table 2). A total of 18 of the 40 calculated Rock-Eval® parameters showed an absolute value of Spearman's ρ above 0.5 with the proportion of the centennially stable SOC fraction in the training set of the $PARTY_{SOC}$v2.0 model (n = 105; Table 2) and were thus retained as predictor variables for the models. The 18 Rock-Eval® parameters retained were the Rock-Eval® temperature parameters $T70_{HC\_PYR}$, $T90_{HC\_PYR}$, $T30_{CO2\_PYR}$, $T50_{CO2\_PYR}$, $T70_{CO2\_PYR}$, $T90_{CO2\_PYR}$, $T70_{CO\_OX}$, $T50_{CO2\_OX}$, $T70_{CO2\_OX}$, and $T90_{CO2\_OX}$ and the Rock-Eval® parameters PseudoS1, S2, S2/PC, HI, HI/OI_{RE6}, PC, PC/TOC_{RE6}, and TOC_{RE6}.

3.3 Performance assessment of the $PARTY_{SOC}$v2.0 and $PARTY_{SOC}$v2.0$_{EU}$ machine-learning models

Using both the internal and the random splitting performance assessment procedures (see Sect. 2.5), the $PARTY_{SOC}$v2.0 and $PARTY_{SOC}$v2.0$_{EU}$ models showed good to very good predictive ability for the proportion of the centennially stable SOC fraction (Fig. 2a; Table 3a). For most of the calculated statistics, the European version of the model $PARTY_{SOC}$v2.0$_{EU}$ showed better performance than the $PARTY_{SOC}$v2.0 model (Table 3). Using the random splitting procedure, the mean R^2 of $PARTY_{SOC}$v2.0$_{EU}$ was 0.87 (0.81 for $PARTY_{SOC}$v2.0); its RMSEP and $_{R}$RMSEP were respectively 0.07 and 0.13 (0.09 and 0.17 for $PARTY_{SOC}$v2.0), and its mean RPIQ was 4.6 (3.6 for $PARTY_{SOC}$v2.0). The bias was low for both models (Table 3a).

The predictive ability of both models decreased when assessed using the leave-one-site-out procedure (see Sect. 2.5; Fig. 2b). Again, $PARTY_{SOC}$v2.0$_{EU}$ showed better performance statistics than the $PARTY_{SOC}$v2.0 model (Table 3; Fig. 2b), with an R^2 of 0.45, an RMSEP of 0.15, an $_{R}$RMSEP of 0.27, and an RPIQ of 2.4. The $PARTY_{SOC}$v2.0 model poorly predicted the proportion of the centennially stable SOC fraction in topsoil samples of two sites (Table 3b; Fig. 2b): La Cabaña (overestimation; with a site-specific RMSEP of 0.28) and Bad Lauchstädt (underestimation; with a site-specific RMSEP of 0.32). The proportion of the centennially stable SOC fraction in topsoil samples of Bad Lauchstädt remained underestimated by the $PARTY_{SOC}$v2.0$_{EU}$ model, though with a reduced site-specific RMSEP (0.23; Table 3b; Fig. 2b). All other site-specific RMSEPs were below 0.18 (0.17 at Versailles for $PARTY_{SOC}$v2.0, 0.18 at Grignon for $PARTY_{SOC}$v2.0$_{EU}$; Table 3b), with remarkably low site-specific RMSEPs for the sites of Askov (below 0.05 for both models) and Ultuna (0.06 for $PARTY_{SOC}$v2.0; 0.09 for $PARTY_{SOC}$v2.0$_{EU}$).

The most important Rock-Eval® parameter for predicting the proportion of the centennially stable SOC fraction is S2 for both $PARTY_{SOC}$v2.0 and $PARTY_{SOC}$v2.0$_{EU}$ (Table 2). Conversely, the two models show only two Rock-Eval® parameters in common of their five most important ones: S2, PC, PC / TOC_{RE6}, $T70_{CO2\_OX}$, and $T90_{HC\_PYR}$ for $PARTY_{SOC}$v2.0 and S2, $T50_{CO2\_PYR}$, PC, S2 / PC, and HI / OI_{RE6} for $PARTY_{SOC}$v2.0$_{EU}$ (Table 2).

3.4 Sensitivity of model performance to the training and test sets

The sensitivity analysis to the training set showed that restricting the model training set to samples from fewer reference sites with pedoclimatic conditions closer to the ones of a fully independent test site changed its performance (Fig. 3). Removing from the training set a reference site with a climate (i.e. La Cabaña) or a soil group (i.e. Bad Lauchstädt) differing strongly from the independent test sites (here, Grignon and Versailles used as examples) reduced the site-specific RMSEP and $_{R}$RMSEP of the model (Supplement Table S5). When Grignon or Versailles were used as independent test sites, the model with the best predictive ability (i.e. the lowest site-specific RMSEP and $_{R}$RMSEP) used a training set composed of 45 topsoil samples from three European reference sites (including the French site with the closest climate, despite its different soil group; Supplement Tables S2 and S5; Fig. 3).

The sensitivity analysis to the test set showed that when excluding Chernozem samples from the test set (i.e. validating the model exclusively with independent samples from Cambisols or Luvisols), the performance statistics of $PARTY_{SOC}$v2.0$_{EU}$ were improved (leave-one-site-out validation procedure: R^2 of 0.56; RMSEP of 0.13; n = 75). The further removal of independent test soils that are unlikely to be encountered in agricultural Cambisols and Luvisols (soils sampled at late dates of bare fallow treatments more than 25 years after the experiment onset) also improved the performance statistics of $PARTY_{SOC}$v2.0$_{EU}$ (Supplement Fig. S1; leave-one-site-out validation procedure: R^2 of 0.71; RMSEP of 0.11; n = 58).

4 Discussion

The second version of the $PARTY_{SOC}$ machine-learning model incorporates a large number of modifications and improvements (Table S1), and its predictive ability was more thoroughly assessed compared to the first version of the model (Cécillon et al., 2018). The critical examination of the

Table 2. Spearman's rank correlation coefficient test between the 40 calculated Rock-Eval® parameters and the proportion of the centennially stable organic carbon fraction in the reference topsoil sample set of the PARTY$_{SOC}$v2.0 model (n = 105), with the variable importance (ranking) of the 18 selected Rock-Eval® parameters for predicting the proportion of the centennially stable SOC fraction in the PARTY$_{SOC}$v2.0 and PARTY$_{SOC}$v2.0$_{EU}$ random forest regression models. See Sect. 2.2 for a description of the units of the 40 Rock-Eval® parameters. The 18 Rock-Eval® parameters retained as predictor variables for the second version of PARTY$_{SOC}$ are shown in bold. SOC: soil organic carbon.

Rock-Eval® parameter	Spearman's ρ with with the proportion of the centennially stable SOC fraction	p value	Variable importance to predict the proportion of the centennially stable SOC fraction in the PARTY$_{SOC}$v2.0 regression model (rank)	Variable importance to predict the proportion of the centennially stable SOC fraction in the PARTY$_{SOC}$v2.0$_{EU}$ regression model (rank)
$T10_{HC\_PYR}$	0.38	0.0001	–	–
$T30_{HC\_PYR}$	0.47	0.0000	–	–
$T50_{HC\_PYR}$	0.46	0.0000	–	–
$T70_{HC\_PYR}$	**0.54**	**0.0000**	**17**	**15**
$T90_{HC\_PYR}$	**0.81**	**0.0000**	**5**	**13**
$T10_{CO\_PYR}$	0.40	0.0000	–	–
$T30_{CO\_PYR}$	0.36	0.0001	–	–
$T50_{CO\_PYR}$	0.33	0.0005	–	–
$T70_{CO\_PYR}$	0.31	0.0014	–	–
$T90_{CO\_PYR}$	0.31	0.0013	–	–
$T10_{CO2\_PYR}$	0.35	0.0003	–	–
$T30_{CO2\_PYR}$	**0.56**	**0.0000**	**12**	**10**
$T50_{CO2\_PYR}$	**0.55**	**0.0000**	**8**	**2**
$T70_{CO2\_PYR}$	**0.55**	**0.0000**	**10**	**7**
$T90_{CO2\_PYR}$	**0.58**	**0.0000**	**11**	**11**
$T10_{CO\_OX}$	0.31	0.0013	–	–
$T30_{CO\_OX}$	0.41	0.0000	–	–
$T50_{CO\_OX}$	0.49	0.0000	–	–
$T70_{CO\_OX}$	**0.58**	**0.0000**	**9**	**16**
$T90_{CO\_OX}$	0.33	0.0007	–	–
$T10_{CO2\_OX}$	0.10	0.3349	–	–
$T30_{CO2\_OX}$	0.39	0.0000	–	–
$T50_{CO2\_OX}$	**0.63**	**0.0000**	**13**	**14**
$T70_{CO2\_OX}$	**0.70**	**0.0000**	**4**	**12**
$T90_{CO2\_OX}$	**0.60**	**0.0000**	**14**	**17**
I index	−0.40	0.0000	–	–
R index	0.47	0.0000	–	–
TLHC index	−0.49	0.0000	–	–
HI	**–0.72**	**0.0000**	**7**	**6**
OI_{RE6}	−0.09	0.3504	–	–
TOC_{RE6}	**–0.55**	**0.0000**	**6**	**9**
MinC	0.03	0.7430	–	–
PC	**–0.85**	**0.0000**	**2**	**3**
S2	**–0.85**	**0.0000**	**1**	**1**
PseudoS1	**–0.50**	**0.0000**	**18**	**18**
PseudoS1 / PC	0.28	0.0033	–	–
PseudoS1 / TOC_{RE6}	−0.06	0.5702	–	–
S2 / PC	**–0.70**	**0.0000**	**16**	**4**
PC / TOC_{RE6}	**–0.71**	**0.0000**	**3**	**8**
HI / OI_{RE6}	**–0.68**	**0.0000**	**15**	**5**

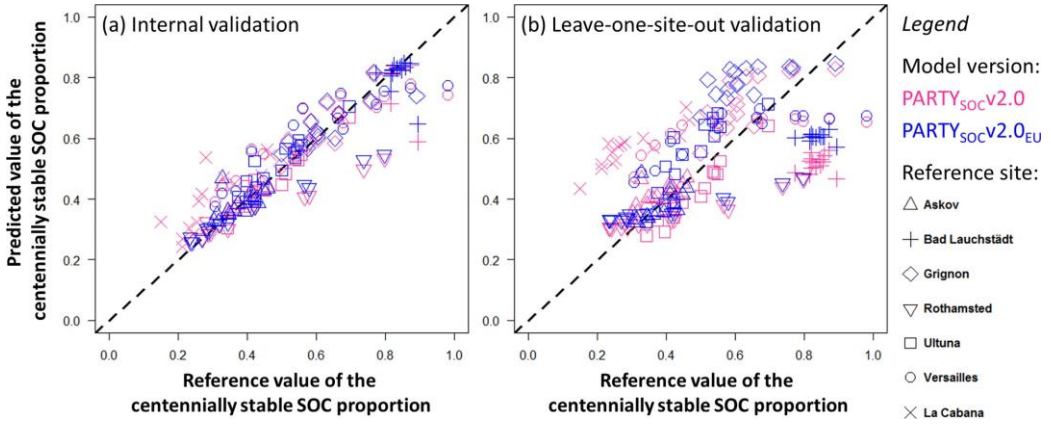


Figure 2. Performance of PARTY$_{SOC}$v2.0 and PARTY$_{SOC}$v2.0$_{EU}$ machine-learning models based on Rock-Eval® thermal analysis for predicting the centennially stable organic carbon proportion in topsoils. **(a)** Results of the internal validation procedure. **(b)** Results of the leave-one-site-out validation procedure (see Sect. 2.5 for more details on model performance assessment). SOC: soil organic carbon.

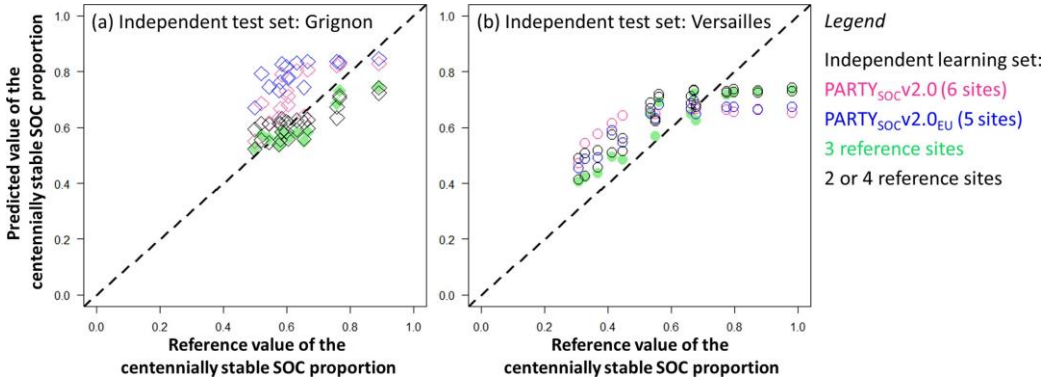


Figure 3. Sensitivity of model performance to the reference sites included in the training set using 15 topsoil samples from the sites of **(a)** Grignon or **(b)** Versailles as independent test sets. Predictions by models showing the lowest RMSEP and RRMSEP are plotted in green (using a training set composed of three independent reference sites to predict Grignon or Versailles as a test set). See Table S5 for more details on the training sets of the different models and their site-specific performance statistics. SOC: soil organic carbon.

performance of PARTY$_{SOC}$v2.0 and PARTY$_{SOC}$v2.0$_{EU}$ provides new insights (1) on the relationships between Rock-Eval® parameters and the century-scale persistence of SOC and (2) on both the current and potential capabilities of the model to partition the centennially stable and active organic carbon fraction in topsoils. Based on those insights, (3) we plan future developments of the PARTY$_{SOC}$ model to further expand its domain of application while reducing its prediction error.

4.1 Rock-Eval® chemical and thermal information is related to the century-scale persistence of SOC

The methodology used to estimate the centennially stable SOC proportion in reference topsoils has been revised for the second version of the PARTY$_{SOC}$ model (see Sect. 2.1 and 2.3 and Supplement Table S1), and the training set now integrates a wider range of centennially stable SOC contents (4.75–15.00 g C kg^{-1}) with a median value of 6.95 g C kg^{-1} (n = 7; Table 1). This range covers most of the published size estimates of this fraction in topsoils, estimated using different methods (Balesdent et al., 1988; Barré et al., 2010; Buyanovsky and Wagner, 1998b; Cécillon et al., 2018; Franko and Merbach, 2017; Hsieh, 1992; Huggins et al., 1998; Jenkinson and Coleman, 1994; Körschens et al., 1998; Rühlmann, 1999). The contribution of each reference site to the training set and the inclusion criteria for topsoil samples were also modified, and 10 Rock-Eval® parameters not considered in the first version of the model were proposed as potential predictor variables for this second version of the model (see Sect. 2.2 and 2.4 and Supplement Table S1).

Using this improved design, all Rock-Eval® temperature parameters showed positive values of Spearman's ρ coefficient with the proportion of the centennially stable SOC fraction in topsoils (Table 2), while a few of them showed counterintuitive significant negative correlations using the training set for the first version of PARTY$_{SOC}$ (Cécillon et al., 2018). This confirms the generic link between SOC ther-

Table 3. Performance of the $PARTY_{SOC}v2.0$ and $PARTY_{SOC}v2.0_{EU}$ random forest regression models based on Rock-Eval® thermal analysis for predicting the proportion of the centennially stable organic carbon fraction in topsoils. **(a)** Performance statistics calculated for the internal, random splitting (mean statistics of 15 different models), and leave-one-site-out validation procedures. **(b)** Site-specific performance statistics calculated for the leave-one-site-out validation procedure. The performance statistics and their abbreviations are defined at Sect. 2.5.

(a)	Internal procedure		Random splitting procedure		Leave-one-site-out procedure	
	$PARTY_{SOC}v2.0$	$PARTY_{SOC}v2.0_{EU}$	$PARTY_{SOC}v2.0$	$PARTY_{SOC}v2.0_{EU}$	$PARTY_{SOC}v2.0$	$PARTY_{SOC}v2.0_{EU}$
R^2_{OOB}	0.83	0.87	0.80	0.84	–	–
$RMSEP_{OOB}$	0.08	0.07	0.09	0.08	–	–
R^2	–	–	0.81	0.87	0.23	0.45
RMSEP	–	–	0.09	0.07	0.18	0.15
$_R$RMSEP	–	–	0.17	0.13	0.36	0.27
RPIQ	–	–	3.59	4.60	1.75	2.39
Bias	–	–	0.005	0.006	< 0.001	−0.003

(b)	Leave-one-site-out procedure									
Test set		Askov	Bad	Lauchstädt	Grignon	Versailles	Rothamsted	Ultuna	La Cabaña	
$PARTY_{SOC}v2.0$	Site-specific RMSEP	0.05		0.32	0.11	0.17	0.14	0.06	0.28	
	Site-specific $_R$RMSEP	0.13		0.38	0.18	0.28	0.36	0.13	0.94	
$PARTY_{SOC}v2.0_{EU}$	Site-specific RMSEP	0.05		0.23	0.18	0.14	0.14	0.09	–	
	Site-specific $_R$RMSEP	0.13		0.28	0.28	0.24	0.35	0.20	–	

mal stability and its in situ biogeochemical stability: centennially stable SOC is thermally stable, even though thermostable SOC fractions are a mixture of centennially stable and active SOC (Fig. 1; Barré et al., 2016; Gregorich et al., 2015; Plante et al., 2013; Sanderman and Grandy, 2020; Schiedung et al., 2017). Some Rock-Eval® temperature parameters were within the five most important predictor variables for both $PARTY_{SOC}v2.0$ ($T70_{CO2\_OX}$, $T90_{HC\_PYR}$) and $PARTY_{SOC}v2.0_{EU}$ ($T50_{CO2\_PYR}$; Table 2).

Contrary to the first version of $PARTY_{SOC}$, the second version tested several Rock-Eval® parameters directly linked to soil carbon content as potential predictor variables. TOC_{RE6} was selected as a meaningful predictor variable for $PARTY_{SOC}v2.0$ and $PARTY_{SOC}v2.0_{EU}$. Its negative correlation with the centennially stable SOC proportion (Table 2) was expected, according to the calculation of the latter (see Sect. 2.3). This is in line with results from SOC-dating techniques and with most multi-compartmental models of SOC dynamics, suggesting that the proportion of the most persistent SOC fraction is a decreasing function of total SOC (Huggins et al., 1998; Rühlmann, 1999). Indeed, the ex post optimized initial value of the proportion of the inert SOC fraction for the simple model of SOC dynamics (AMG) is higher (0.60 on average) for SOC-depleted temperate topsoils with a long-term arable history than for SOC-rich temperate topsoils with a long-term grassland history (0.47 on average; Clivot et al., 2019). Contrarily, the empirical function commonly used to initialize the size of the inert SOC fraction of the multi-compartmental RothC model predicts an increased proportion of inert SOC with increased total SOC (Falloon et al., 1998). This empirical function needs to be examined in light of these results.

Interestingly, S2 (pyrolyzable volatile hydrocarbon effluent) and PC (total pyrolyzable organic carbon), two other Rock-Eval® parameters linked to SOC content, showed a stronger negative relationship than TOC_{RE6} with the proportion of the centennially stable SOC fraction. Both variables are among the three most important predictor variables for $PARTY_{SOC}v2.0$ and $PARTY_{SOC}v2.0_{EU}$, while TOC_{RE6} was ranked sixth or ninth out of the 18 predictor variables (Table 2). Other Rock-Eval® parameters related to the pyrolyzable SOC fraction (PC / TOC_{RE6} and HI, both negatively related to the centennially stable SOC proportion) were also important predictor variables for both models. The results suggest that a simple decreasing function of total SOC content cannot accurately predict the centennially stable SOC proportion in topsoils according to the recent report by Clivot et al. (2019). They also confirm the generic elemental stoichiometry of the centennially stable SOC fraction in that it is consistently depleted in hydrogen (Barré et al., 2016; Gregorich et al., 2015; Poeplau et al., 2019), and they illustrate the usefulness of the pyrolysis step of Rock-Eval® thermal analysis and its volatile hydrocarbon effluent quantification to infer the proportion of the centennially stable SOC fraction in unknown topsoils.

4.2 Capability of the second version of PARTY$_{SOC}$ to partition the centennially stable and active SOC fractions

The training set for the second version of PARTY$_{SOC}$ was significantly diversified compared with the first version. It now represents wider pedoclimatic conditions (Table S2), and it includes one long-term vegetation change site as a reference site (La Cabaña). Reference topsoils from the Colombian site of La Cabaña fit well into the training set of the global model: they did not alter its overall performance, as the root mean square errors of PARTY$_{SOC}$v2.0 (internal or random splitting validation procedures) are comparable to the ones of the model's first version, in which the content of the centennially stable SOC fraction was inferred exclusively from plant-free soils (Fig. 2a, Table 3; Cécillon et al.,2018). Similarly, the expansion of the training set to new pedoclimates (Supplement Table S2) did not alter the performance of the model when assessed using the internal or random splitting validation procedures (Fig. 2a, Table 3).

The predictive ability of the second version of PARTY$_{SOC}$ was more thoroughly assessed compared to the first version of the model. Specifically, the sensitivity of model performance to the reference sites included in the training set demonstrates that local models – with training sets composed of soils from pedoclimates similar to the ones of the soils from the prediction set – showed better predictive ability for the centennially stable SOC proportion compared to a global model (Fig. 3). While the current training set is composed of too few reference sites to implement local modelling, this suggests that the European version PARTY$_{SOC}$v2.0$_{EU}$ should be preferred to the global version PARTY$_{SOC}$v2.0 when predicting the centennially stable SOC proportion in unknown soils from Europe.

On the other hand, the leave-one-site-out validation procedure, the most robust validation procedure (see Sect. 2.5), demonstrates that the second version of PARTY$_{SOC}$ is currently not capable of accurately partitioning SOC into its centennially stable and active fractions in soil samples from pedoclimates that differ strongly from the ones included in thetraining set (Fig. 2b, Table 3b). This indicates that, like all machine-learning approaches, the PARTY$_{SOC}$ model gains progressively more genericity (i.e. capability to fairly predict the centennially stable SOC proportion in unknown soils) as its training set integrates soils from new pedoclimates. In this respect, we consider applying the second version of PARTY$_{SOC}$ to unknown soils from pedoclimates outside its training set not recommended. The sensitivity analysis to the test set, however, shows that PARTY$_{SOC}$v2.0$_{EU}$ reliably partitions SOC kinetic fractions at its validation sites for Cambisols and Luvisols (with a mean prediction error in the centennially stable SOC proportion of 0.11; see Sect. 3.4 and Fig. S1). Cambisols and Luvisols are the two dominant reference soil groups in Europe, covering more than 41 % of European land areas (European Commission, 2008). Though the model test set does not include all the within-group pedological variability of Cambisols and Luvisols (FAO, 2014), this suggests that PARTY$_{SOC}$v2.0$_{EU}$ can accurately partition SOC into its centennially stable and active fractions for a significant portion of northwestern European agricultural soils. The relatively high prediction error, however, of both PARTY$_{SOC}$v2.0 and PARTY$_{SOC}$v2.0$_{EU}$ models at Rothamsted (high $_{R}$RMSEP; Table 3), a site developed on a Chromic Luvisol, may be due to an inaccurate estimate (overestimation) of the centennially stable SOC content at this site. Indeed, a report from an ancient LTBF trial at Rothamsted (drain gauge experiment; Jenkinson and Coleman, 1994), on the same soil unit as the Highfield bare fallow experiment, showed a measured total SOC content of 7.9 g C kg^{-1}, which is lower than our current estimate of the centennially stable SOC content (9.72 g C kg^{-1}; Table 1). Yet, the conditions of the drain gauge experiment, with a basic soil pH value of 7.9 due to heavy dressing of chalk on Rothamsted's arable lands before the 19th century (Avery and Catt, 1995; Jenkinson and Coleman, 1994), may not be directly comparable to the conditions of the Highfield bare fallow experiment, showing acidic pH values ranging from 5.2 to 6.3 (Supplement Table S2).

4.3 Future developments of the PARTY$_{SOC}$ model

The very first future improvements to the PARTY$_{SOC}$ machine-learning model are to increase the size and further expand the pedoclimatic diversity of its training set. A few additional LTBF sites and several C_3-to-C_4-plant (or C_4 to C_3) long-term vegetation change sites (including space-for-time substitution, like the site of La Cabaña) could be used to achieve this goal. A potential complement lies in a few long-term experimental sites with soil archives and treatments experiencing contrasting SOC stock changes. Radiocarbon measurements on recent and archived soil samples from such sites can be used to infer the content of the centennially stable SOC fraction in topsoils (Hsieh, 1992), but also in subsoils, to allow extending the model to deeper soil horizons. Following the method developed by Buyanovsky and Wagner (1998b, a) and Huggins et al. (1998), the content of the centennially stable SOC fraction can also be estimated in a few additional long-term experiments with contrasting SOC inputs. A promising complement to these strategies comes from numerous long-term sites where time series of SOC inputs, outputs, and stocks are well constrained (i.e. long-term experiments or long-term monitoring sites in various types of ecosystems including arable land, grassland, and forest). It is possible to reliably infer the content of the centennially stable SOC fraction at these sites using simple models of SOC dynamics like AMG (Clivot et al., 2019). Combining all these strategies could help significantly expand the training set of PARTY$_{SOC}$ to soil samples from diverse climates, ecosystems, soil types, and soil depths. When the training setfor PARTY$_{SOC}$ integrates a sufficient diversity of soil samples, a second future improvement of the

model lies in the comparison of different machine-learning algorithms and a test of local modelling approaches, as commonly used in soil spectroscopy studies (Dangal et al., 2019; Gogé et al., 2012; Ramirez-Lopez et al., 2013b, a).

The independent validation of $PARTY_{SOC}v2.0_{EU}$ at five sites with the two dominant reference soil groups in north-western Europe presented here (Figs. 2 and S1) constitutes significant progress in the metrology of SOC kinetic pools. It represents an improvement compared to other approaches that consistently fail to achieve a proper separation of active from stable SOC (Fig. 1; Hsieh, 1992; von Lützow et al., 2007). Methods such as physical and physico-chemical SOC fractionation schemes have been developed to initialize the size of SOC kinetic pools of models (Skjemstad et al., 2004; Zimmermann et al., 2007a), and some of them are now implemented for large topsoil sample sets at the national or continental scale in Europe (Cotrufo et al., 2019; Vos et al., 2018) and Australia (Gray et al., 2019; Viscarra Rossel et al., 2019). A similar implementation in soil monitoring networks of Rock-Eval® measurements combined with the second version of $PARTY_{SOC}$ can provide a more accurate quantification of the functionally different SOC fractions that are centennially stable or active (Fig. 1), at least for a portion of northwestern European agricultural land areas with Cambisols and Luvisols. Large-scale Rock-Eval® measurements and the combined application of $PARTY_{SOC}v2.0_{EU}$ are already ongoing in the French soil monitoring network for soil quality assessment (RMQS; Jolivet et al., 2018). We recommend undertaking similar works in other national and international soil monitoring networks. The second version of $PARTY_{SOC}$ could also be directly employed as an SOC pool partitioning method for simple models of SOC dynamics that are built on the same dualistic conceptual approach of SOC persistence (i.e. active vs. stable SOC pools). The accuracy of these simple models, such as AMG, is highly sensitive to the proper partitioning of SOC kinetic pools (Clivot et al., 2019) and could thus strongly benefit from the second version of $PARTY_{SOC}$.

We envision a significant contribution of the $PARTY_{SOC}$ machine-learning model based on Rock-Eval® thermal analysis to the forthcoming large-scale availability of accurate information on the size of the centennially stable and active SOC fractions. Such accurate information will foster (1) initiatives for soil health assessment and monitoring as well as (2) modelling works of SOC dynamics and of the climate regulation function of soils.

Code and data availability. The Rock-Eval® data for the 105 reference topsoil samples in $PARTY_{SOC}v2.0$ are provided in Table S3 as a CSV file. The R script used to extract Rock-Eval® raw data and calculate Rock-Eval® parameters, the Rock-Eval® data and the R script used to build the $PARTY_{SOC}v2.0$ and $PARTY_{SOC}v2.0_{EU}$ models and test their performance, and the $PARTY_{SOC}v2.0_{EU}$ model (available as an R script and an R data file; please note that predictions of the centennially stable and active SOC contents – in $g\,C\,kg^{-1}$ – are obtained by multiplying their respective proportions by the TOC_{RE6} Rock-Eval® parameter) can be accessed on GitHub at https://github.com/lauric-cecillon/PARTYsoc (last access: 17 June 2021) and on Zenodo at the permanent link https://doi.org/10.5281/zenodo.4446138 (Cécillon, 2021).

Supplement. The supplement related to this article is available on-line at: https://doi.org/10.5194/gmd-14-3879-2021-supplement.

Author contributions. LC and PB designed the study with contributions from CC and FB. FB and FS performed the Rock-Eval® measurements. LC wrote the R scripts used to calculate Rock-Eval® parameters and built the second version of the $PARTY_{SOC}$ model with contributions from PB, LNS, and EK. BTC, UF, SH, TK, IM, FvO, CP, and JCQ provided the topsoil samples and the metadata for the reference sites. LC and PB wrote the paper with contributions from all authors.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We are indebted to the generations of technicians and scientists that started and managed the long-term experiments and archives of soil samples used in this work. We thank Rothamsted Research for access to samples and data from the Rothamsted Sample Archive and the electronic Rothamsted Archive (e-RA). We thank David Montagne and Joël Michelin (AgroParis-Tech, France), who provided information on the soil characteristics at Grignon. We thank our colleagues in the ROMENS research group at Ecole normale supérieure (Paris, France), especially Samuel Abiven, Núria Catalán, Bertrand Guenet, and MarcusSchiedung, who provided advice that improved this paper. Finally, we thank Emanuele Lugato (JRC, Ispra, Italy) and an anonymous reviewer for their constructive comments that further improved ourpaper.

Financial support. This research has been supported by the French Agence nationale de la recherche (StoreSoilC project, grant ANR-17-CE32-0005), the French Agence de la transition écologique (ADEME), and Ville de Paris (SOCUTE project, emergence(s) programme). The Rothamsted Long-term Experiments are supported by the UK Biotechnology and Biological Sciences Research Council under the National Capabilities programme grant (BBS/E/C/000J0300) and by the Lawes Agricultural Trust.

Review statement. This paper was edited by Tomomichi Kato and reviewed by Emanuele Lugato and one anonymous referee.

References

Abiven, S., Menasseri, S., and Chenu, C.: The effects of organic inputs over time on soil aggregate stability – A literature analysis, Soil Biol. Biochem., 41, 1–12, https://doi.org/10.1016/j.soilbio.2008.09.015, 2009.

Amundson, R., Berhe, A. A., Hopmans, J. W., Olson, C., Sztein, A. E., and Sparks, D. L.: Soil and human secu-rity in the 21st century, Science, 348, 1261071–1261071, https://doi.org/10.1126/science.1261071, 2015.

Ansorge, H.: Die Wirkung des Stallmistes im "Statischen Düngungsversuch" Lauchstädt, 2. Mitteilung: Veränderung des Humusgehaltes im Boden, 10, 401–412, 1966.

Avery, B. W. and Catt, J. A.: The soil at Rothamsted, Lawes Agricultural Trust, Harpenden, 1995.

Baldock, J. A., Hawke, B., Sanderman, J., and Macdonald, L. M.: Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra, Soil Res., 51, 577, https://doi.org/10.1071/SR13077, 2013.

Balesdent, J.: The significance of organic separates to car- bon dynamics and its modelling in some cultivated soils, Eur. J. Soil Sci., 47, 485–493, https://doi.org/10.1111/j.1365-2389.1996.tb01848.x, 1996.

Balesdent, J. and Guillet, B.: Les datations par le 14C des matières organiques des sols. Contribution à l'étude de l'humification et du renouvellement des substances humiques, Science du sol, 2, 93–112, 1982.

Balesdent, J. and Mariotti, A.: Measurement of soil organic mat- ter turnover using 13C natural abundance, in: Mass spectrometryof soils, edited by: Boutton, T. W. and Yamasaki, S. I., 83–111, 1996.

Balesdent, J., Mariotti, A., and Guillet, B.: Natural 13C abun- dance as a tracer for studies of soil organic matter dynamics, Soil Biol. Biochem., 19, 25–30, https://doi.org/10.1016/0038-0717(87)90120-9, 1987.

Balesdent, J., Wagner, G. H., and Mariotti, A.: Soil organic matter turnover in long-term field experiments as revealed by carbon-13 natural abundance, Soil Science Society of America Journal, 52, 118–124, https://doi.org/10.2136/sssaj1988.03615995005200010021x, 1988.

Balesdent, J., Basile-Doelsch, I., Chadoeuf, J., Cornu, S., Derrien, D., Fekiacova, Z., and Hatté, C.: Atmosphere–soil car- bon transfer as a function of soil depth, Nature, 559, 599–602, https://doi.org/10.1038/s41586-018-0328-3, 2018.

Barré, P., Eglin, T., Christensen, B. T., Ciais, P., Houot, S., Kätterer, T., van Oort, F., Peylin, P., Poulton, P. R., Romanenkov, V., and Chenu, C.: Quantifying and isolating stable soil organic carbon using long-term bare fallow experiments, Biogeosciences, 7, 3839–3850, https://doi.org/10.5194/bg-7-3839-2010, 2010.

Barré, P., Plante, A. F., Cécillon, L., Lutfalla, S., Baudin, F., Bernard, S., Christensen, B. T., Eglin, T., Fernandez, J. M., Houot, S., Kätterer, T., Le Guillou, C., Macdonald, A., van Oort, F., and Chenu, C.: The energetic and chemical signatures of persistent soil organic matter, Biogeochemistry, 130, 1–12, https://doi.org/10.1007/s10533-016-0246-0, 2016.

Behar, F., Beaumont, V., and De B. Penteado, H. L.: Rock-Eval 6 technology: performances and developments, Oil Gas Sci. Technol., 56, 111–134, https://doi.org/10.2516/ogst:2001013, 2001.

Beleites, C. and Sergo, V.: hyperSpec: a package to handle hyperspectral data sets in R, R package version 0.99-20201127, available at: https://github.com/cbeleites/hyperSpecm (last access: 15 June 2021), 2020.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., and McBratney, A.: Critical review of chemometric indica- tors commonly used for assessing the quality of the prediction ofsoil attributes by NIR spectroscopy, TrAC-Trend Anal. Chem., 29, 1073–1081, https://doi.org/10.1016/j.trac.2010.05.006, 2010.

Borchers, H. W.: racma: Practical Numerical Math Functions. R package version 2.2.9, available at: https://CRAN.R-project.org/package=pracma (last access: 22 June 2021), 2019.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Buyanovsky, G. A. and Wagner, G. H.: Carbon cycling in cultivated land and its global significance, Glob. Change Biol., 4, 131–141, https://doi.org/10.1046/j.1365-2486.1998.00130.x, 1998a.

Buyanovsky, G. A. and Wagner, G. H.: Changing role of cultivated land in the global carbon cycle, Biol. Fert Soils, 27, 242–245, https://doi.org/10.1007/s003740050427, 1998b.

Canty, A. and Ripley, B.: boot: Bootstrap R (S-Plus) Functions, R package version 1.3-28, 2020.

Cardinael, R., Eglin, T., Guenet, B., Neill, C., Houot, S., and Chenu, C.: Is priming effect a significant process for long-term SOC dynamics? Analysis of a 52-years old experiment, Biogeochemistry, 123, 203–219, https://doi.org/10.1007/s10533-014-0063-2, 2015.

Cécillon, L.: lauric-cecillon/PARTYsoc: Second version of the PARTYsoc statistical model (Version v2.0), Zenodo, https://doi.org/10.5281/zenodo.4446138, 2021.

Cécillon, L., Cassagne, N., Czarnes, S., Gros, R., and Brun, J.-J.: Variable selection in near infrared spectra for the biological characterization of soil and earthworm casts, Soil Biol. Biochem., 40, 1975–1979, https://doi.org/10.1016/j.soilbio.2008.03.016, 2008.

Cécillon, L., Baudin, F., Chenu, C., Houot, S., Jolivet, R., Kätterer, T., Lutfalla, S., Macdonald, A., van Oort, F., Plante, A. F., Savignac, F., Soucémarianadin, L. N., and Barré, P.: A model based on Rock-Eval thermal analysis to quantify the size of the centennially persistent organic carbon pool in temperate soils, Biogeosciences, 15, 2835–2849, https://doi.org/10.5194/bg-15-2835-2018, 2018.

Cerri, C., Feller, C., Balesdent, J., Victoria, R., and Plenccassagne, A.: Application du traçage isotopique naturel en 13C, à l'étude de la dynamique de la matière organique dans les sols, Cr. Acad. Sci., 300, 423–428, 1985.

Christensen, B. T. and Johnston, A. E.: Soil organic matter and soil quality – Lessons learned from long-term experi- ments at Askov and Rothamsted, Dev. Soil Sci., 25, 399–430, https://doi.org/10.1016/S0166-2481(97)80045-1, 1997.

Christensen, B. T., Thomsen, I. K., and Eriksen, J.: The Askov long-term experiments: 1894–2019: a unique research platform turns 125 years, DCA – Nationalt Center for Fødevarer og Jordbrug, Tjele, 2019.

Clivot, H., Mouny, J.-C., Duparque, A., Dinh, J.-L., Denoroy, P., Houot, S., Vertès, F., Trochard, R., Bouthier, A., Sagot, S., and Mary, B.: Modeling soil organic carbon evolution in long-term arable experiments with AMG model, Environ. Modell. Softw., 118, 99–113, https://doi.org/10.1016/j.envsoft.2019.04.004, 2019.

Cotrufo, M. F., Ranalli, M. G., Haddix, M. L., Six, J., and Lugato, E.: Soil carbon storage informed by particulate and mineral-associated organic matter, Nat. Geosci., 12, 989–994, https://doi.org/10.1038/s41561-019-0484-6, 2019.

Coulston, J. W., Blinn, C. E., Thomas, V. A., and Wynne, R. H.: Approximating prediction uncertainty for random forest regression models, Photogramm. Eng. Rem. S., 82, 189–197, https://doi.org/10.14358/PERS.82.3.189, 2016.

Dangal, S., Sanderman, J., Wills, S., and Ramirez-Lopez, L.: Accurate and precise prediction of soil properties from a large mid-infrared spectral library, Soil Syst., 3, 11, https://doi.org/10.3390/soilsystems3010011, 2019.

Davison, A. C. and Hinkley, D. V.: Bootstrap methods and their application, Cambridge University Press, Cambridge, New York, NY, USA, 582 pp., 1997.

Disnar, J. R., Guillet, B., Keravis, D., Di-Giovanni, C., and Sebag, D.: Soil organic matter (SOM) characterization by Rock-Eval pyrolysis: scope and limitations, Org. Geochem., 34, 327–343, https://doi.org/10.1016/S0146-6380(02)00239-5, 2003.

European Commission: Soils of the European Union, Joint Research Centre, Institute for Environment and Sustainability, Publications Office, LU, 2008.

Falloon, P., Smith, P., Coleman, K., and Marshall, S.: Estimating the size of the inert organic matter pool from total soil organic carbon content for use in the Rothamsted carbon model, Soil Biol. Biochem., 30, 1207–1211, https://doi.org/10.1016/S0038-0717(97)00256-3, 1998.

Falloon, P. D. and Smith, P.: Modelling refractory soil organic matter, Biol. Fert. Soils, 30, 388–398, https://doi.org/10.1007/s003740050019, 2000.

FAO: World reference base for soil resources 2014: international soil classification system for naming soils and creating legends for soil maps, FAO, Rome, 2014.

Franko, U. and Merbach, I.: Modelling soil organic matter dynamics on a bare fallow Chernozem soil in Central Germany, Geoderma, 303, 93–98, https://doi.org/10.1016/j.geoderma.2017.05.013, 2017.

Genuer, R. and Poggi, J.-M.: Random Forests with R, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-56485-8, 2020.

Gogé, F., Joffre, R., Jolivet, C., Ross, I., and Ranjard, L.: Optimization criteria in sample selection step of lo-cal regression for quantitative analysis of large soilNIRS database, Chemometr. Intell. Lab., 110, 168–176, https://doi.org/10.1016/j.chemolab.2011.11.003, 2012.

Gray, J., Karunaratne, S., Bishop, T., Wilson, B., and Veeragathipillai, M.: Driving factors of soil organic carbon fractions over New South Wales, Australia, Geoderma, 353, 213–226, https://doi.org/10.1016/j.geoderma.2019.06.032, 2019.

Gregorich, E. G., Gillespie, A. W., Beare, M. H., Curtin,D., Sanei, H., and Yanni, S. F.: Evaluating biodegrad-ability of soil organic matter by its thermal stability and chemical composition, Soil Biol. Biochem., 91, 182–191, https://doi.org/10.1016/j.soilbio.2015.08.032, 2015.

He, Y., Trumbore, S. E., Torn, M. S., Harden, J. W., Vaughn, L. J. S., Allison, S. D., and Randerson, J. T.: Radiocarbon constraints imply reduced carbon uptake by soils during the 21st century, Science, 353, 1419–1424, https://doi.org/10.1126/science.aad4273, 2016.

Hénin, S. and Dupuis, M.: Bilan de la matière organique des sols, Annales Agronomiques, 1, 17–29, 1945.

Hénin, S. and Turc, L.: Essai de fractionnement des matières organiques du sol, Comptes rendus de l'Académie d'agriculture de Francem 35, 41–43, 1949.

Houot, S., Molina, J. A. E., Chaussod, R., and Clapp, C. E.: Simulation by NCSOIL of net mineralization in soils from the Deherain and 36 parcelles fields at Grignon, Soil Sci. Soc. Am. J., 53, 451–455, https://doi.org/10.2136/sssaj1989.03615995005300020023x, 1989.

Hsieh, Y.-P.: Pool size and mean age of stable soil organic carbon in croplands, Soil Sci. Soc. Am. J., 56, 460–464, https://doi.org/10.2136/sssaj1992.03615995005600020049x, 1992.

Huggins, D. R., Buyanovsky, G. A., Wagner, G. H., Brown, J. R., Darmody, R. G., Peck, T. R., Lesoing, G. W., Vanotti, M. B., and Bundy, L. G.: Soil organic C in the tallgrass prairie-derived region of the corn belt: effects of long-term crop management, Soil Till. Res., 47, 219–234, https://doi.org/10.1016/S0167-1987(98)00108-1, 1998.

IPBES: Summary for policymakers of the assessment report on land degradation and restoration of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, edited by: Scholes, R. J., Montanarella, L., Brainich, E., Brainich, E., Barger, N., ten Brink, B., Cantele, M., Erasmus, B., Fisher, J., Gardner, T., Holland, T. G., Kohler, F., Kotiaho, S., von Maltitz, G., Nangendo, G., Pandit, R., Parrotta, J., Potts, M. D., Prince, S., Sankaran, M., and Willemen, L., Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, 2018.

IPCC: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems, edited by: Shukla, P. R., Skea, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D. C., Zhai, P., Slade, R., Connors, S., van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Portugal Pereira, J., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M., and Malley, J., available at: https://www.ipcc.ch/srccl/ (last access: 22 June 2021), 2019.

ISO 10694: Soil quality – Determination of organic and total carbon after dry combustion (elementary analysis), available at: https://www.iso.org/standard/18782.html (last access: 22 June 2021), 1995.

Jaconi, A., Poeplau, C., Ramirez-Lopez, L., Van Wesemael, B., and Don, A.: Log-ratio transformation is the key to determining soil organic carbon fractions with near-infrared spectroscopy, Eur. J. Soil. Sci., 70, 127–139, https://doi.org/10.1111/ejss.12761, 2019.

Janzen, H. H.: The soil carbon dilemma: shall we hoardit or use it?, Soil Biol. Biochem., 38, 419–424, https://doi.org/10.1016/j.soilbio.2005.10.008, 2006.

Jenkinson, D. S.: The turnover of organic carbon and nitro- gen in soil, Philos. T. R. Soc. Lond. B, 329, 361–368, https://doi.org/10.1098/rstb.1990.0177, 1990.

Jenkinson, D. S. and Coleman, K.: Calculating the annual in-put of organic matter to soil from measurements of total or- ganic carbon and radiocarbon, Eur. J. Soil Sci., 45, 167–174, https://doi.org/10.1111/j.1365-2389.1994.tb00498.x, 1994.

Jenkinson, D. S., Adams, D. E., and Wild, A.: Model estimates of CO_2 emissions from soil in response to global warming, Nature, 351, 304–306, https://doi.org/10.1038/351304a0, 1991.

Johnston, A. E., Poulton, P. R., and Coleman, K.: Soil or- ganic matter: its importance in sustainable agriculture and carbon dioxide fluxes, in: Adv. Agronom., 101, 1–57, https://doi.org/10.1016/S0065-2113(08)00801-8, 2009.

Jolivet, C., Almeida-Falcon, J. L., Berché, P., Boulonne, L., Fontaine, M., Gouny, L., Lehmann, S., Maître, B., Ratié, C., Schellenberger, E., and Soler-Dominguez, N.: Manuel du Réseau de mesures de la qualité des sols, RMQS2: deuxième campagne métropolitaine, 2016 – 2027, Version 3, INRA, US 1106 InfoSol, Orléans, France, 2018.

Kätterer, T., Bolinder, M. A., Andrén, O., Kirchmann, H., and Menichetti, L.: Roots contribute more to refractory soil organic matter than above-ground crop residues, as revealed by a long-term field experiment, Agriculture, Ecosyst. Environ., 141, 184–192, https://doi.org/10.1016/j.agee.2011.02.029, 2011.

Keesstra, S. D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., Cerdà, A., Montanarella, L., Quinton, J. N., Pachepsky,Y., van der Putten, W. H., Bardgett, R. D., Moolenaar, S.,Mol, G., Jansen, B., and Fresco, L. O.: The significanceof soils and soil science towards realization of the United Nations Sustainable Development Goals, SOIL, 2, 111–128, https://doi.org/10.5194/soil-2-111-2016, 2016.

Khedim, N., Cécillon, L., Poulenard, J., Barré, P., Baudin, F., Marta, S., Rabatel, A., Dentant, C., Cauvy-Fraunié, S., Anthelme, F., Gielly, L., Ambrosini, R., Franzetti, A., Azzoni, R. S., Caccianiga, M. S., Compostella, C., Clague, J., Tielidze, L., Messager, E., Choler, P., and Ficetola, G. F.: Topsoil organic matter build-up in glacier forelands around the world, Glob. Change Biol., 27, 1662–1677, https://doi.org/10.1111/gcb.15496, 2021.

Koch, A., McBratney, A., Adams, M., Field, D., Hill, R., Crawford, J., Minasny, B., Lal, R., Abbott, L., O'Donnell, A., Angers, D., Baldock, J., Barbier, E., Binkley, D., Parton, W., Wall, D. H., Bird, M., Bouma, J., Chenu, C., Flora, C. B., Goulding, K., Grunwald, S., Hempel, J., Jastrow, J., Lehmann, J., Lorenz, K., Morgan, C. L., Rice, C. W., Whitehead, D., Young, I., and Zimmermann, M.: Soil security: solving the global soil crisis, Glob. Policy, 4, 434–441, https://doi.org/10.1111/1758-5899.12096, 2013.

Körschens, M., Weigel, A., and Schulz, E.: Turnover of soil organic matter (SOM) and long-term balances – tools for evaluating sustainable productivity of soils, Z. Pflanzenernaehr. Bodenk., 161, 409–424, https://doi.org/10.1002/jpln.1998.3581610409, 1998.

Lal, R.: Soil carbon sequestration impacts on global climate change and food security, Science, 304, 1623–1627, https://doi.org/10.1126/science.1097396, 2004.

Lavallee, J. M., Soong, J. L., and Cotrufo, M. F.: Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century, Glob. Change Biol., 26, 261–273, https://doi.org/10.1111/gcb.14859, 2020.

Liaw, A. and Wiener, M.: Classification and regression by randomForest, R News, 2, 18–22, 2002.

Ludwig, B., Schulz, E., Rethemeyer, J., Merbach, I., andFlessa, H.: Predictive modelling of C dynamics in the long-term fertilization experiment at Bad Lauchstädt with the Rothamsted Carbon Model, Eur. J. Soil Sci., 58, 1155–1163, https://doi.org/10.1111/j.1365-2389.2007.00907.x, 2007.

Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J. W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R. B., Jones, C. D., Lara, M. J., Liang, J., McGuire, A. D., Parton, W., Peng, C., Randerson, J. T., Salazar, A., Sierra, C. A., Smith, M. J., Tian, H., Todd-Brown, K. E. O., Torn, M., van Groenigen, K. J., Wang, Y. P., West, T. O., Wei, Y., Wieder, W. R., Xia, J., Xu, X., Xu, X., and Zhou, T.: Toward more realistic projections of soil carbon dynamics by Earth system models, Global Biogeochem. Cy., 30, 40–56, https://doi.org/10.1002/2015GB005239, 2016.

Monnier, G., Turc, C., and Jeanson Luusinang, C.: Une methode de fractionnement densimétrique par centrifugation des matières organiques du sol, Annales Agronomiques, 13, 55–63, 1962.

Nikiforoff, C. C.: Some General Aspects of the Chernozem Formation, Soil Sci. Soc. Am. J., 1, 333–342, https://doi.org/10.2136/sssaj1937.03615995000100000060x, 1936.

Patil, A., Huard, D., and Fonnesbeck, C.: PyMC: Bayesian stochastic modelling in Python, J. Stat. Softw., 35, https://doi.org/10.18637/jss.v035.i04, 2010.

Pellerin, S., Bamière, L., Launay, C., Martin, R., Schiavo, M., Angers, D., Augusto, L., Balesdent, J., Basile-Doelsch, I., Bellassen, V., Cardinael, R., Cécillon, L., Ceschia, E., Chenu, C., Constantin, J., Darroussin, J., Delacote, P., Delame, N., Gastal, F., Gilbert, D., Graux, A.-I., Guenet, B., Houot, S., Klumpp, K., Letort, E., Litrico, I., Martin, M., Menasseri-Aubry, S., Meziere, D., Morvan, T., Mosnier, C., Roger-Estrade, J., Saint-André, L., Sierra, J., Therond, O., Viaud, V., Grateau, R., Le Perchec, S., Savini, I., and Rechauchère, O.: Stocker du carbone dans les sols français, Quel potentiel au regard de l'objectif 4 pour 1000 et à quel coût? Rapport scientifique de l'étude, INRA (France), 540 pp., available at: https://www.inrae.fr/sites/default/files/pdf/RapportEtude4p1000.pdf (last access: 22 June 2021), 2020.

Petersen, B. M., Berntsen, J., Hansen, S., and Jensen, L. S.: CN-SIM – a model for the turnover of soil organic matter. I. Long-term carbon and radiocarbon development, Soil Biol. Biochem., 37, 359–374, https://doi.org/10.1016/j.soilbio.2004.08.006, 2005.

Plante, A. F., Beaupré, S. R., Roberts, M. L., and Baisden, T.: Distribution of radiocarbon ages in soil organic mat-ter by thermal fractionation, Radiocarbon, 55, 1077–1083, https://doi.org/10.1017/S0033822200058215, 2013.

Poeplau, C., Don, A., Dondini, M., Leifeld, J., Nemo, R., Schumacher, J., Senapati, N., and Wiesmeier, M.: Reproducibility of a soil organic carbon fractionation method to derive RothC carbon pools: Soil carbon fractionation ring trial, Eur. J. Soil Sci., 64, 735–746, https://doi.org/10.1111/ejss.12088, 2013.

Poeplau, C., Don, A., Six, J., Kaiser, M., Benbi, D., Chenu, C., Cotrufo, M. F., Derrien, D., Gioacchini, P., Grand, S., Gregorich, E., Griepentrog, M., Gunina, A., Haddix, M., Kuzyakov, Y., Küh- nel, A., Macdonald, L. M., Soong, J., Trigalet, S., Vermeire, M.-L., Rovira, P., van Wesemael, B., Wiesmeier, M., Yeasmin, S.,

Yevdokimov, I., and Nieder, R.: Isolating organic carbon fractions with varying turnover rates in temperate agricultural soils – A comprehensive method comparison, Soil Biol. Biochem., 125, 10–26, https://doi.org/10.1016/j.soilbio.2018.06.025, 2018.

Poeplau, C., Barré, P., Cécillon, L., Baudin, F., and Sig- urdsson, B. D.: Changes in the Rock-Eval signature ofsoil organic carbon upon extreme soil warming and chem- ical oxidation - A comparison, Geoderma, 337, 181–190, https://doi.org/10.1016/j.geoderma.2018.09.025, 2019.

Quezada, J. C., Etter, A., Ghazoul, J., Buttler, A., and Guillaume, T.: Carbon neutral expansion of oil palm plantations in the Neotropics, Sci. Adv., 5, eaaw4418, https://doi.org/10.1126/sciadv.aaw4418, 2019.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Rossel, R. A. V., Demattê, J. A. M., and Scholten, T.: Distance and similarity-search metrics for use with soil vis–NIR spectra, Geoderma, 199, 43–53, https://doi.org/10.1016/j.geoderma.2012.08.035, 2013a.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J. A. M., and Scholten, T.: The spectrum-basedlearner: A new local approach for modeling soil vis–NIR spectra of complex datasets, Geoderma, 195–196, 268–279, https://doi.org/10.1016/j.geoderma.2012.12.014, 2013b.

R Core Team: R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2020.

RStudio Team: RStudio: integrated development for R, RStudio, Inc., Boston, MA, 2020.

Rühlmann, J.: A new approach to estimating the poolof stable organic matter in soil using data from long-term field experiments, Plant Soil, 213, 149–160, https://doi.org/10.1023/A:1004552016182, 1999.

Saenger, A., Cécillon, L., Sebag, D., and Brun, J.-J.: Soil organic carbon quantity, chemistry and thermal stability in a mountainous landscape: A Rock–Eval pyrolysis survey, Org. Geochem., 54, 101–114, https://doi.org/10.1016/j.orggeochem.2012.10.008, 2013.

Saenger, A., Cécillon, L., Poulenard, J., Bureau, F., De Daniéli, S., Gonzalez, J.-M., and Brun, J.-J.: Surveying the carbon pools of mountain soils: A comparison of physical fractiona- tion and Rock-Eval pyrolysis, Geoderma, 241–242, 279–288, https://doi.org/10.1016/j.geoderma.2014.12.001, 2015.

Sanderman, J. and Grandy, A. S.: Ramped thermal analy-sis for isolating biologically meaningful soil organic matterfractions with distinct residence times, SOIL, 6, 131–144, https://doi.org/10.5194/soil-6-131-2020, 2020.

Sanderman, J., Hengl, T., and Fiske, G. J.: Soil carbon debt of 12,000 years of human land use, P. Natl. Acad. Sci. USA, 114, 9575–9580, https://doi.org/10.1073/pnas.1706103114, 2017.

Schiedung, M., Don, A., Wordell-Dietrich, P., Alcántara, V., Kuner, P., and Guggenberger, G.: Thermal oxidation does not fractionate soil organic carbon with differing bio- logical stabilities, J. Plant Nutr. Soil Sci., 180, 18–26, https://doi.org/10.1002/jpln.201600172, 2017.

Schulte, R. P. O., Creamer, R. E., Donnellan, T., Far- relly, N., Fealy, R., O'Donoghue, C., and O'hUallachain, D.: Functional land management: A framework for man- aging soil-based ecosystem services for the sustainable in- tensification of agriculture, Environ. Sci. Policy, 38, 45–58, https://doi.org/10.1016/j.envsci.2013.10.002, 2014.

Sebag, D., Verrecchia, E. P., Cécillon, L., Adatte, T., Al- brecht, R., Aubert, M., Bureau, F., Cailleau, G., Co- pard, Y., Decaens, T., Disnar, J.-R., Hetényi, M., Nyilas, T., and Trombino, L.: Dynamics of soil organic matter based on new Rock-Eval indices, Geoderma, 284, 185–203, https://doi.org/10.1016/j.geoderma.2016.08.025, 2016.

Shi, Z., Allison, S. D., He, Y., Levine, P. A., Hoyt, A. M., Beem-Miller, J., Zhu, Q., Wieder, W. R., Trumbore, S., and Randerson, J. T.: The age distribution of global soil carbon inferred from radiocarbon measurements, Nat. Geosci., 13, 555–559, https://doi.org/10.1038/s41561-020-0596-z, 2020.

Skjemstad, J. O., Spouncer, L. R., Cowie, B., and Swift, R. S.: Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools, Soil Res., 42, 79–88, https://doi.org/10.1071/SR03013, 2004.

Soucémarianadin, L., Cécillon, L., Chenu, C., Baudin, F., Nicolas, M., Girardin, C., and Barré, P.: Is Rock-Eval 6 thermal analysis a good indicator of soil organic carbon lability? – A method-comparison study in forest soils, Soil Biol. Biochem., 117, 108–116, https://doi.org/10.1016/j.soilbio.2017.10.025, 2018a.

Soucémarianadin, L. N., Cécillon, L., Guenet, B., Chenu, C., Baudin, F., Nicolas, M., Girardin, C., and Barré, P.: Environmental factors controlling soil organic carbon sta- bility in French forest soils, Plant Soil, 426, 267–286, https://doi.org/10.1007/s11104-018-3613-x, 2018b.

Stoorvogel, J. J., Bakkenes, M., Brink, B. J. E., and Temme, A. J. A. M.: To what extent did we change our soils? A global comparison of natural and current conditions, Land Degrad. Develop., 28, 1982–1991, https://doi.org/10.1002/ldr.2721, 2017.

Strobl, C., Malley, J., and Tutz, G.: An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests, Psychol. Meth., 14, 323–348, https://doi.org/10.1037/a0016973, 2009.

Taghizadeh-Toosi, A., Cong, W.-F., Eriksen, J., Mayer, J., Ole- sen, J. E., Keel, S. G., Glendining, M., Kätterer, T., and Chris- tensen, B. T.: Visiting dark sides of model simulation of car- bon stocks in European temperate agricultural soils: allomet-ric function and model initialization, Plant Soil, 450, 255–272, https://doi.org/10.1007/s11104-020-04500-9, 2020.

Trumbore, S. E., Vogel, J. S., and Southon, J. R.: AMS 14C measurements of fractionated soil organic matter: an approach to deciphering the soil carbon cycle, Radiocarbon, 31, 644–654, https://doi.org/10.1017/S0033822200012248, 1989.

van Oort, F., Paradelo, R., Proix, N., Delarue, G., Baize,D., and Monna, F.: Centennial fertilization-induced soil processes control trace metal dynamics. Lessons froma long-term bare fallow experiment, Soil Syst., 2, 23, https://doi.org/10.3390/soilsystems2020023, 2018.

Viscarra Rossel, R. A. and Hicks, W. S.: Soil organic carbon and its fractions estimated by visible-near infrared transfer functions: Vis-NIR estimates of organic carbon and its fractions, Eur. J. Soil Sci., 66, 438–450, https://doi.org/10.1111/ejss.12237, 2015.

Viscarra Rossel, R. A., Lee, J., Behrens, T., Luo, Z., Baldock, J., and Richards, A.: Continental-scale soil carbon composition and vulnerability modulated by regional environmental controls, Nat. Geosci., 12, 547–552, https://doi.org/10.1038/s41561-019-0373-z, 2019.

von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E., and Marschner, B.: SOM frac-

tionation methods: Relevance to functional pools and to stabilization mechanisms, Soil Biol. Biochem., 39, 2183–2207, https://doi.org/10.1016/j.soilbio.2007.03.007, 2007.

Vos, C., Jaconi, A., Jacobs, A., and Don, A.: Hot regionsof labile and stable soil organic carbon in Germany – Spatial variability and driving factors, SOIL, 4, 153–167, https://doi.org/10.5194/soil-4-153-2018, 2018.

Wehrens, R.: Chemometrics with R: Multivariate Data Analysis in the Natural and Life Sciences, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-62027-4, 2020.

Wickham, H.: tringr: Simple, consistent wrappers for common string operations, R package version 1.4.0, available at: https://CRAN.R-project.org/package=stringr (last access: 22 June 2021), 2019.

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., and Kögel-Knabner, I.: Soil organic carbon storage as a key function of soils – A review of drivers and indicators at various scales, Geoderma, 333, 149–162, https://doi.org/10.1016/j.geoderma.2018.07.026, 2019.

Zimmermann, M., Leifeld, J., Schmidt, M. W. I., Smith, P., and Fuhrer, J.: Measured soil organic matter fractions can be related to pools in the RothC model, Eur. J. Soil Sci., 58, 658–667, https://doi.org/10.1111/j.1365-2389.2006.00855.x, 2007a.

Zimmermann, M., Leifeld, J., and Fuhrer, J.: Quan- tifying soil organic carbon fractions by infrared- spectroscopy, Soil Biol. Biochem., 39, 224–231, https://doi.org/10.1016/j.soilbio.2006.07.010, 2007b.

ANNEX 2

A quick analysis of AMG residuals for three initialization approaches

Objective:

Here we examined if the residuals of AMG predictions (i.e., the difference between the AMG model predictions and the observed SOC change) for the nine LTE sites introduced in Chapter 1 of this manuscript were related to any of the pedoclimatic parameters included in the model. Moreover, we compared the behaviour of the residuals for three different initialization methods; using optimized, default, and $PARTY_{SOC}$-based pool partitioning to obtain SOC simulations.

How-to:

Residuals represent the portion of the variability in SOC that is not explained by the model. Here we studied the relationship between the residuals and ten variables including time (yr^{-1}), temperature (°C), potential evapotranspiration (P_ETP; mm yr^{-1}), initial SOC content ($SOC_{initial}$; g C kg $soil^{-1}$), clay content (g kg $soil^{-1}$), $CaCO_3$ content (g kg $soil^{-1}$), soil pH, bulk density (g cm^{-3}), C/N ratio, and initial SOC stock ($QC_{initial}$; tC ha^{-1}).

We calculated the residuals for a total of 32 trials from the nine LTEs, as the difference between **simulated** and **observed** total carbon stock change:

$$\boldsymbol{AMG\ dCS\ residuals\ =\ dCS_{sim}\ -\ dCS_{obs}}$$

$$dCS_{obs} = CS_{obs,t_2} - CS_{obs,t_1}$$

$$dCS_{sim} = CS_{sim,t_2} - CS_{obs,t_1}$$

where CS_{obs}=observed carbon stock at time t, CS_{sim}=carbon stock at time t simulated with AMG, t_1=start of simulation period, t_2=end of simulation period

Plotting the residuals against the different parameters for each trial allowed us to find the best model describing the observed relationship; here linear regression. Only in cases where the correlation was strong and the linear regression parameters significant, could a part of the variability not accounted for by the model be explained by the corresponding variable.

Results:

As expected due to the exceptional performance of AMG when correctly initialized, we found very low values for the residuals of the AMG model initialized by optimized pool partitioning (Fig. 1). These were very weakly correlated to two out of ten parameters examined here, clay content (R^2=0.12) and soil pH (R^2=0.15; Fig.1; Table 1). For the default AMG initialization as we show in Chapter 1 the accuracy of the model was less good (i.e., residuals become more important; Fig. 2). The source of this uncertainty seems to be related to the consideration of initial SOC content (or initial SOC stocks) since there was a strong correlation between AMG residuals and both initial SOC content (R^2=0.53) and SOC stocks (R^2=0.42; Fig. 2; Table 2). As for the $PARTY_{SOC}$-based initialization, it also increased the residuals of AMG predictions compared to the AMG initialization using optimized pool partitioning (Fig. 3). However, the correlations of the residuals to various parameters, namely time (R^2=0.22), potential evapotranspiration (R^2=0.37), clay content (R^2=0.36), soil pH (R^2=0.25), and bulk density (R^2=0.16), was weaker compared to the default AMG initialization (Fig. 3; Table 3).

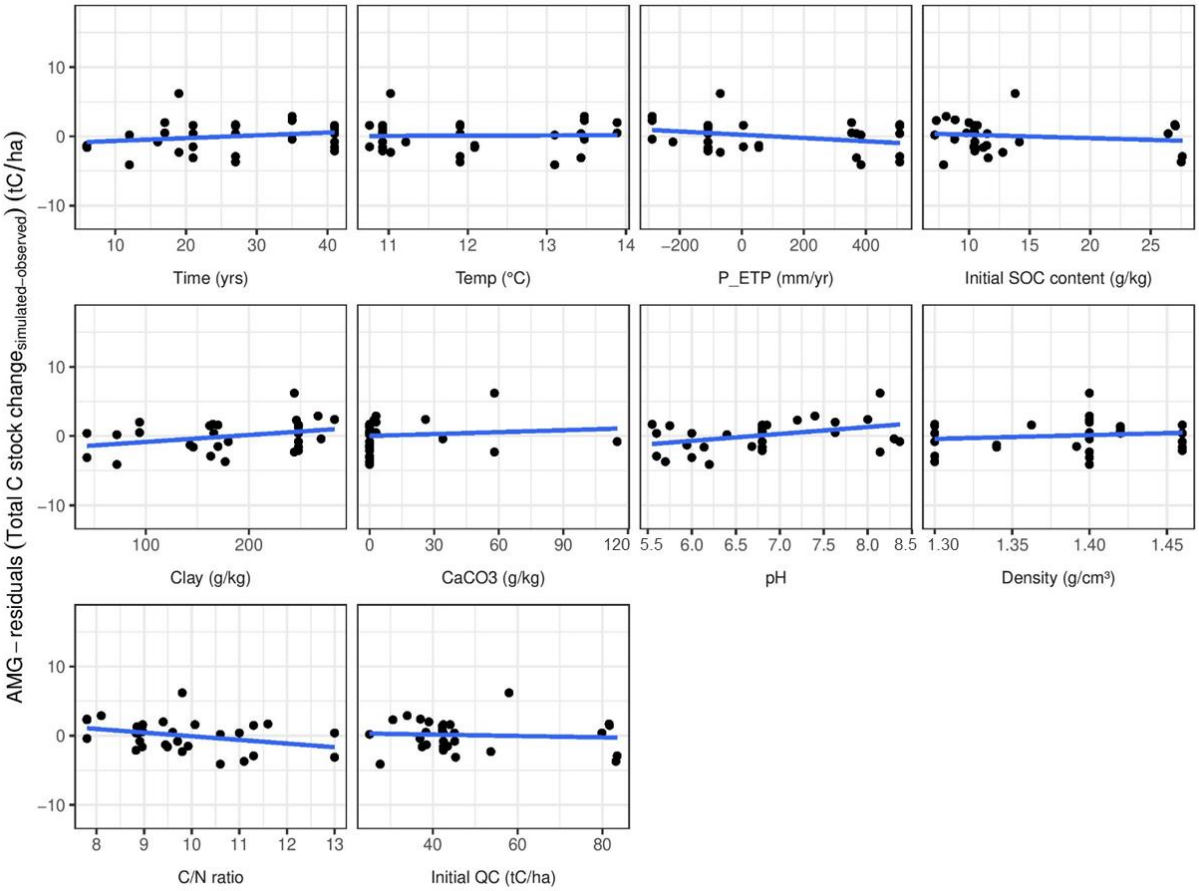


Figure 1: Residual analysis of the AMG model initialized by optimized pool partitioning.

Table 1: Linear regression parameters of the relationships between variables and AMG residuals after optimized AMG pool partitioning. Significant parameters are shown in bold.

Variable	R^2	Slope	p.value	Intercept	p.value
Time	0.06	0.05	1.47e-01	-1.23	2.07e-01
Temp	0.00	0.03	9.27e-01	-0.27	9.45e-01
P_ETP	0.09	0	7.68e-02	0.23	5.20e-01
$SOC_{initial}$	0.03	0.06	3.58e-01	-0.62	4.69e-01
Clay	**0.12**	**0.01**	**4.86e-02**	-1.86	7.54e-02
$CaCO_3$	0.01	0.01	5.53e-01	0.01	9.81e-01
pH	**0.15**	**1.00**	**2.59e-02**	**-6.69**	**2.88e-02**
Density	0.02	5.65	4.3e-01	-7.78	4.35e-01
C/N ratio	0.11	-0.53	5.77e-02	5.17	5.56e-02
$QC_{initial}$	0.01	0.02	5.07e-01	-0.64	5.82e-01

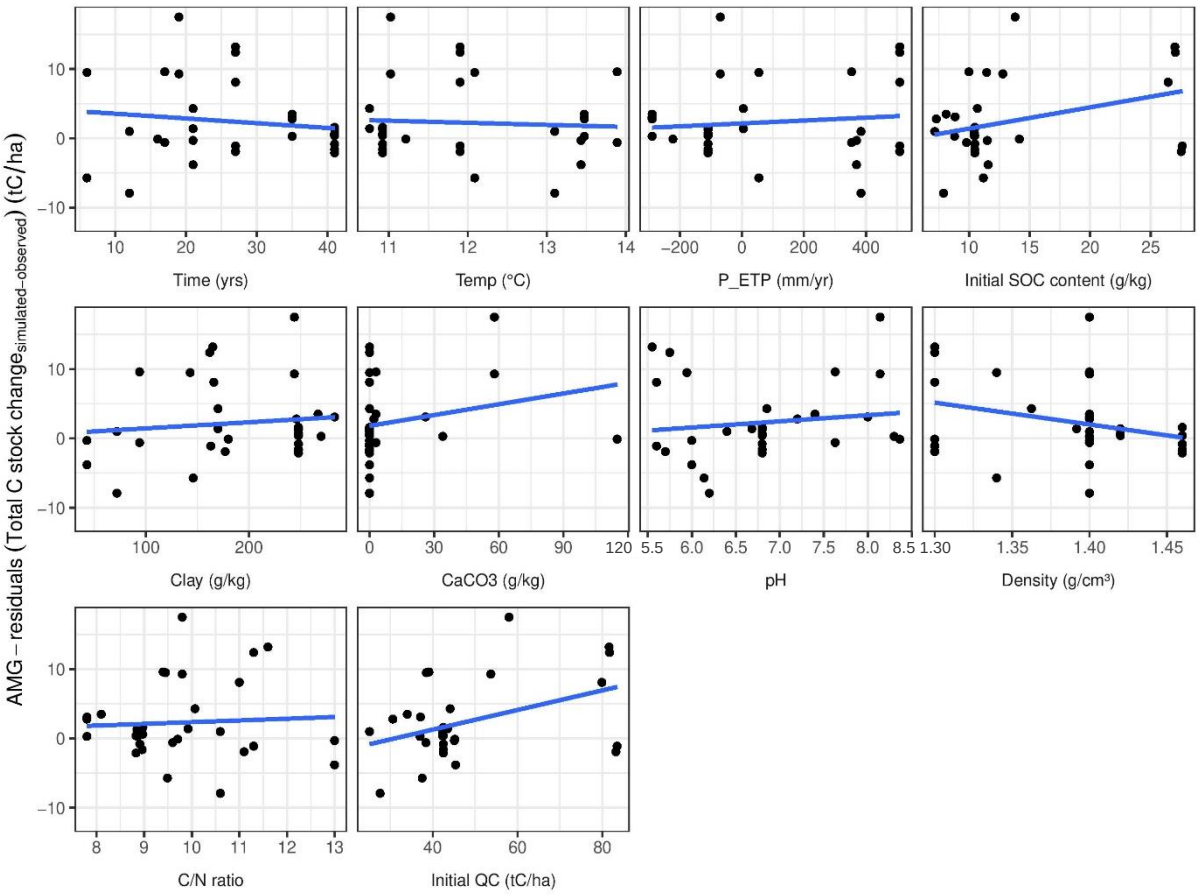


Figure 2: Residual analysis of the AMG model initialized by default pool partitioning.

Table 2: Linear regression parameters of the relationships between variables and AMG residuals after default AMG pool partitioning. Significant parameters are shown in bold.

Variable	R^2	Slope	p.value	Intercept	p.value
Time	0.00	-0.02	8.21e-01	2.81	2.79e-01
Temp	0.00	-0.31	7.2e-01	5.92	5.63e-01
P_ETP	0.01	0.00	5.67e-01	**2.15**	**3.26e-02**
$SOC_{initial}$	**0.53**	**0.64**	**9.73e-07**	**-5.93**	**4.15e-04**
Clay	0.01	0.01	5.16e-01	0.57	8.39e-01
$CaCO_3$	0.05	0.05	1.92e-01	1.81	7.44e-02
pH	0.02	0.88	4.6e-01	-3.73	6.47e-01
Density	0.09	-31.46	8.26e-02	46.07	6.87e-02
C/N ratio	0.00	0.25	7.34e-01	-0.14	9.85e-01
$QC_{initial}$	**0.42**	**0.22**	**2.98e-05**	**-8.18**	**1.04e-03**

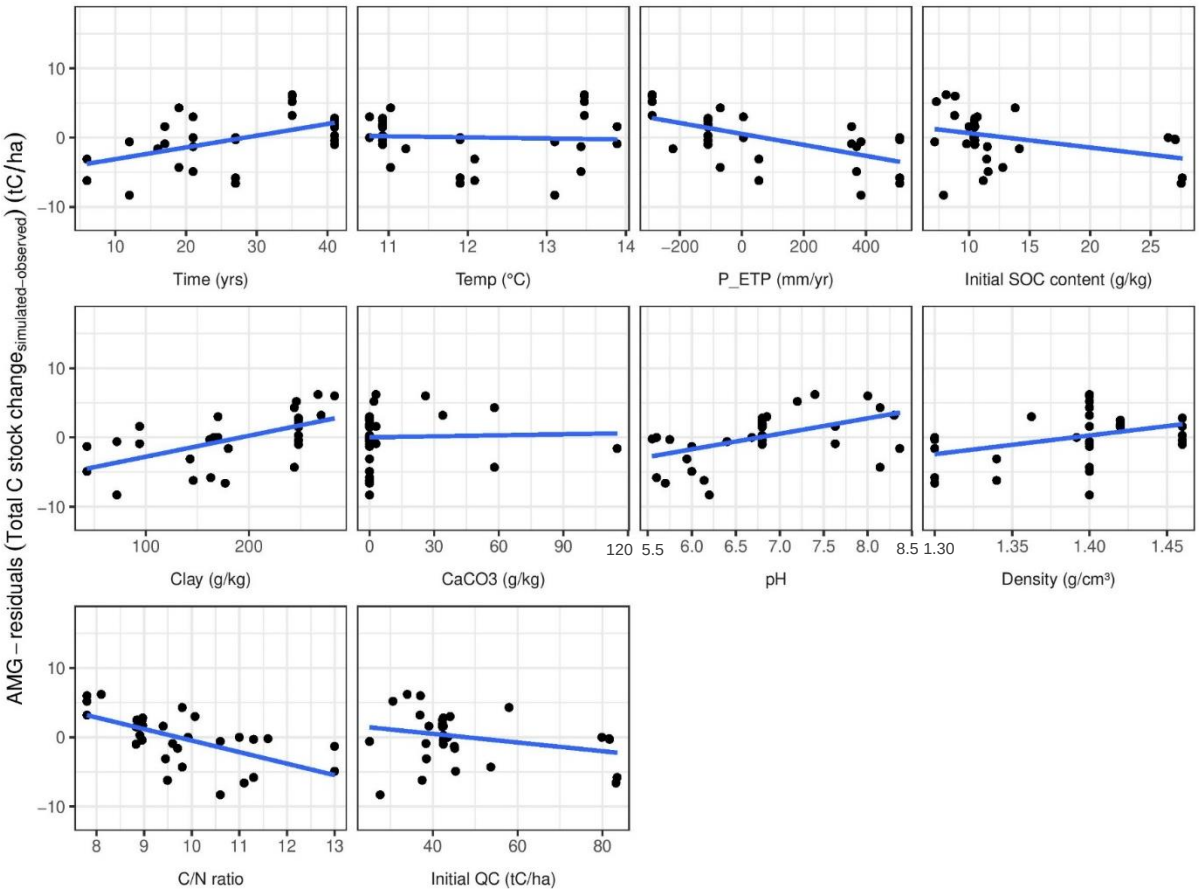


Figure 3: Residual analysis of the AMG model initialized by $PARTY_{SOC}$-based pool partitioning.

Table 3: Linear regression parameters of the relationships between variables and AMG residuals after AMG pool partitioning according to $PARTY_{SOC}$. Significant parameters are shown in bold.

Variable	R^2	Slope	p.value	Intercept	p.value
Time	**0.22**	**0.14**	**5.25e-03**	**-4.08**	**9.83e-03**
Temp	0.00	-0.16	7.82e-01	1.91	7.76e-01
P_ETP	**0.37**	**-0.01**	**1.28e-04**	0.53	3.01e-01
$SOC_{initial}$	0.03	-0.11	3.02e-01	1.41	3.31e-01
Clay	**0.36**	**0.03**	**1.64e-04**	**-5.80**	**3.76e-04**
$CaCO_3$	0.00	0.00	8.51e-01	0.01	9.86e-01
pH	**0.25**	**2.23**	**2.44e-03**	**-15.1**	**2.67e-03**
Density	**0.16**	**27.54**	**1.85e-02**	**-38.29**	**1.87e-02**
C/N ratio	**0.37**	**-1.66**	**1.24e-04**	**16.09**	**1.32e-04**
$QC_{initial}$	0.05	-0.05	1.89e-01	2.49	2.02e-01

Discussion and conclusions

First, weak and marginally significant correlations between examined variables and residuals of AMG initialized using ex post optimized pool partitioning show that this version of the model is able to account for practically the entire variability in SOC. Even though the correlations between clay content and soil pH could be an indication of the importance of these two parameter as controls of SOC, their weak value shows that only very little space for improvement is left regarding the parametrization of the model. Better calibration of the model functions related to pH and clay content could eventually slightly improve the performance of the model, at least for the sites tested here. Second, for the initialization of AMG using default pool partitioning strong and significant correlations between residuals and initial SOC content and SOC stocks were observed. Imprecise information regarding initial SOC pool partitioning resulted in higher residuals. By definition, higher residuals can also be expressed as an underestimation of the observed change in SOC stocks. We interpret the weaker correlations observed between residuals and various parameters as a weaker and less biased uncertainty caused by the $PARTY_{SOC}$-based initialization compared to the initialization by default.

Overall the AMG model performs well, although the initialization method can influence the model performance and the behaviour of residuals. In agreement with Clivot et al. (2019) we underline once more the importance of correct pool initialization for the precision of simulations.

DISCLAIMER: This part was conducted as a conceptual exercise and concerns the model performance only for the sites used here.