Understanding and multi-scale modelling of clay shrinkage-swelling

Sophie Barthélemy

▶ **To cite this version:**

Sophie Barthélemy. Understanding and multi-scale modelling of clay shrinkage-swelling. Earth Sciences. Université de Toulouse, 2024. English. NNT : 2024TLSEP091 . tel-04845472

HAL Id: tel-04845472
https://theses.hal.science/tel-04845472v1

Submitted on 18 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

Université de Toulouse

Doctorat de l'Université de Toulouse

préparé à Toulouse INP

Compréhension et modélisation multi-échelle du phénomène de retrait-gonflement des argiles

---

Thèse présentée et soutenue, le 13 novembre 2024 par

Sophie BARTHELEMY

École doctorale
SDU2E - Sciences de l'Univers, de l'Environnement et de l'Espace

Spécialité
Océan, Atmosphère, Climat

Unité de recherche
CNRM - Centre National de Recherches Météorologiques

Thèse dirigée par
Jean-Christophe CALVET et GILLES GRANDJEAN

Composition du jury
M. Lionel JARLAN, Président, IRD
Mme Catherine OTTLE, Rapporteuse, CNRS Île-de-France Gif-sur-Yvette
Mme Florence HABETS, Rapporteuse, CNRS Paris-Centre
M. Guillaume CHAMBON, Examinateur, INRAE Lyon-Grenoble Auvergne-Rhône-Alpes
M. Olivier CUISINIER, Examinateur, Université de Lorraine
M. Jean-Christophe CALVET, Directeur de thèse, Météo-France
M. Gilles GRANDJEAN, Co-directeur de thèse, Bureau de Recherches Géologiques et Minières

Membres invités
M. Jean-Philippe Naulin, Caisse Centrale de Réassurance

Résumé

Le retrait-gonflement des argiles (RGA), aussi appelé sécheresse géotechnique, est un phénomène naturel de rétractation des sols riches en minéraux argileux gonflants lors de périodes de sécheresse. Le bâti peut être affecté si le dessèchement se propage en profondeur dans le sol sous les fondations provoquant la fissuration des façades. En France, les dégâts sont couverts par les assureurs dans le cadre du régime d'indemnisation des catastrophes naturelles (CatNat). Il s'agit du deuxième risque le plus coûteux après les inondations, avec des dommages estimés à plus de 20,8 milliards d'euros depuis 1989. Les coûts sont particulièrement élevés depuis 2017, avec un nouveau record en 2022 estimé à plus de 3 milliards d'euros d'après la Caisse Centrale de Réassurance (CCR). Bien que ce phénomène ait été décrit localement, il reste imprévisible à de grandes échelles spatiales, en partie à cause de l'hétérogénéité inhérente aux dépôts argileux. La thèse vise à améliorer la caractérisation des facteurs d'occurrence du RGA et à quantifier statistiquement le nombre global de sinistres. Il ne s'agit pas ici de déterminer un risque de sinistre pour chaque maison. Différents types de données sont comparés : des simulations de l'humidité du sol (Météo-France), des données géologiques et des acquisitions faites sur un site instrumenté dans le Loiret (BRGM), et une base de données de sinistres (CCR). Le travail est abordé sous l'angle du facteur déclenchant principal : la sécheresse. Un indicateur sur-mesure est développé, la magnitude annuelle de sécheresse (YDMI), basée sur l'humidité du sol simulée par le modèle de surface terrestre ISBA. Le YDMI intègre les fortes anomalies négatives d'humidité du sol sur une année. Ses caractéristiques (définition de la sécheresse, profondeur de la couche de sol et méthode de calcul) ont été ajustées par comparaison avec un échantillon de données de sinistres. L'objectif de cet indicateur est d'identifier les conditions de sécheresse susceptibles d'induire du retrait. Une comparaison des humidités du sol simulées par ISBA et des YDMI associés avec des données du site instrumenté a permis de caractériser les incertitudes liées à l'échelle macroscopique du modèle. Des YDMI sont ensuite calculés de façon spatialisée sur la France, d'abord pour une période historique allant de 2000 à 2022, puis pour une période future allant jusqu'à 2065. Pour cela, ISBA est forcé successivement par la réanalyse SAFRAN, et par les champs atmosphériques simulés par plusieurs modèles de climat sous les scénarios RCP 4.5 et 8.5. L'analyse historique identifie des événements extrêmes qui ont frappé la France en 2003, 2018, 2019, 2020 et particulièrement en 2022. D'autre part, les YDMI projetés montrent que la fréquence et l'intensité médianes des sécheresses favorisant le RGA sont amenées à augmenter dans le futur, surtout dans le scénario le plus pessimiste RCP8.5. Le dernier volet de la thèse met en interaction le YDMI et trois composantes géologiques de susceptibilité des sols au RGA identifiées par le BRGM, qui sont des critères qualifiant leur minéralogie, lithologie et propriétés géotechniques. Un des enjeux pour les assureurs est de prédire le nombre de sinistres causés par une année de sécheresse. C'est pourquoi le YDMI par années et les critères géologiques ont été croisés avec des données assurantielles pour entraîner un modèle ensembliste d'apprentissage random forest. Les résultats soulignent le pouvoir prédictif supérieur des YDMI et du critère minéralogique pour prédire le nombre de sinistres global. Ces travaux constituent une avancée vers une meilleure compréhension du phénomène RGA et ouvrent des perspectives pour la prédiction opérationnelle de la sinistralité assurantielle. La modélisation pourrait être améliorée avec des données d'aléa plus fines et une meilleure caractérisation de la vulnérabilité des bâtiments.

Abstract

Clay shrinkage is the retraction movements of soils rich in expansive minerals during dry periods. Buildings can be affected if the desiccation extends in depth into the ground below the foundations, causing cracks to appear in facades. In France, this damage, known as clay shrinkage-induced subsidence, is covered by insurers in the framework of the national compensation scheme for natural disasters (NatCat regime). It is the second most costly risk after floods, with damages estimated at more than 20.8 billion euros since 1989. Costs have been particularly high since 2017, with a new record set in 2022 for which damages were estimated at more than 3 billion euros, according to the Caisse Centrale de Réassurance (CCR). Although this phenomenon has been described locally, it remains unpredictable at large spatial scales, in part due to the inherent heterogeneity of clayey soils. The aim of this PhD work is to improve the characterization of clay shrinkage occurrence factors and to statistically quantify the global number of insurance claims. The aim here is not to determine a risk of damage for each house. Different types of data are compared: soil moisture simulations (Météo-France), geological data and acquisitions made at an instrumented site in the Loiret French department (BRGM), and insurance claims (CCR). To address the problem, we focus first on the main driver of clay shrinkage: drought. A custom drought index, the year drought magnitude index (YDMI), is developed based on soil moisture simulated by the ISBA land surface model. The YDMI integrates strong negative soil moisture anomalies over one year. Its characteristics (drought definition, soil layer depth and calculation method) were adjusted by comparison with a sample of insurance claims data. The purpose of such an index is to identify drought conditions that are likely to trigger clay shrinkage. A comparison of soil moisture simulated by ISBA and associated YDMI with data from the instrumented site enabled us to characterize the uncertainties associated with the macroscopic scale of the model. The YDMI is calculated for the whole country, first for a historical period 2000-2022, and then for a future period until 2065. For this purpose, ISBA is successively forced by the SAFRAN reanalysis, and by atmospheric fields simulated by several downscaled climate models under RCP scenarios 4.5 and 8.5. The historical assessment identifies extreme events that hit France in 2003, 2018, 2019, 2020 and particularly 2022. On the other hand, the projected magnitudes show that the median frequency and intensity of droughts favoring clay shrinkage are bound to increase in the future, especially under the most pessimistic scenario RCP8.5. The last part of the thesis brings YDMI into interaction with three geological components of soil susceptibility to shrinkage identified by BRGM, which are criteria qualifying mineralogy, lithology and geotechnical properties. One of the challenges for insurers is to predict the number of claims caused by a drought episode. For this reason, YDMI per year and geological criteria were cross-tabulated with insurance data to train a random forest supervised learning ensemble model. The results underline the superior predictive power of YDMI and of the mineralogy criterion. This work therefore represents a step towards a better understanding of the clay shrink-swell phenomenon, and opens up new prospects for the operational prediction of the global number of insurance claims. The modeling could be improved with finer hazard data and better characterization of building vulnerability.

Remerciements

Une page bien remplie se tourne avec la publication de ce manuscrit. Trois années, deux villes, deux équipes – cette thèse fut riche en rencontres et l'expérience n'en a été que plus belle. Cette page est dédiée à celles et à ceux avec qui j'ai partagé un bout de chemin. J'oublierai sûrement du monde, je m'en excuse par avance !

Pour commencer, je souhaite remercier les rapporteures et les membres du jury pour leur intérêt pour les travaux et les échanges riches lors de la soutenance. Ensuite, merci Jean-Christophe de m'avoir fait confiance en encadrant la thèse. J'ai eu beaucoup de chance d'avoir un directeur aussi impliqué et disponible ! J'espère que l'occasion de travailler ensemble se représentera à l'avenir. Merci Gilles, d'avoir été un co-directeur si bienveillant et positif. Ce sera un plaisir de continuer à se croiser au BRGM. Mes pensées se tournent ensuite vers David, disparu en 2023, pour son encadrement dans la bonne humeur. Merci Dorothée pour le suivi, et merci Jean-Philippe d'avoir su être présent lors de la dernière ligne droite tout en ayant pris le flambeau en cours de route.

Ensuite, je voudrais remercier mes collègues, sans qui le quotidien aurait été plus morne. D'abord les toulousain(e)s du couloir Végéo-Surface, qui m'ont fait voir la vie en rose pendant deux ans. Les docteur(e)s : Timothée, Tanguy, Gaëtan, Malak, (Adèle, bientôt !) – on l'a fait ! Merci aux AA pour les pauses inter-équipes, sorties, et week-ends chez Kirk. Thanks Daniel and Oscar for the ephemeral Vegeo running club. Merci Gaëtan pour notre amitié, sur fond de critiques ciné/bouquins et The Office. Pardon Bertrand de t'avoir torturé avec mes questions. Et une dernière pensée pour Belén, Alexis, Emma, Jazmin, je vous souhaite le meilleur pour la suite et j'espère que nous resterons en contact !

Ensuite les RIGien(ne)s, merci pour votre énergie et votre bonne humeur ! Ça me fait vraiment plaisir de continuer l'aventure avec vous. En particulier, en plus de l'encadrement, merci Séverine de m'avoir fait confiance depuis le début, et Sébastien pour la suite ! Une dédicace à Claire, notre happiness manager (*« assistante plus »*) indispensable. Bravo encore, les Dr Lucas et Lucie G., et courage, Lucie A. et César, vous allez cartonner !

Quelques lignes dédiées à mes co-bureaux d'exception. Timothée, ton seul défaut aura été de ne pas apprécier Mr Chatouille à sa juste valeur. Fengfeng, tu es une amie géniale, vivement que tu reviennes à Orléans pour qu'on fasse des raviolis et qu'on décore des bureaux !

Merci aux autres ami(e)s d'Orléans : Lucie G. pour les discussions-champignons, David et Lucie A. pour l'escalade, et Luca, Thibault, Lucas, Lucille (+Juju !) pour les soirées jeux, fromage ou séries, éclaircies dans la grisaille loirétaine. Luca, on a fait le même parcours, je suis très heureuse de te connaître, et j'espère qu'on coïncidera encore un moment !

Pour continuer, merci Adri pour tant de choses, entre autres de m'avoir soutenue pendant cette rédaction infernale, de faire preuve d'une patience *infinie* pour mes problèmes informatiques et LaTeX (eux aussi *infinis*), en plus d'être mon fidèle acolyte de vélo. Trois lignes c'est trop court pour dire à quel point tu es extra.

Enfin, merci à la famille, Maman, Papa, Chantal, Grégoire, Marguerite, Nicolas, Hélène et Irène (+ les petits !) qui êtes (presque tous) présents depuis le début. Maman, c'est à toi que je veux dédicacer ce manuscrit, je te dois beaucoup !

Contents

List of Figures

List of Tables

Acronyms

C_s Swelling Index.

E_w Shrink-swell modulus.

W_{fc} field capacity.

W_{wilt} wilting point.

AUC area under curve.

BRE Building Research Establishment.

BRGM Bureau de Recherches Géologiques et Minières.

CCR Caisse Centrale de Réassurance.

CNRM Centre National de Recherches Météorologiques.

EDO European Drought Observatory.

FDR Frequency Domain Reflectometry.

FN false negative.

FP false positive.

FPR false positive rate.

GCM Global Climate Model.

HWSD Harmonized World Soil Database.

INSEE Institut National de la Statistique et des Études Économiques.

IPCC Intergovernmental Panel for Climate Change.

ISBA Interactions Sol-Biosphère-Atmosphère.

LAI Leaf Area Index.

LL Liquid Limit.

LSM Land Surface Model.

MBA Methylene Blue Adsorption.

MRN Mission Risques Naturels.

PFT Plant Functional Type.

PI Plasticity Index.

PL Plastic Limit.

RCM Regional Climate Model.

RCP Representative Concentration Pathway.

RGA retrait-gonflement des argiles.

RMSE root mean square error.

ROC Receiver Operating Characteristic.

SIM SAFRAN-ISBA-MODCOU.

SL Shrinkage Limit.

SURFEX SURFace EXternalisée.

SWCC Soil Water Characteristic Curve.

SWI Soil Wetness Index.

TN true negative.

TP true positive.

TPR true positive rate.

WG Soil Liquid Water Content.

XRD X-Ray Diffraction.

YDMI Yearly Drought Magnitude Index.

Introduction Générale

La Terre se distingue dans le Système Solaire par la quantité d'eau importante à sa surface, ce qui lui vaut le surnom de « Planète Bleue ». Elle y est présente sous les trois états solide, liquide et gazeux, au sein de différents réservoirs formant l'hydrosphère. Le volume le plus important (96,5%) se trouve à l'état liquide et salé dans les océans. L'eau douce, vitale pour les besoins de l'humanité, ne représente que 2,5% de l'eau sur Terre, ce qui correspond à 35 millions de km^3. La majorité (68,7%) de cette eau douce est à l'état solide dans les glaciers et neiges permanentes. Un tiers (30,15%) se trouve dans le sol, dont 0,05% sous forme d'humidité dans la zone non saturée située entre la surface et les nappes souterraines [Gleick et al., 1993].

Les différents réservoirs d'eau interagissent les uns avec les autres tandis que l'eau circule des continents aux océans et inversement sous l'effet de la gravité et de l'énergie solaire. On appelle cela le cycle de l'eau. Il est étroitement lié au bilan énergétique de la Terre. Les flux entre les réservoirs du cycle de l'eau varient au fil du temps. Une partie de cette variabilité s'explique par des phénomènes climatiques de grande échelle, tels que l'Oscillation Nord-Atlantique (NAO) [Rodó et al., 1997]. Il est désormais établi que le changement climatique influence le cycle de l'eau, car l'augmentation de la concentration de gaz à effet de serre modifie le bilan énergétique de la Terre [IPCC, 2023]. Une augmentation ou une diminution des flux entre les réservoirs peut entraîner des événements extrêmes du type inondations ou sécheresses, respectivement [Oki et al., 2004].

L'Organisation Météorologique Mondiale (OMM) définit une sécheresse comme une période sèche prolongée dans le cycle climatique naturel, qui peut survenir n'importe où dans le monde. Il s'agit d'une définition opérationnelle parmi d'autres, car à ce jour, il n'existe pas de critère universel sur la question. D'après Dracup et al. [1980], c'est en raison de la complexité du phénomène. En effet, une sécheresse peut concerner différentes étapes du cycle de l'eau, des précipitations aux débits de rivière, à des échelles de temps différentes. Cette complexité est exacerbée par la nécessité de définir des seuils par rapport aux normales. Le point commun à toutes les définitions est que la sécheresse exprime un manque d'eau par rapport à un besoin. Selon Wilhite and Glantz [1985], quatre types de sécheresse peuvent être distingués : (i) sécheresse météorologique, définie par rapport au régime de précipitations, (ii) sécheresse hydrologique, définie par rapport aux niveaux des lacs et des rivières, (iii) sécheresse agricole, définie par rapport à l'eau dans le sol et particulièrement ses impacts sur la végétation, et (iv) sécheresse socio-économique, définie par rapport aux effets du manque d'eau sur les activités humaines. Mishra and Singh [2010] ajoutent à cette liste (v) la sécheresse des eaux souterraines, aux dynamiques très lentes. Ces définitions, associées à des niveaux différents du cycle de l'eau, ne sont pas indépendantes, les réservoirs étant connectés [Changnon, 1987, Farahmand et al., 2021].

Ce que l'on nomme sécheresse agricole caractérise un déficit d'eau par rapport aux normales dans la zone non saturée du sol. Les impacts d'une sécheresse agricole peuvent être multiples, notamment car c'est dans ce réservoir que les plantes prélèvent de l'eau par le biais de leurs racines, sans laquelle elles flétrissent. La baisse de la productivité végétale et donc du rendement agricole sont les impacts les plus évidents de l'assèchement des sols superficiels. On

recense également des impacts sur le climat, car la végétation contribue à l'évapotranspiration globale [Seneviratne et al., 2010] et intervient dans le cycle du carbone [Ciais et al., 2005]. Les sécheresses agricoles augmentent également le risque de feux de forêt en asséchant la biomasse combustible. Il existe un dernier impact, moins connu mais d'une importance majeure : le phénomène de retrait-gonflement des argiles.

Le phénomène de retrait-gonflement des argiles désigne les variations de volume des sols riches en minéraux argileux gonflants sous l'effet de fluctuations d'humidité du sol. En particulier, le sol se rétracte en période de sécheresse agricole, ce qui cause un affaissement susceptible d'endommager les constructions, sous des climats à dominante humide. Des dégâts liés à ce phénomène ont été recensés et font l'objet de recherches scientifiques dans le monde entier, notamment au Royaume-Uni [MacQueen et al., 2023], aux États-Unis [Mostafiz et al., 2021], au Kenya [Kariuki and Woldai, 2004], en Italie [Meisina et al., 2006], ou encore à Chypre [Tzampoglou et al., 2022]. Les assureurs qualifient ce risque de « sécheresse géotechnique ».

En France, la sécheresse géotechnique occasionne des dégâts considérables. La Caisse Centrale de Réassurance (CCR) a estimé la sinistralité cumulée entre 1989 et 2023 à 23,9 milliards d'euros . Ils soulignent que le coût moyen annuel est particulièrement élevé depuis 2017. L'année 2022, marquée par une forte sécheresse, a accusé une sinistralité exceptionnelle estimée entre 3 et 3,5 milliards d'euros [CCR, 2023a]. L'indemnisation des dégâts liés à la sécheresse géotechnique est assurée dans le cadre d'un dispositif d'État appelé *Régime d'indemnisation des catastrophes naturelles*, ou régime CatNat. Plusieurs études comme Gourdier and Plat [2018] ou CCR [2023b] ont montré qu'une hausse des coûts liés à ce phénomène était attendue sous l'effet du changement climatique, en raison de l'assèchement du territoire combiné à l'augmentation du nombre de biens exposés.

Malgré ses impacts considérables, le phénomène de retrait-gonflement des argiles reste mal compris. Il est complexe par sa nature multifactorielle : l'amplitude du déplacement du sol autour d'un bâtiment dépend de sa nature, des conditions hydrologiques et de son environnement immédiat. Tous ces éléments présentent un niveau de variabilité spatiale élevé, ce qui complique les études à échelles régionales et nationales, pourtant nécessaires à la gestion du risque. De plus, il n'existe à ce jour aucune métrique de sécheresse qui soit adaptée de manière spécifique au suivi du phénomène de retrait-gonflement des argiles.

Il apparaît nécessaire de développer les connaissances sur le retrait-gonflement des argiles afin de mieux gérer ce risque, dont les impacts, déjà considérables, sont amenés à s'aggraver à l'avenir. En réponse à ces enjeux opérationels, cette thèse vise à mieux comprendre le phénomène et ses conséquences. Nous nous intéresserons à la fois au facteur déclencheur et aux facteurs de prédisposition. L'approche scientifique et le plan de thèse correspondant sont présentés en figure 1.

Le chapitre 1 donne des éléments de contexte sur le phénomène de retrait-gonflement des argiles et ses impacts. En plus du cadre scientifique, nous détaillerons la structure institutionnelle propre à la France pour son indemnisation.

Le chapitre 2 présente les données et outils de natures différentes qui serviront de base aux résultats. Ces données rassemblent des simulations d'humidité des sols, des cartes d'aléa liées aux natures de sols, des mesures provenant d'un site instrumenté et des données assurantielles.

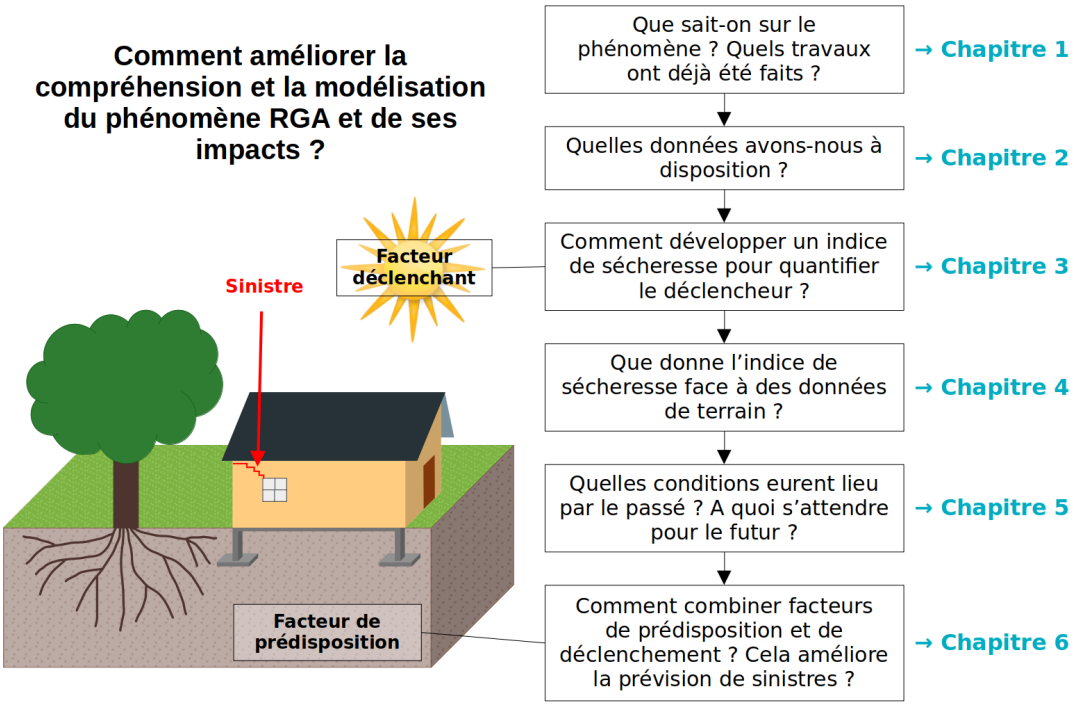


Figure 1 : Travail réalisé au cours de la thèse

Le chapitre 3 détaille les étapes de développement d'un indice de sécheresse, le Yearly Drought Magnitude Index (YDMI), qui vise à quantifier le facteur déclenchant du phénomène RGA. Le YDMI est basé sur les simulations d'humidité des sols et a été calibré par comparaison à un échantillon des données assurantielles.

Le chapitre 4 développe une comparaison effectuée entre simulations d'humidité, YDMI associés et données du site instrumenté. Cette analyse permet de caractériser les incertitudes inhérentes à la modélisation.

Le chapitre 5 présente les YDMI obtenus sur la France pour une période historique de 2000 à 2022, à partir d'une réanalyse atmosphérique, et pour une période future de 2006 à 2065, à partir de projections climatiques. La rétrospective permet de caractériser les évènements passés tandis que les projections montrent comment les déclencheurs sont susceptibles d'évoluer sous l'effet du changement climatique.

Le chapitre 6 décrit des expériences combinant YDMI et cartes d'aléa RGA pour prédire la sinistralité par le biais d'un modèle d'apprentissage automatique. Les résultats permettent d'évaluer la valeur ajoutée des travaux dans une perspective de prévision assurantielle.

General Introduction

The Earth stands out in the Solar System for the large quantity of water on its surface, earning it the nickname "Blue Planet". Water is present in three states: solid, liquid and gaseous, in the various reservoirs that make up the hydrosphere. The largest volume (96.5%) is found in oceans in a liquid and salty state. Freshwater, vital for human needs, represents only 2.5% of the Earth's water, or 35 million km^3. The majority (68.7%) of this freshwater is in solid form in glaciers and permanent snow. A third (30.15%) is in the soil, with 0.05% in the form of moisture in the unsaturated zone between the surface and the groundwater table [Gleick et al., 1993].

The various water reservoirs interact with each other as water flows from the continents to the oceans and back again, driven by gravity and solar energy. This is known as the water cycle. It is closely linked to the Earth's energy balance. Flows between water cycle reservoirs vary over time. Part of this variability is explained by large-scale climatic phenomena, such as the North Atlantic Oscillation (NAO) [Rodó et al., 1997]. It has now been established that climate change influences the water cycle, as rising greenhouse gas concentrations modify the Earth's energy balance [IPCC, 2023]. An increase or decrease in flows between reservoirs can lead to extreme events such as floods or droughts, respectively [Oki et al., 2004].

The World Meteorological Organization (WMO) defines a drought as a prolonged dry period in the natural climatic cycle, which can occur anywhere in the world. This is just one operational definition, as to date there is no universal criterion. According to Dracup et al. [1980], this is due to the complexity of the phenomenon. Indeed, a drought can affect different stages of the water cycle, from precipitation to river flow, at different timescales. This complexity is exacerbated by the need to define thresholds in relation to normals. What all definitions have in common is that drought expresses a lack of water in relation to a need. According to Wilhite and Glantz [1985], four types of drought can be distinguished: (i) meteorological drought, defined in relation to precipitation patterns, (ii) hydrological drought, defined in relation to lake and river levels, (iii) agricultural drought, defined in relation to water in the soil and particularly its impact on vegetation, and (iv) socio-economic drought, defined in relation to the effects of water scarcity on human activities. Mishra and Singh [2010] add to this list (v) groundwater drought, with its very slow dynamics. These definitions, associated with different levels of the water cycle, are not independent, reservoirs being connected [Changnon, 1987, Farahmand et al., 2021].

What we call agricultural drought characterizes a deficit of water compared with normal levels in the unsaturated zone of the soil. The impact of an agricultural drought can be diverse, in part because it is from this reservoir that plants draw water through their roots, without which they wither. The most obvious impact of surface soil drying is a drop in plant productivity and, consequently, agricultural yields. There are also impacts on the climate, as vegetation contributes to global evapotranspiration [Seneviratne et al., 2010] and plays a part in the carbon cycle [Ciais et al., 2005]. Agricultural drought also increases the risk of forest fires by drying out biomass. A final, lesser-known impact of major importance is the phenomenon of clay shrink-swell.

The phenomenon of clay shrink-swell refers to variations in the volume of soils rich in expansive clay minerals due to fluctuations in soil moisture. In particular, the soil shrinks during periods of agricultural drought, causing subsidence that can damage buildings in predominantly humid climates. Damage linked to this phenomenon has been recorded and is the subject of scientific research worldwide, including in the UK [MacQueen et al., 2023], the USA [Mostafiz et al., 2021], Kenya [Kariuki and Woldai, 2004], Italy [Meisina et al., 2006], and Cyprus [Tzampoglou et al., 2022]. Insurers refer to this risk as "geotechnical drought".

In France, geotechnical drought is causing considerable damage. The Caisse Centrale de Réassurance (CCR) has estimated the cumulative loss between 1989 and 2023 at 23.9 billion euros. They point out that the average annual cost has been particularly high since 2017. The year 2022, marked by severe drought, had damage estimated at between 3 and 3.5 billion euros. Compensation for damage caused by geotechnical drought is provided in the framework of a government scheme known as *Régime d'indemnisation des catastrophes naturelles*, or NatCat regime. A number of studies, such as Gourdier and Plat [2018] or CCR [2023b], have shown that the costs associated with this phenomenon are expected to rise as a result of climate change, due to the drying-up of the soil combined with an increase in the number of houses exposed.

Despite its considerable impact, the phenomenon of clay shrink-swell remains poorly understood. It is complex because of its multi-factorial nature: the amplitude of soil movement around a building depends on its nature, hydrological conditions and the immediate environment. All these elements present a high level of spatial variability, which complicates studies on regional and national scales, which are nonetheless necessary for risk management. In addition, there is currently no drought metric specifically adapted to monitoring the phenomenon of clay shrink-swell.

It appears necessary to develop knowledge of clay shrink-swell in order to better manage this risk, whose already considerable impacts are set to worsen in the future. In response to these operational challenges, this thesis aims at gaining a better understanding of the phenomenon and its consequences. We will look at both the triggering factor and the predisposing factors. The scientific approach and corresponding thesis plan are presented in Figure 2.

In chapter 1, we will first provide background information on the phenomenon of clay shrinkage and swelling and its consequences. In addition to the scientific context, the institutional situation specific to France will be presented.

In chapter 2, we will present the different types of data and tools on which the work will be based. These include soil moisture simulations, expansive soil hazard maps, measurements from an instrumented site and insurance data.

In chapter 3, we will describe the development of a drought index, called Yearly Drought Magnitude Index (YDMI), to quantify the triggering factor for clay shrinkage. It is based on the soil moisture simulations and has been calibrated by comparison with a sample of insurance data.

In chapter 4, we develop a comparison between the soil moisture simulations, derived YDMI and data from the instrumented site. This analysis allows to characterize the uncertainties inherent to modeling.

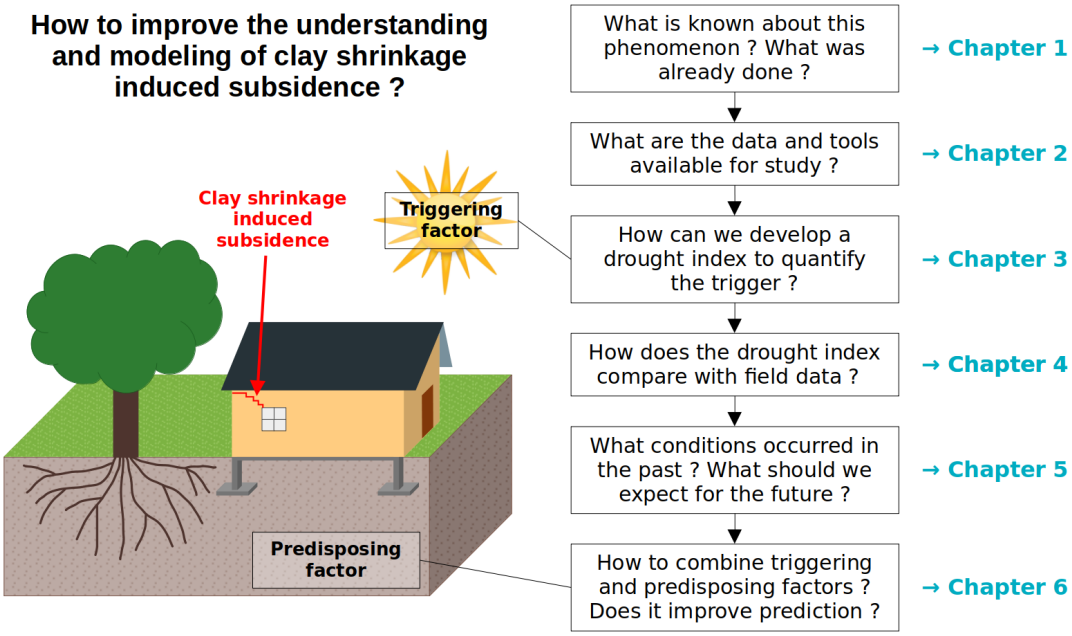


Figure 2: Work done during the thesis

In chapter 5, we will present the YDMI obtained all over France for a historical period using an atmospheric reanalysis, and for a future period using climate models. The retrospective will enable to identify past trends while the projections will show how the triggering conditions are likely to evolve under the influence of climate change.

In chapter 6, we will describe an experiment combining YDMI and expansive soil hazard maps to predict claims using a machine learning model. The results will allow us to assess the added value of the work from an insurance forecasting perspective.

Chapter 1

Scientific and institutional context

Contents

1.1 Introduction

Clay shrinkage and swelling result from interactions between soils and the atmosphere. It can damage buildings, depending on their characteristics. The multi-factor nature of this phenomenon will be detailed in section 1.2 to give an exhaustive overview of the issue. Knowing these elements, different methods exist for identifying soils subject to the phenomenon, tracking its evolution over time, and model it. We will detail some of these in section 1.3, which is useful as field data will be processed later on in the thesis. Then, since clay shrinkage and swelling is a climate-driven phenomenon, it is likely to be affected by global warming. We will give a review of the state of the art on the topic in section 1.4 as this aspect will be addressed in the thesis. Finally, the institutional context by which natural disasters are compensated in France is described in section 1.5. Knowing how this system works brings elements for understanding results better.

1.2 A multi-factor phenomenon

Clay shrink-swell occurs when soils with a structure that allows volume variations are exposed to specific environmental conditions. The factors of occurrence can be divided into two categories: predisposing factors, changing little over time, and triggering factors, which condition the onset of the phenomenon. In a specific context, the volume variations can damage buildings, causing cracks of varying depth to appear on walls.

1.2.1 Predisposing factor

1.2.1.1 Expansive soils

Swelling clay minerals

Clay minerals, which belong to the phyllosilicate (*"layered silicate"*) group, are composed of layers formed by rows of tetrahedra of silicon atoms and octahedra of aluminum, iron or magnesium atoms. They are the finest particles found in soils, with a size in the micrometer range. There is a considerable diversity of clay minerals due to the possible variations in layer composition and arrangement. Three groups have been defined according to the nature of the layers (shown on Figure 1.1):

- 1:1 clays, where the layers consist of a row of tetrahedra joined to a row of octahedra. Kaolinite belongs to this family.
- 2:1 clays, where the layers consist of a row of octahedra intercalated between two rows of tetrahedra. Smectites, including montmorillonite, or illite, belong to this family. Only 2:1 layers carry a negative electrical charge: the neutrality of the mineral is ensured

by the presence of cations (e.g. sodium or calcium ions for smectites) in the interlayer space.

- 2:1:1 clays, alternating 2:1 layers and rows of octahedra. Chlorite belongs to this family.

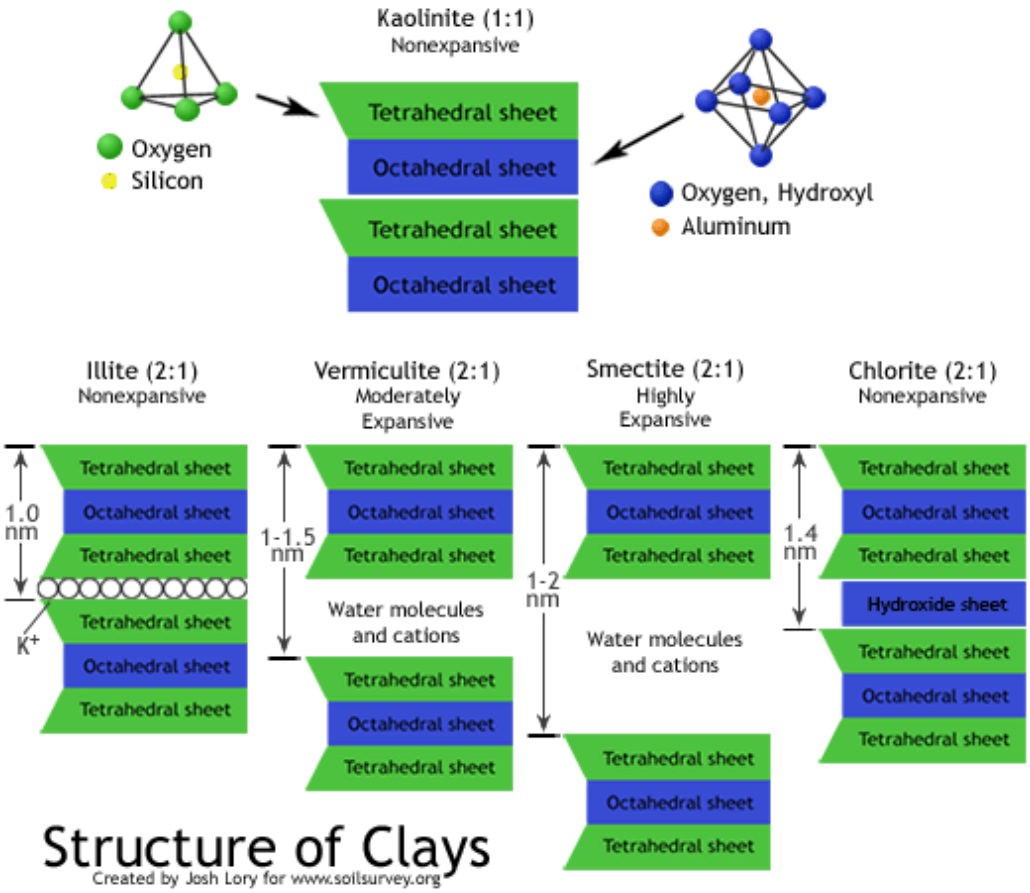


Figure 1.1: Structure of different clay minerals and consequences on the swelling potential. Figure from Soil Observation Institute [2024].

Because of this layered structure, clay minerals have a large surface area, called the *specific surface area*, and therefore an important water-holding capacity. It is especially high for smectite minerals which absorb water not only on their outer surface but also between layers, since the interlayer space is large and the interactions weak enough [Earle, 2021]. Water adsorption causes swelling of the mineral: its volume can double or even triple [Tessier et al., 2006, Olive et al., 1989]. The affinity to water molecules also depends on electrical charge and therefore on the nature of interlayer cations: for instance, sodium montmorillonite swells more than calcium montmorillonite [Nowamooz, 2007].

Within natural soils

Natural soils consist of a skeleton made of various particles *(clay minerals, sand, silt and organic matter)* and voids forming the porosity, filled either by water or air. Only clay minerals and organic matters are concerned by volume changes [Seidel et al., 2023] occurring

as in Figure 1.2. Volume variations occur in all three dimensions. The overall swelling capacity of the soil is positively correlated with the proportion of expansive elements.

If the expansive clayey layers are discontinuous, volume variations will not be as pronounced, as they are damped by inter-layers. The presence of a surface load can also limit expansion by countering swelling pressures. Third, cementation (with, for example, carbonates, silica, or iron oxide) limits swelling potential due to both the rigid structure formed by the cement and the reduced porosity [Olive et al., 1989, Vincent et al., 2009].

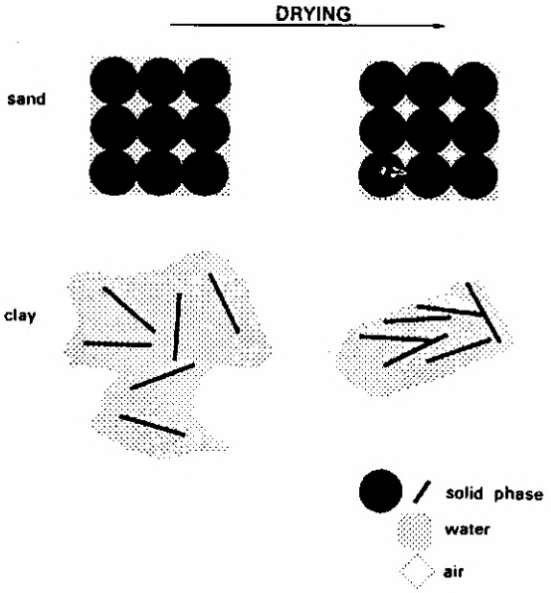


Figure 1.2: Schematic representation of the process of drying and air entry in sandy soils and clay soils. Figure from Bronswijk [1989].

1.2.1.2 Location of expansive soils

The geographical distribution of expansive soils is not random, but is a consequence of their formation process and transport mode.

Geological formation of clay minerals

Clay minerals are formed by the hydrolysis (alteration by water) of silicate minerals, which are widely distributed, being the major constituents of 90% of the rocks that make up the Earth's crust [Klein et al., 2002]. This hydrolysis can occur either at the surface or at depth. The type of mineral formed depends on the initial silicate, the pressure and temperature conditions, and the chemistry of the alteration fluid. For example, smectite and kaolinite are formed under surface conditions, and deep burial is required to transform them into chlorite [Earle, 2021]. It is a slow process, dependent on climatic conditions, occurring on the time scales of tens of thousands to hundreds of thousands of years [Price et al., 2005].

Transport and sedimentation

Because of their small size, clay particles are easily transported. They settle in calm conditions, when their weight prevails over the current in the balance of forces (e.g. in a lake or puddle). The agglomeration of clay minerals in the presence of salt water, called flocculation, accelerates this process. Given the transport dynamics of clay minerals, topography should be considered when mapping expansive soils [Tessier et al., 2006].

Distribution in soils

The vertical distribution of clay minerals within the soil column is affected by leaching, a process by which fine particles are carried by rainwater from the surface (eluviation) to the deeper layers (illuviation). It is favored by a weak structure and a high connected porosity [JRC, 2024].

1.2.1.3 Modifying the expansive potential

Research is aimed at modifying soil characteristics, including expansive properties, for engineering application. For example, Stoltz et al. [2012] has shown that lime treatment reduces the swelling potential of the soil by modifying its microstructure. However, it is not effective in reducing the amplitude of shrinkage. Also, such treatment has not proven to be stable over time, based on a long-term analysis of an embankment [Chabrat et al., 2023].

1.2.2 Triggering factor

1.2.2.1 Drought

Exceptional variations in soil volume can be caused by either shrinkage or swelling, depending on the usual moisture content. Western Europe is characterized by a climate that is wet all year to dry only in summer [Kottek et al., 2006], which means that soils are in a hydrated state and swollen most of the time. In this region the triggering element being dry conditions [Cojean, 2007, Meisina et al., 2006]. Reports of building damage occurring in the summer and fall [Vincent et al., 2009, Fernandes et al., 2015] support this assertion.

At which depth ?

Several studies based on instrumented sites have investigated the depth to which the soil dries every year. Fernandes et al. [2015] found significant variations up to 3 meters, favored by the presence of local sandy drains at their study site in southwestern France. Using roadside suction measurements, Ighil Ameur [2023] observed that the dry front reached 1 m in 2020 and 3 m in the exceptionally hot and dry 2022 in central France. The presence of a stable water table limits the vertical extent of drying, as shown by Welch et al. [2024] based on a 4 meter deep water table in their study site in Texas.

1.2.2.2 Unsaturated soil hydraulics

As previously stated in section 1.2.1.1, soils are composed of particles of varying shapes and sizes, as well as voids in which water or air can circulate. The water can be in two states: bound water, which is strongly attached to the surface of the grains by capillary forces, and free water, which is out of their range of attraction [De Marsily, 1986].

Soil moisture and matric potential

The force that binds the water to the grains is a negative pressure referred to as suction, or matric potential (abbreviated h or ψ). It is positive when given as suction and negative when expressed as potential. The drier the soil, the stronger the force, which means that matric potential varies with soil moisture (abbreviated θ). The relationship between the two is non-linear, and is referred to as the Soil Water Characteristic Curve (SWCC). The intensity of suction is also positively correlated with the specific surface area, which is particularly important for clay minerals, hence the SWCC varies with soil texture.

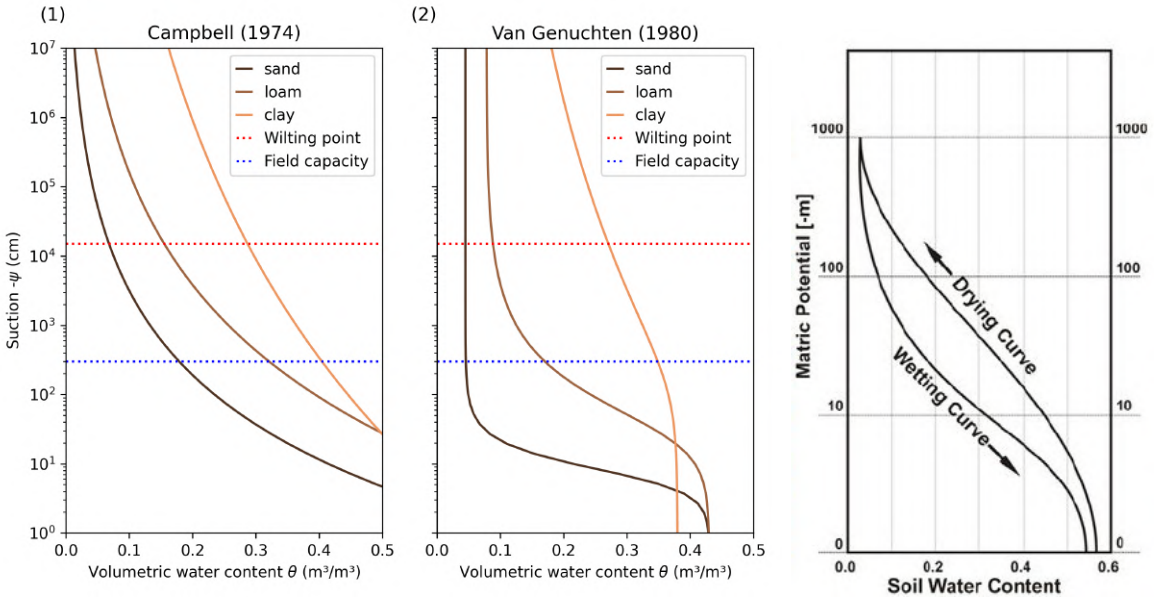
Equations for the Soil Water Characteristic Curve have been published in several works. Campbell [1974] proposed a power law equation, specified in Equation 1.1. On the other hand, Van Genuchten [1980] defined a more complex equation, developed in Equation 1.2. Both

equations are commonly used. The parameters θ_{sat} and ψ_{sat} refer to values of soil moisture and matric potential at saturation, respectively, θ_{res} is the residual water content in the dry soil, while α is the inverse of the air entry potential and n is the pore-size distribution index. The hydraulic parameters θ_{sat}, ψ_{sat}, θ_{res} and shape parameters b, α and n are dependent on the soil texture. Values have been empirically derived for the different soil classes defined by the United States Department of Agriculture (USDA) (see Figure A.1). They are given in Table A.1.

$$\psi = \psi_{sat} \left(\frac{\theta}{\theta_{sat}} \right)^{-b} \tag{1.1}$$

$$\frac{\theta - \theta_{res}}{\theta_{sat} - \theta_{res}} = (1 + (\alpha|\psi|)^n)^{-m}, \; with \; m = 1 - \frac{1}{n} \tag{1.2}$$

The SWCC for both equations and for different soil texture classes is represented in Figure 1.3a. The suction values associated with notable moisture states are indicated as well: -3 m for the field capacity (W_{fc}), the state of maximum water retention after gravitational flow, and -150 m for the permanent wilting point (W_{wilt}), the limit below which plant roots are no longer able to absorb water. Both of these states are associated with higher moisture contents for clay than for sand, due to the greater capillary forces exerted by particles.



(a) SWCC for (1) Campbell [1974] and (2) Van Genuchten [1980] equations

(b) SWCC for a wetting and drying path

Figure 1.3: Soil Water Characteristic Curve (SWCC) considering (a) different equations and soil textures and (b) wetting and drying paths. Subfigure (b) is from Tuller [2005].

It is interesting to note that the SWCC depends on the direction of the water content variation, as shown by the hysteresis on Figure 1.3b. The difference is attributed to the presence

of air bubbles that remain trapped when a dry soil is wetting, modifying the contact surface between grains and water. It is possible to move from the wetting curve to the drying curve by waiting long enough for the bubbles to disappear. Between these two curves exist countless possibilities for intermediate cycles.

Hydraulic conductivity

The hydraulic conductivity (abbreviated K) of a soil is defined as its capacity to let a fluid pass through under the influence of a pressure gradient. K decreases with increasing matric potential (h or ψ) and therefore decreasing moisture content. This phenomenon can be attributed to the filling of pores by air: under these conditions water circulates mainly by vapor diffusion, which is much slower than liquid flow.

As for the SWCC, the relation between hydraulic conductivity K and soil water content θ is non-linear and changes with soil texture. Campbell [1974] and Van Genuchten [1980] have also defined equations for the $K(\theta)$ curve. They are detailed in Equation 1.3 and Equation 1.4, respectively.

$$K = K_{sat} \left(\frac{\theta}{\theta_{sat}} \right)^{2b+3} \tag{1.3}$$

$$K = K_{sat} \left(\frac{\theta - \theta_{res}}{\theta_{sat} - \theta_{res}} \right)^{\frac{1}{2}} \left(1 - \left(1 - \left(\frac{\theta - \theta_{res}}{\theta_{sat} - \theta_{res}} \right)^{\frac{1}{m}} \right)^{m} \right)^{2}, \; with \; m = 1 - \frac{1}{n} \tag{1.4}$$

The parameters θ_{sat}, θ_{res}, b and n are derived from the soil texture as explicited in Table A.1. The evolution of hydraulic conductivity as a function of soil water content for both equations and different soil textures is represented on Figure 1.4.

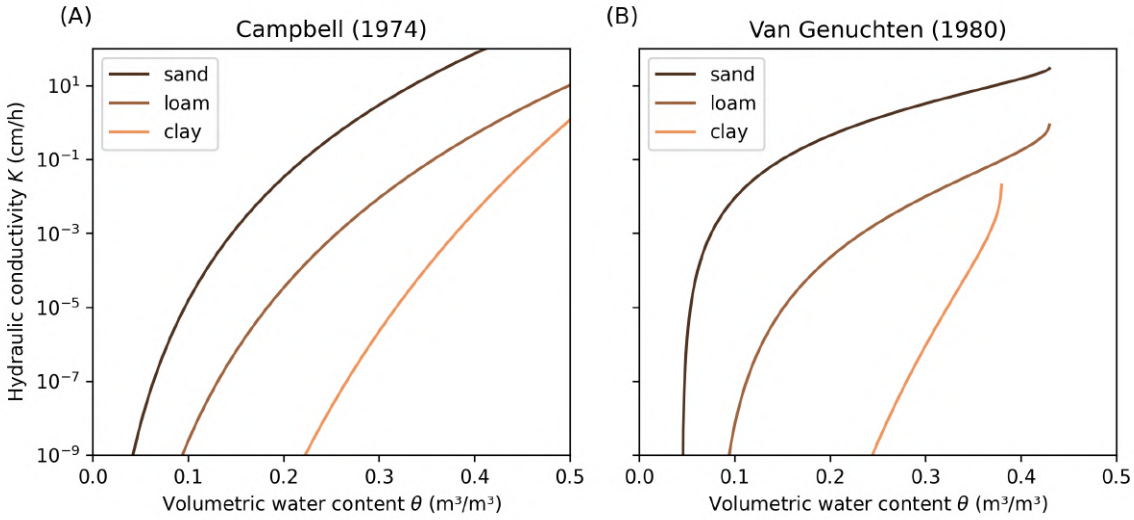


Figure 1.4: Evolution of hydraulic conductivity K as a function of soil water content θ from (A) the Campbell [1974] and (B) the Van Genuchten [1980] equations, and different soil textures.

As explained above, hydraulic conductivity decreases with moisture content. Also, sandy soil has a higher conductivity than clay soil.

Kinetics

Water circulation in soils is a very slow process [De Marsily, 1986] exacerbated by the low permeability of clay soils. For this reason, soils that theoretically have a high expansive potential may not shrink or swell much in the field [Olive et al., 1989].

1.2.2.3 The effect of vegetation

The presence of vegetation favors shrinkage...

Plants have a significant surface area in contact with the atmosphere through their leaves, resulting in increased evaporation called transpiration. They compensate for the deficit by pumping water through their root network, increasing soil drying, and therefore shrinkage if the soil has expansive properties.

This effect is particularly pronounced for trees, which dry the soil to a greater depth than shorter vegetation types: the desiccation can reach a maximum of 6 m for certain tree species, versus 1.5 m for grass [Vincent et al., 2009]. This was measured on the field in the 1990s by the British Building Research Establishment (BRE), as shown on Figure 1.5 [Hawkins, 2013, Freeman et al., 1992].

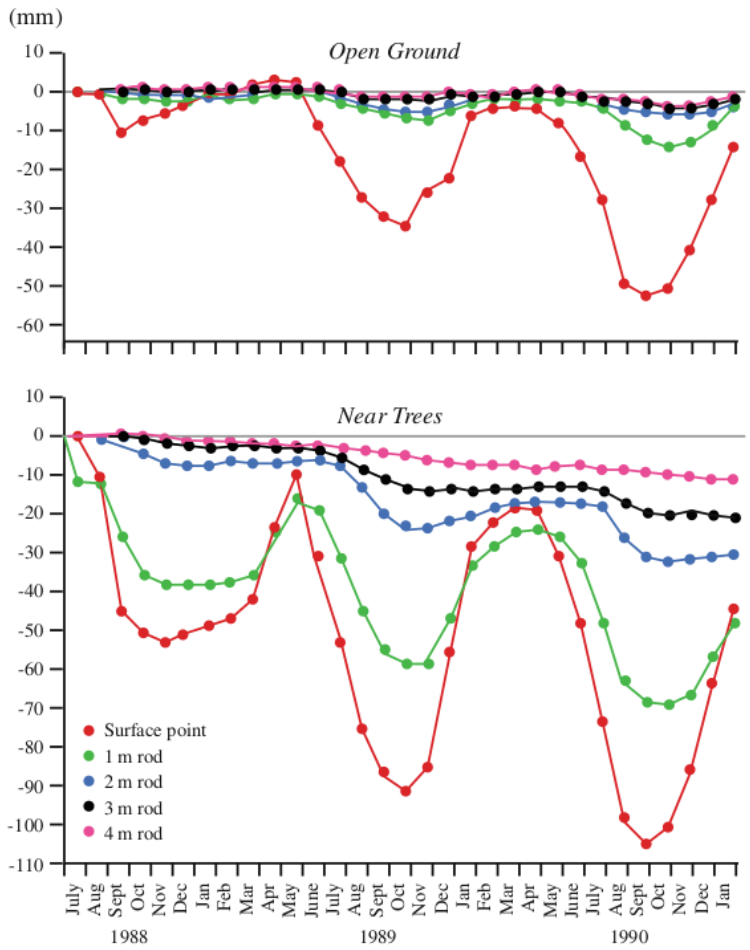


Figure 1.5: Effect of trees on vertical soil movement investigated by the BRE: variations of the level at different depths, in open ground and near trees. Figure from Hawkins [2013] based on Freeman et al. [1992]. The vertical axis is in mm.

Water flow is conditioned by species-specific characteristics related to leaves (phenology, leaf area) and roots (depth, spatial distribution and extraction capacity), as well as the woody character of stems [Tessier et al., 2006]. Drying is

more intense for broad-leaved trees such as oaks, poplars, willows, or elms, with high moisture demand, than for conifers [Page, 1998].

Roots of isolated trees can extend laterally beyond the area circumscribed by the canopy, when in forests it is limited by competition [Ravina, 1983]. Because clayey layers have a greater water retention capacity, they are favorable to the development of roots [Tessier et al., 2006]. Conversely, soil structure elements such as texture heterogeneities can influence root network development [Vincent et al., 2009].

... and its removal causes swelling

When the soil permeability is very low, winter precipitation is not always sufficient to recharge the subsurface, which over time can result in permanently dry soil around plant roots. When this vegetation is removed, swelling occurs. Cases of building damage caused by swelling have been identified in France and the UK in this context [Page, 1998, Driscoll, 1983], although it is much less common than damage caused by shrinkage.

1.2.2.4 Anthropogenic factors

In addition to the presence of vegetation, certain anthropogenic factors can affect the water content around a building, causing differential soil movement. For example, localized swelling can be caused by garden irrigation or leaks in the water or sewer system, and shrinkage can be caused by the proximity of heating systems or fireplaces [Olive et al., 1989].

1.2.3 Shrinkage-induced cracking feedback

Shrinkage of expansive soils causes 3-dimensional cracking, illustrated in Figure 1.6. The formation of a superficial network of cracks increases the permeability of the soil, allowing the drying and thus the shrinkage to extend even deeper [Ravina, 1983]. In fact, with cracking, exchange surfaces greatly expand as evaporation no longer occurs only on the surface of the soil, but also on the crack walls [Song and Cui, 2020]. Cracks can reach depths between 1 m and 1.5 m [Hawkins, 2013]. They close after the summer, as the soil rehydrates and swells annually, and reopen later at each drought episode.

Figure 1.6: Dessication cracks formed in the summer at the north-east border of the city of Toulouse (Photo by Bertrand Bonan, 22 August 2022)

The leaching of clay minerals in soils (mentioned in section 1.2.1.2) is exacerbated by cracking in the context of successive wet and dry seasons.

As explained by Page [1998], when shrinkage cracks open next to a building and contribute to its damage, they tend not to close completely afterwards due to the shifting of the foundation and the accumulation of debris in the crack.

1.2.4 Displacement

The amplitude of displacement caused by soil shrinkage and swelling has been quantified in several studies, all of them based on field measurements.

A first experiment carried out by Cooling and Ward [1948] in Great Britain measured the displacement of plates installed at different depths during a single summer. A subsidence of 27 mm was noted for the most superficial plate located at a depth of 30 cm. Also in Great Britain, Freeman et al. [1992] measured a surface displacement of 4 cm without trees and up to 10 cm in the presence of trees. In both cases, details of the instrumentation are unknown. In southwestern France, Fernandes et al. [2015] measured millimeter-scale displacement cycles (between 1 and 3 mm) at their study site, using devices called extensometers, that measure relative displacement to a reference point in depth. Finally, Welch et al. [2024] measured subsidence in the order of 10 to 20 cm consecutively to drought, at their study site in Texas. The instrumentation captured displacement based on GPS technology.

The disparity calls for attention to the measurement technique, some of which will be detailed in section 1.3.2.2.

1.2.4.1 Cumulative effect

Some of the studies cited above describe an irreversible pattern in soil volume variations. Fernandes et al. [2015] notice a cumulative shrinkage of their soil of several millimeters after a few drought-rewetting cycles, which they attribute to insufficient winter recharge. Welch et al. [2024], who make the same observation (see Figure 1.7), estimate that 10% of the shrinkage is irreversible. According to them, this inelastic deformation differs from natural compaction by its discontinuous and abrupt nature, and its timing in the wake of drought episodes.

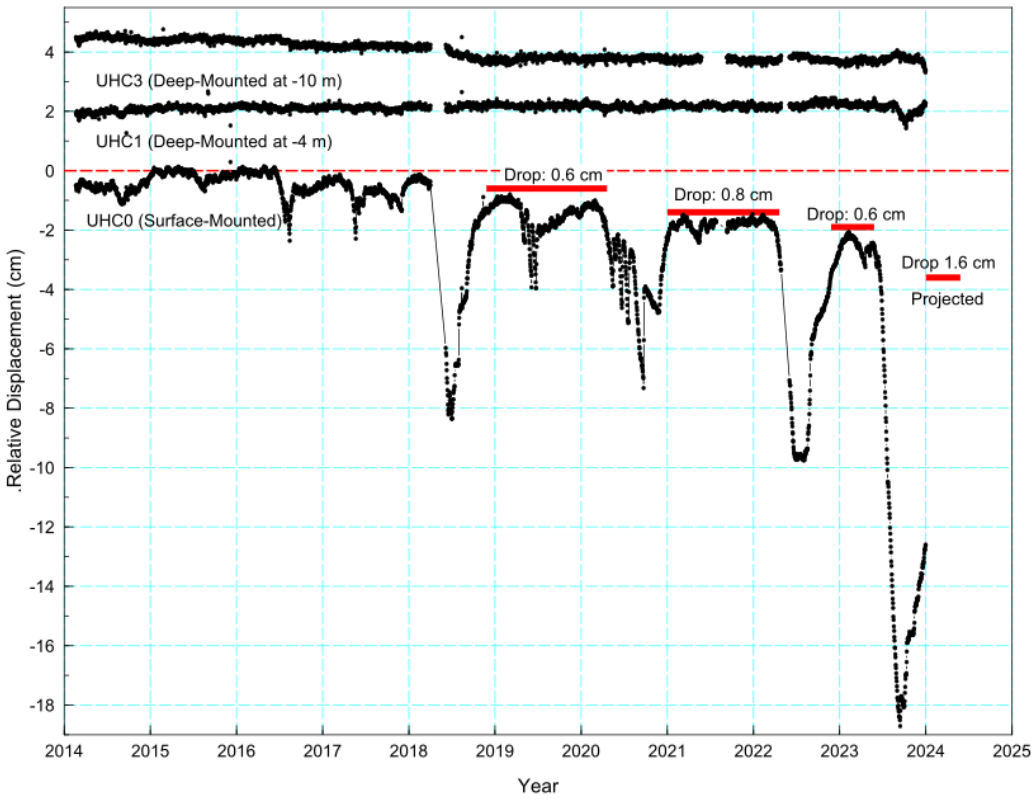


Figure 1.7: Cumulative soil shrinkage after several drought-rewetting cycles. Figure from Welch et al. [2024].

1.2.5 Damage to buildings

Clay shrinkage or swelling can damage buildings when differential displacement occurs beneath foundations. The damage is either referred to as clay shrinkage induced subsidence or clay swelling induced heave, based on which movement is at its source.

This situation is caused by differences in soil moisture content under the building, for example between a central part protected from evaporation, and the periphery dried out by vegetation roots. The resulting mechanical stress on the structure is absorbed as far as possible by its elastic deformation, up to the point of brittle fracture. Beyond this point, cracks of varying depth appear, ranging from simple cosmetic defects to structural failure [Page, 1998]. Figure 1.8 shows examples of building damage documented by the BRGM, and Figure 1.9 shows the simplified underlying mechanism.

Theoretically, the opening of cracks on a building provides information on the cause of the damage: a large opening at the top combined with a small opening at the foot of the building indicates subsidence, and vice versa for heave.

Detached houses are the type of building the most affected by this phenomenon due to their shallow foundations (less than 1 meter deep). In 2019, there were 19,230,227 detached houses in mainland France, 54.2% of which were located in areas affected by clay shrink-swell [MTES, 2021].

Figure 1.8: Two examples of building damage caused by clay shrinkage or swelling (photos of BRGM)

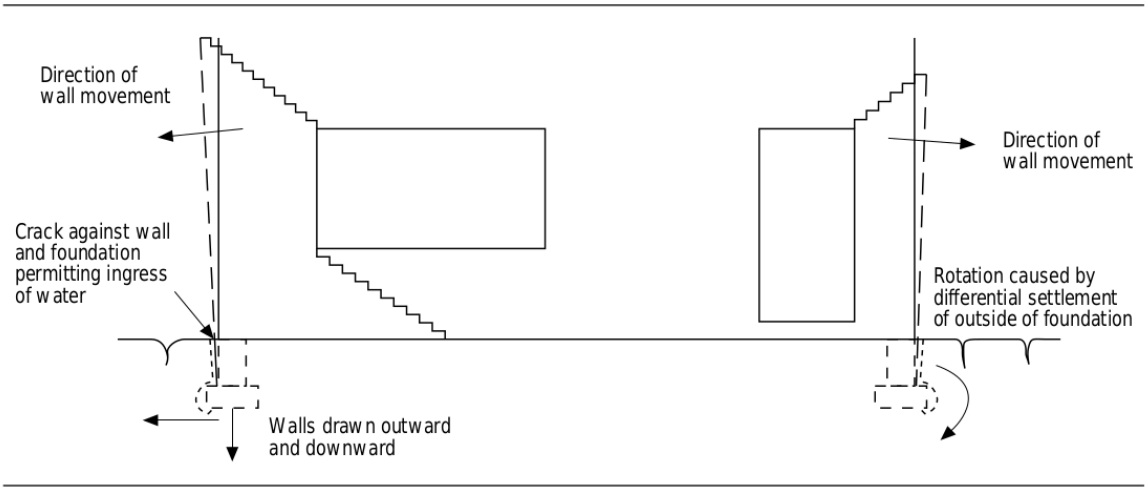


Figure 1.9: Building damage mechanism associated with shallow foundations in shrinkable clay. Figure from Page [1998].

The type of urban area also seems to play a role in the occurrence of damage: Corti et al. [2011] found that the urban centers are not affected by the phenomenon, while rural areas are twice as susceptible as discontinuous urban areas, based on insurance records. This could be due to the dominant type of building in these areas.

1.3 How to investigate clay shrinkage and swelling ?

In this chapter, we will present different approaches developed to study the phenomenon of clay shrinkage and swelling. First, some techniques for identifying expansive soils will be detailed in section 1.3.1. Then, we will comment on methods for tracking the phenomenon over time *in situ* in section 1.3.2. Thirdly, in section 1.3.3, we will present several modeling approaches that can be applied to help understanding clay shrinkage induced subsidence better.

1.3.1 Identification methods

Several methods, either done in the laboratory or on the field, have been developed to characterize the swelling potential of soils samples. These can directly characterize the mineralogy of the soil (section 1.3.1.1 and section 1.3.1.2), or indirectly quantify its mechanical (*geotechnical*) properties (section 1.3.1.3).

1.3.1.1 X-Ray Diffraction (XRD)

The XRD method consists in characterizing the crystalline structure of a mineral by analyzing the angle and intensity of the X-rays it diffracts. It can be used to determine the composition of a mixture of minerals within a small sample. Applied to the context of clay shrinkage and swelling, this technique enables to quantify the proportion of expansive clay minerals in a soil, to assess its overall expansive potential. However, it remains a punctual and costly method, difficult to apply on large spatial scales [Ducasse, 2019].

1.3.1.2 Towards remote identification

It has been shown by Grandjean et al. [2019] that the mineralogy of a soil can also be determined by spectroscopy, i.e. a technique based on the interaction of light with matter. Hyperspectral imaging is based on this principle. This technique has already been implemented on airborne sensors or onboard satellites (e.g. the ENMAP satellite). Although not yet operational, there are prospects for using remote sensing to characterize soils sensitive to clay shrink-swell at large spatial scales.

1.3.1.3 Geotechnical analyses

Several laboratory methods have been developed in geotechnical engineering to assess the potential of a soil to shrink and swell, resulting in specific metrics. Such investigations are frequently conducted in the context of construction works.

Methylene Blue Adsorption (MBA)

The MBA estimates the absorption surface of clay minerals. The test involves counting the amount of methylene blue dropped onto the surface of a soil sample until saturation occurs. Because swelling occurs due to water fixation on the surface of clay minerals, the MBA is positively correlated with the swelling potential [Çokça, 2002]. The experimental protocol is specified in the French NF EN 933-9 AFNOR standard (2022).

Plasticity Index (PI)

Atterberg limits are water content values that characterize phase transitions of fine-grained soils. When a soil goes from a liquid state to a plastic state, its water content falls below the Liquid Limit (LL). From a plastic state, it changes to a solid state below the Plastic Limit (PL). Finally, shrinkage stops below the Shrinkage Limit (SL). The Plasticity Index (PI), obtained by subtracting LL and PL, represents the extent of the plastic domain, and therefore indicates the ability of the soil to retain water. It is positively correlated with the expansive potential. The LL can be determined by two standardized experimental methods: either by counting the number of blows required to settle a soil sample in a cup (Casagrande

method), or by measuring the depth of penetration of a metal cone into the soil under its own weight. The PL is determined experimentally by shaping the soil in a fine cylinder and determining the water content at which it cracks. All these methods are described in detail by Öser [2020]. The experimental protocol is specified in the French NF EN ISO 17892-12 AFNOR standard (2018).

Swelling Index (C_s)

The swelling potential of the soil can also be estimated using the oedometer test. It consists of measuring the volume changes of a soil sample under mechanical stress, in drained conditions, and through load-discharge cycles. The resulting Swelling Index (C_s) quantifies the ability to return to the initial volume state after a first loading-unloading cycle. The test protocol is specified in the French XP P94-091 AFNOR standard (1995).

Shrink-swell modulus (E_w)

Soil shrinkage occurs in two stages. First, drying brings the soil particles closer together while the soil remains saturated. The volume shrinks significantly during this stage. Once all the soil particles are brought together, the interstitial water evaporates, without a significant change in volume. The slope of the first phase on a water content/deformation plane is called the Shrink-swell modulus (E_w) [Briaud et al., 2003]. It is determined experimentally by measuring the shrinkage of a standardized soil sample left to dry in the open air. The test protocol is specified in the French XP P94-060-2 AFNOR standard (1997).

1.3.2 Observation methods

Clay shrinkage and swelling can be tracked directly or indirectly, either by monitoring the vertical displacement of the soil or by monitoring soil moisture variations. In both cases, local and remote sensing techniques can be used. Each technique has its own advantages and disadvantages, which are discussed below. Soil moisture monitoring techniques are discussed in detail in section 1.3.2.1, and displacement tracking techniques in section 1.3.2.2.

1.3.2.1 Indirect observation through soil moisture

Soil moisture variations are the main driver of clay shrinkage and swelling. Therefore, tracking them allows to trace back to the phenomenon indirectly.

In situ with soil moisture probes

The water content of soils can be estimated *in situ* using probes that can rely on different technologies. Frequency Domain Reflectometry (FDR), Time Domain Reflectometry (TDR), or capacitive probes are most commonly used in instrumented sites. These three types of sensors estimate the volumetric water content of the soil through its electromagnetic property of permittivity, and all require prior calibration with respect to the soil texture. These probes have the advantage of measuring a reliable signal that is accurate over time, but remains very local.

There are operational soil moisture measurement networks based on such probes. For example, the Soil Moisture Observing System - Meteorological Automatic Network Integrated

Application called SMOSMANIA [Calvet et al., 2007], operated by Météo-France since 2007, comprises 20 stations distributed across southwestern France measuring soil moisture quasi-continuously down to a depth of 30 cm.

Figure 1.10 shows an example of how FDR probes are set up into the ground, in the Montaut site of the SMOSMANIA network. The pit is backfilled after the installation of the sensors.

Figure 1.10: FDR probes (Tethaprobes) installed at depths of 5 cm, 10 cm, 20 cm and 30 cm in the SMOSMANIA Montaut site. Picture from the supplement of Calvet et al. [2016].

Satellite data

Some satellites acquire data that can be processed into soil moisture products. This is the case of the SMOS (Soil Moisture and Ocean Salinity) satellite, and the Metop and Sentinel constellations, all three of which have been operated by the European Spatial Agency (ESA) since the 2010s, using sensors based on different technologies (2D interferometric radiometer, ASCAT scatterometer and synthetic aperture radar (SAR), respectively). The final products have spatial resolutions on the order of one to ten kilometers and are representative of the first few centimeters of soil. Measurements are taken at intervals of several days, depending on the satellite's orbit. These products have the advantage of large spatial coverage, but represent a part of the soil that is too superficial to describe the phenomenon under study.

1.3.2.2 Direct observation through displacement

The most direct way to monitor clay shrinkage and swelling is to track surface displacement. These observational methods only consider vertical displacements, although volume variations occur in all three dimensions.

Relative displacement with extensometers

Some experimental devices have been developed to track surface level variations locally. Extensometers, which belong to this family of instruments, measure over time the distance between a sensor at the surface and a supposedly stationary anchor point at depth. This is equivalent to tracking thickness variations, and captures shrinkage and swelling phenomena. The concept of extensometers is illustrated in Figure 1.11. This system is implemented by Fernandes et al. [2015], for instance.

The advantages of extensometers are their reliability (better than millimeter accuracy) and their fine temporal precision. However, they are complex to implement, as the installation requires drilling, and reflect very local information. It is important to note that extensometers measure the displacement of the surface relative to the anchor point, which might differ from absolute displacement.

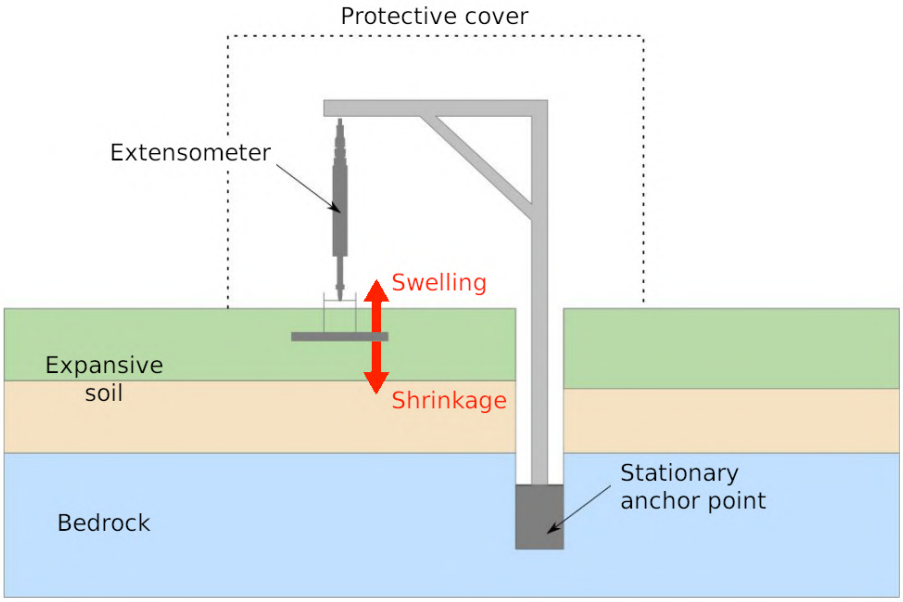


Figure 1.11: Concept of extensometer. Example of the instrumentation deployed on a BRGM site in the municipality of Chaingy, Loiret. Figure modified from Gourdier [2018].

Absolute displacement with GPS technology

There is a variant to the system cited above, which also uses an installation on-site, but returns absolute displacements. It is described in Welch et al. [2024], and involves embedding GPS sensors in the ground through boreholes, and tracking their position over time to infer displacement. As stated in section 1.2.4, this method results in different displacement values because it measures absolute rather than relative displacement. Despite this difference, both types of instruments are capable of detecting annual shrinkage and swelling episodes. Since the phenomenon only concerns the surface soils, as they are the only ones affected by meteorological variations, relative displacement seems more relevant than absolute displacement.

Satellite data

Recently, research effort was focused towards satellite-derived vertical displacement products. These include the European Ground Motion Service (EGMS) [Crosetto et al., 2021], developed under the European Union's Copernicus program. The EGMS, operational since 2018, provides vertical and horizontal displacement data over Europe, acquired by Sentinel satellites using InSAR (Interferometric Synthetic Aperture Radar) technology. A major advantage of these data is their high temporal (6 days) and spatial (1,000 points per square kilometer) resolution, the latter resulting from the use of widespread urban elements (buildings, poles...) as reflectors (Permanent Scatterers).

Such data can be used to analyze clay shrinkage and swelling cycles, as Burnol et al. [2021] has done and shown it captured trends. However, the interpretation of these data is not trivial. This arises from the fact that the vertical displacement caused by shrinkage and swelling is non-linear (seasonal periodicity), of small amplitude, and spatially heterogeneous,

as clayey soils are. A contrasting response is also expected from different types reflectors- tall buildings and poles have deep foundations, and are therefore not affected by this superficial phenomenon [Tzampoglou et al., 2022]. For all of these reasons, the signal corresponding to clay shrinkage and swelling may be difficult to separate from other phenomena such as subsidence induced by water pumping [Meisina et al., 2006].

Although remotely sensed vertical displacement data have interesting potential, they were not used in this work because of the difficulties cited above and the short temporal overlap with our data.

1.3.3 Modeling

Finally, clay shrinkage and swelling can be studied using models that help provide a deeper understanding of the phenomenon. We will describe three types of models, developed by different communities, that can be useful for this research topic. Land surface models, which are used in particular by the climate science community, will be introduced in section 1.3.3.1. Then we will mention geotechnical models used by the soil mechanics community in section 1.3.3.2. Finally, we will describe catastrophe models used by the insurance industry in section 1.3.3.3.

1.3.3.1 Land Surface Models

General information

A Land Surface Model (LSM) is used to understand and predict how the land interacts with the atmosphere. It relies on numerically implemented physical equations to reproduce processes such as how water moves in the soil and back into the atmosphere, or how solar energy is absorbed or reflected by the land surface. LSMs are typically defined at global to continental scales, with resolutions on the order of 1 to 10 kilometers. Such models are very complex because of the interdependence of several processes (in particular, the water, energy, and carbon cycles are coupled at the surface). LSMs are used for both research and operational purposes, such as weather forecasting, climate change, land use studies or water resource management. Compared to the previous generation of models, LSMs include the possibility to simulate (i) soil heat and water diffusion, (ii) photosynthesis and interactive vegetation, (iii) carbon fluxes and carbon storage.

Well-known examples of LSMs include Noah [Ek et al., 2003], used operationally in the United States by the National Centers for Environmental Predictions (NCEP), the Community Land Model (CLM) [Oleson et al., 2013], developed by the American National Center for Atmospheric Research (NCAR) for climate research, or the Joint UK Land Environment Simulator (JULES) [Best et al., 2011] developed by the UK Met Office for weather forecasting, climate modeling and hydrological studies.

Application to clay shrinkage and swelling

LSMs simulate soil moisture, the main driver of clay shrinkage and swelling, and can therefore be used to study this issue. They are more simplistic and less spatially precise than field data, but cover large areas, allowing generalization to regional or national scales.

The ISBA Land Surface Model

Météo-France has developed its own LSM, Interactions Sol-Biosphère-Atmosphère (ISBA) [Noilhan and Planton, 1989, Noilhan and Mahfouf, 1996], for both operational and research use. Its horizontal resolution is of 8 km, covering France with 8925 grid cells, but it can be applied worldwide if the atmospheric forcing is available. ISBA includes separate soil and vegetation schemes, with the possibility of differentiating land cover.

Soil moisture simulated by ISBA will be used in the thesis, therefore the model is described in detail in section 2.1. ISBA was extensively validated, in particular by comparing ISBA-derived soil water contents with field soil moisture data, either collected by Frequency Domain Reflectometry soil moisture probes [Decharme et al., 2011] or lysimeters [Sobaga et al., 2023].

1.3.3.2 Geotechnical models

Other types of models are specifically oriented at simulating soil volume changes due to shrinkage and swelling from the hydric state. The review by Adem and Vanapalli [2015] identifies three types of such models, which differ in their input variables and the governing equations. First, consolidation theory based methods are physical approaches that consider the soil as either an elastic or elastoplastic material, and use matric suction as a state variable. The other two types of methods consist of more empirical approaches, and are based on either water content or suction. However, these types of models require precise knowledge of the geotechnical properties of the soil, and can therefore only be implemented at a very local scale.

1.3.3.3 Catastrophe models

Finally, the purpose of the last type of model, catastrophe models, is to predict losses resulting from disasters. There are specific models for each peril, including models for subsidence. Such models are particularly useful to the insurance industry because they provide insight into the risk, which helps them plan their activities on the long term. These models are applied at large spatial scales, often over the insurer's portfolio, and are based on statistical approaches. In particular, historical loss data is used in the calibration phase [Langreney et al., 2023]. Many of these models are developed as part of the internal activities of insurance companies, in which case not much information is known about them. However, we have identified several published papers that present such models, which we will describe below.

Corti et al. [2009] developed a model using monthly precipitation and temperature as proxies for climate conditions to simulate annual subsidence losses for France between 1989 and 2002 by calibrating vulnerability curves. The method has been refined in Corti et al. [2011]. According to them, they manage to estimate damages correctly, but difficulties remain, as they struggle to characterize the 2003 drought event accurately and they identify disparities between French departments.

A similar but more advanced approach was used by Charpentier et al. [2022] to estimate the number of claims and total damages for France between 2001 and 2018. They used as climatic predictors standardized indicators of monthly precipitation, temperatures and soil moisture. Different models are tested, including regression and ensemble models. In conclusion, they

manage to assess the severity of a given year, but lack precision, which is explained by the heterogeneity of the data and the lack of information on local characteristics.

1.4 The impact of climate change

Increasing atmospheric CO_2 concentrations are already causing significant global climate changes [Lashof and Ahuja, 1990, Solomon et al., 2009, Hansen et al., 2013]. The phenomenon of clay shrinkage and swelling is expected to be affected as it is climate-driven. Various studies provide insight into this point, covering its different aspects. These studies can be based on the analysis of historical data, climate projections, or both.

1.4.1 Climate change scenarios

Projections are based on climate change scenarios, whose diversity enables to consider a range of possible futures. Over the years, there have been several generations of climate change scenarios, with different approaches.

Special Report on Emissions Scenarios (SRES) [Nakićenović and IPCC, 2000] based future climatic evolutions on different assumptions about economic growth, population growth, and technological development, without considering specific climate policies.

They have been replaced in climate impact studies by the concept of Representative Concentration Pathway (RCP) [Moss et al., 2010, Van Vuuren et al., 2011], which consists of trajectories for the evolution of global greenhouse gas concentrations up to 2100. Four RCP scenarios were initially defined: RCP 2.6, RCP 4.5, RCP 6 and RCP 8.5, named according to the level of radiative forcing. Radiative forcing, expressed in $W.m^{-2}$, represents changes in the Earth's energy balance. Higher values correspond to higher greenhouse gas concentrations, and thus higher temperatures and more pronounced climate change effects. RCPs have been widely used in climate modeling since the IPCC AR5 report [IPCC, 2014]. They have been supplemented by three additional pathways (RCP 1.9, RCP 3.4 and RCP 7) following the IPCC AR6 report [IPCC, 2023].

Shared Socio-economic Pathways (SSPs) have also been defined as an alternative to RCPs. They consist in deriving greenhouse gas emissions from projected socio-economic changes including climate policies.

1.4.2 Consequences for clay shrinkage and swelling

1.4.2.1 Meteorological conditions

In Europe, changes in precipitation and temperature patterns are expected. This can already be seen by comparing past and present data. Gudmundsson and Seneviratne [2015] found an increase in precipitation in northern Europe, while southern Europe is becoming drier. Manning et al. [2019] identified an increase in the frequency of hot and dry events in recent years compared to the second half of the 20^{th} century, as a result of rising temperatures.

The same diagnosis is made based on climate projections. For example, Spinoni et al. [2018] note an increase in the frequency and intensity of meteorological droughts over Europe under RCP 4.5 and especially under RCP 8.5, based on EURO-CORDEX simulations [Jacob et al.,

2014, 2020]. Although located in the transition zone between the wet north and the dry south, France is expected to be affected by the increase in the frequency of summer droughts.

1.4.2.2 Soil moisture trends

The impact of climate change on soil moisture trends has been specifically investigated by several works resorting to LSMs. The frequency of extreme soil moisture drought events is expected to increase in Europe [Samaniego et al., 2018], in line with the precipitation and temperature evolutions mentioned in the preceding section. Studies carried out at the scale of France agree on an increase in drying trends for soil moisture [Vidal et al., 2012, Dayon et al., 2018, Soubeyroux et al., 2023]. The largest impacts are expected for the most severe scenarios.

Merk et al. [2021] have already identified decreasing soil moisture levels at depth over the past two decades, based on lysimeter field measurements in Germany.

These soil moisture trends are expected to propagate to suctions. Assadollahi and Nowamooz [2020] projected future suction, under RCP 2.6, 4.5, and 8.5, for a study site in southwestern France, modeling local soil-vegetation-atmosphere interactions. They found that suction near trees would become more important in the long term for all scenarios.

1.4.2.3 Clay shrinkage hazard

While the studies mentioned above relate to the projected evolution of the climatic conditions that trigger clay shrinkage, others focus on the hazard in itself.

Several of these have been carried out for the UK. In particular, Harrison et al. [2012] first identify characteristics of triggering climatic conditions by comparing precipitation and temperature data from several weather stations with an insurance claims dataset, then apply their identification method to projected conditions under different emission scenarios and finally merge the results with a soil susceptibility map. On the other hand, Pritchard et al. [2015] assess future drought conditions by subtracting precipitation and evapotranspiration outputs from a stochastic weather generator under SRES A1B, equivalent to RCP 6 and combine the resulting variable with a soil susceptibility map. Both studies agree on a worsening of the hazard in southern England.

To our knowledge, no such study has ever been done for France.

1.4.2.4 Subsidence damage

A final type of work aims to quantify future damage to buildings due to subsidence under climate change. Technical reports are more common in this category than scientific articles.

Many of such reports focusing on France have been published recently due to the recurring droughts of the past years. These include Gourdier and Plat [2018], France Assureurs [2021], Covéa and RiskWeatherTech [2022] and most recently CCR [2023b]. To develop this last example, the study involved generating 400 years of climate conditions under RCP 4.5 and RCP 8.5 and inputting them into a catastrophe model to estimate subsidence damage under a future climate. This type of study requires making assumptions about changes in exposed

buildings and in the institutional framework, and on inflation. As a result, they quantify an increase in average annual losses of 59% under RCP 4.5 and 190% under RCP 8.5, by 2050.

1.5 Institutional context

France has a specific framework for claim compensation. Developing the way it works is useful since the work exploits insurance claim data.

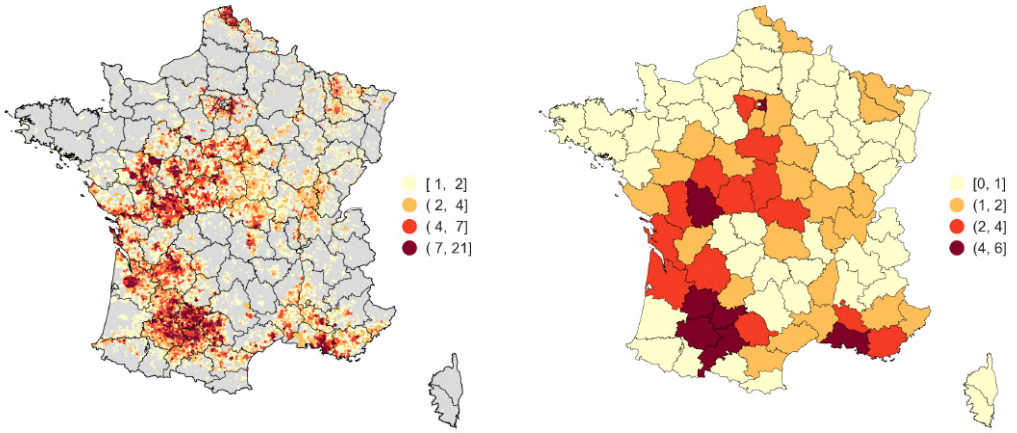
1.5.1 The NatCat regime

In France, clay shrinkage-induced subsidence is compensated by insurers through a scheme called the *Régime d'indemnisation des catastrophes naturelles*, or NatCat regime. Created in 1982, it is based on a principle of national solidarity: a part of the compensation that victims of natural disasters receive indirectly originates from a fund financed by a share of the car and home insurance premiums of all policyholders. This fund is managed by the Caisse Centrale de Réassurance (CCR). Perils covered by this regime include flooding, phenomena related to the action of the sea, earthquakes, avalanches, and land movements including clay shrink-swell (referred to as drought), since 1989.

Compensation is subject to the issuance of a peril-specific decree, that recognizes the state of natural disaster for a municipality for a specified period of time. The procedure is that the mayor of the municipality submits a request, which is examined by a special government committee. Requests and related decisions are available, grouped by decree, on the CCR website (https://catastrophes-naturelles.ccr.fr/les-arretes). If the application is accepted and results in a decree, the residents of the municipality can apply to their insurer for compensation, which may result in a claim.

In the specific case of clay shrinkage-induced subsidence, the decision of the committee rests on two technical criteria: a geotechnical criterion and a meteorological criterion. The geotechnical criterion, implemented in 1989, requires the municipality to have at least 3% of its surface mapped as expansive clay, based on the Bureau de Recherches Géologiques et Minières (BRGM) maps. The meteorological criterion, developed by Météo-France, aims to characterize exceptional soil moisture drought events. Its definition changed successively over time as detailed in section 1.5.1.1.

Figure 1.12 plots statistics on accepted requests relative to clay shrinkage-induced subsidence between 1989 and 2022. The frequency of recognition is especially high in the Center-West, South-West and South-East of the country. The detail of accepted and refused requests for each year is available in Figure A.2.



(a) Number of accepted requests by municipality

(b) Average number of accepted requests per municipality by department

Figure 1.12: Statistics on accepted NatCat requests relative to drought between 1989 and 2022, (a) by municipality and (b) by department. The data is from https://catastrophes-naturelles.ccr.fr/les-arretes.

1.5.1.1 Evolution of the meteorological criterion

The meteorological criterion aiming to identify exceptional drought events has evolved successively over time. The first criterion, introduced in 2000 was a "winter" criterion that estimated the water deficit in the period preceding the drought event. It was revised, in 2003, 2011, and 2018 to include summer, spring and fall water deficit periods, respectively. Initially, these criteria were based on a "double-reservoir water balance", which required data from Météo-France instrumented stations. The main drawback was the insufficient number of stations. For this reason, it was replaced in 2009 by the "SWI uniforme" modeled by the ISBA LSM, detailed in section 2.1.4. [De Solère, 2007, Salagnac, 2007, Vincent et al., 2009].

The current method for calculating the meteorological criterion is detailed in the legal document République Française [2019], and the upcoming changes are mentioned in République Française [2024].

1.5.2 Prevention techniques

In this context, risk prevention efforts are made to limit the damage caused by clay shrinkage and swelling. The *Géorisques* website [MTES, 2024] informs the general public about the phenomenon and makes recommendations.

Operational techniques to prevent clay shrinkage induced subsidence damage consist of either neutralizing the triggering factor or strengthening the building. The first type of technique involves removing or pruning trees around houses, or protecting the outline of the house with paving to limit evaporation. The second type of technique consists, for example, of stiffening the structure with chain links.

Chapter 2

Tools and data

Contents

This chapter presents the main tools and datasets used in this work: the ISBA Land Surface Model (section 2.1), the ATHENA insurance databases (section 2.2), clay shrinkage hazard maps (section 2.3), and measurements from an instrumented site (section 2.4). We will also give details on the programming environment used to obtain the results in section 2.5, in particular the Python modules.

ISBA simulations will be used to explore past and future drought conditions, insurance data will be used as observations to guide experiments, hazard maps will provide complementary information to understand the phenomenon, and data from the instrumented site will be used for a case study of the accuracy of ISBA at the local scale.

2.1 ISBA Land Surface Model

Interactions Sol-Biosphère-Atmosphère (ISBA) [Noilhan and Planton, 1989, Noilhan and Mahfouf, 1996] is the Land Surface Model (LSM) developed by Météo-France. Such tools estimate water and energy budgets at the surface by modeling the physical processes occurring between the atmosphere, soil and vegetation (e.g. precipitation interception, evapotranspiration, runoff, and infiltration). Figure 2.1 summarizes the concept of ISBA, and specifies its different components, that will be described below in section 2.1.1 and section 2.1.2. ISBA is integrated into the SURFEX modeling platform [Masson et al., 2013] where it simulates the water and energy fluxes of natural surfaces (nature tile).

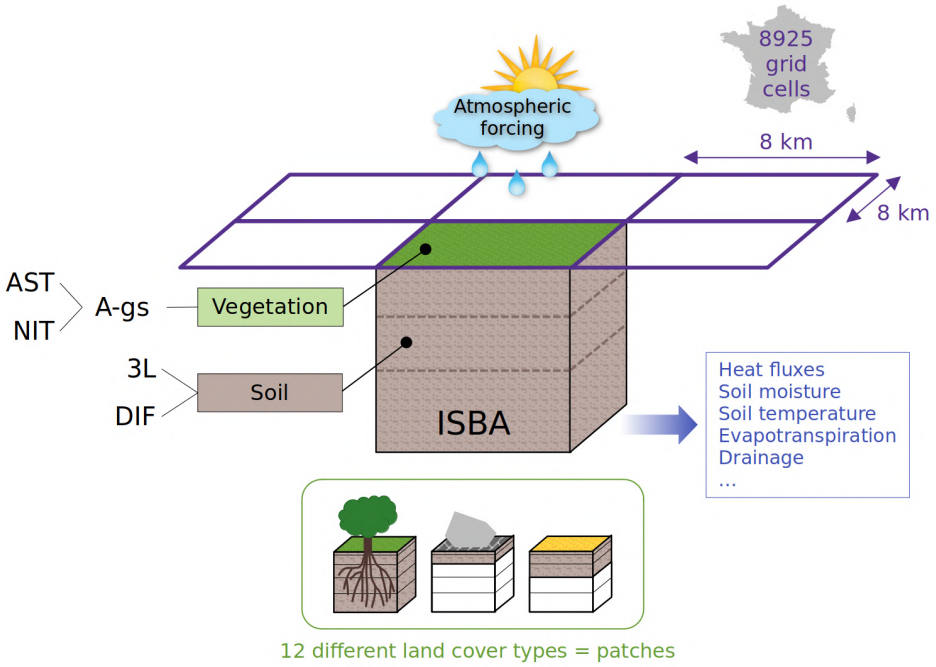


Figure 2.1: Components of the ISBA Land Surface Model. The acronyms refer to model options that will be described below.

ISBA covers mainland France on a 8 km resolution grid, totaling 8925 grid cells. Atmospheric forcing, such as radiation, air temperature, wind speed or air humidity is required to drive the model. Outputs, that include fluxes and state variables of the system (e.g. sensible and latent heat fluxes, soil liquid water content...) will be described in section 2.1.3.

Current operational uses of ISBA by Météo-France will be detailed in section 2.1.4.

2.1.1 Soil

2.1.1.1 Physical representations

A summary of the different ISBA soil schemes useful to this work and their associated references is presented in Table 2.1.

Table 2.1: ISBA soil schemes

Abbreviation	Soil scheme	References	Number of layers
ISBA-3L	Force-Restore	Noilhan and Planton [1989] Boone et al. [1999]	3
ISBA-DIF	Diffusion	Boone et al. [2000] Decharme et al. [2011]	maximum 14

ISBA-3L is based on the force-restore approach [Bhumralkar, 1975, Blackadar, 1976], which theorizes that the evolution of the surface temperature is driven (*force*) by energy exchanges with the atmosphere (i.e., net radiation, latent and sensible heat fluxes), but equilibrates at slower time scales with a deeper soil temperature (*restore*). This scheme is applied in ISBA by dividing the soil into two conceptual layers, a surface layer and a deeper layer. The representation is extended to water transfers [Deardorff, 1978]: in this case, the *restore* term depends on the contrast between surface and root zone soil moisture with a correction for gravitational effects. Three layers are considered, surface, root zone and below the root zone for a better representation of dry periods.

ISBA-DIF adopts on the other hand a diffusive approach, which requires a larger number of layers. It rests on both a finite-element approach of the mixed form of the Richards equation for water transfers in unsaturated soils [Richards, 1931], and a 1-D explicit Fourier law resolution for thermal processes [Fourier, 1822, Le Roy, 1897]. The mixed-form of the Richards equation, explicited in Equation 2.1, is derived from Darcy's law [Darcy, 1856], considering the law of mass conservation. This equation is commonly used to model water transfer in the vadose zone.

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\theta)\frac{\partial \psi}{\partial z}\right] + \frac{\partial K}{\partial z} \tag{2.1}$$

The default number of layers of ISBA-DIF is 14 with depths of 0.01, 0.04, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 1.5, 2, 3, 5, 8, and 12 meters. The number of layers is adjustable by the user. The thinness of the most superficial layers is designed to minimize errors associated to the important surface gradients due to the diurnal cycle. In practice, the layers considered for

hydrological simulation are those intersected by roots, which can extend to a depth of 2 meters in the case of trees.

2.1.1.2 Parameters

Physical parameters required for modeling, such as the matric potential at saturation ψ_{sat} or the saturated hydraulic conductivity K_{sat} for soil moisture representation in ISBA-DIF, are derived from the soil texture using empirical pedotransfer functions from Clapp and Hornberger [1978] by Noilhan and Lacarrère [1995].

Soil texture refers to particle size, the categories clay, silt and sand being defined by diameters inferior to 2 μm, 50 μm and 2 mm, respectively. The texture information is provided by the Harmonized World Soil Database (HWSD) [Nachtergaele et al., 2012], at the kilometer resolution and constant in depth. It is averaged to match the ISBA 8-km grid, as plotted on Figure 2.2.

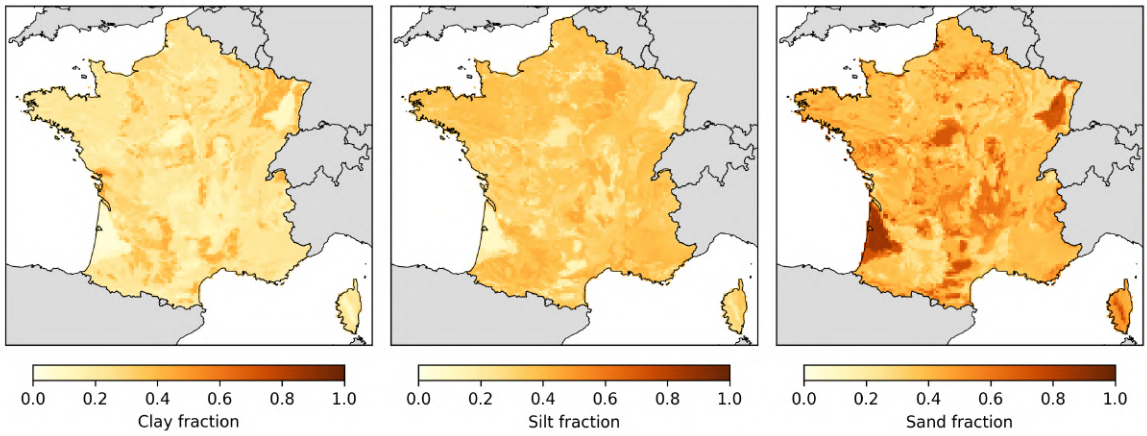


Figure 2.2: ISBA texture parameters

2.1.2 Vegetation

The role of vegetation in LSMs is crucial as it is the main contributor to total land evapotranspiration on the global scale [Seneviratne et al., 2010]. Several schemes were implemented in ISBA to replicate plant influence on surface exchanges.

2.1.2.1 Model options

ISBA-A-gs [Calvet et al., 1998] consists of an implementation of a CO_2-dependent plant stomatal conductance scheme in ISBA. It is based on a simple growth submodel. Calvet et al. [2012] have shown that ISBA-A-gs reproduces the inter-annual variability of above-ground biomass for grassland very well by comparison with agricultural statistics.

With the **AST option**, the Leaf Area Index (LAI), which quantifies the leaf surface available for exchanges with the atmosphere, is prescribed, for instance from satellite data. AST also implements a more detailed representation of hydric stress on photosynthesis, with drought-tolerant of drought-avoidant strategies, for both herbaceous plants [Calvet, 2000] and trees [Calvet et al., 2004].

With the **NIT option**, phenology is driven by the photosynthesis simulated in AST: the LAI is not prescribed anymore but simulated from leaf biomass considering the leaf nitrogen concentration [Calvet and Soussana, 2001]. This means that with this scheme, plants are sensitive to environmental conditions, for example they wither consequently to a drought. The output LAI has been validated at global [Gibelin et al., 2006], continental [Szczypta et al., 2014] and regional [Brut et al., 2009] scales.

2.1.2.2 Patches

To consider surface heterogeneity within an 8 km × 8 km grid cell, ISBA runs separately for 12 land cover types called *patches*. The water budget is simulated separately for each patch. A description of the 12 patches is available in Table 2.2, and their spatial repartition in Figure 2.3. The geographic distribution of patches and vegetation parameters are derived from the ECOCLIMAP-II database at a kilometer resolution [Faroux et al., 2013].

Of the 12 patches, only 8 are widespread over the country: bare soil (patch 1), rocks (patch 2), broadleaf trees (patch 4), coniferous trees (patch 5), C_3 crops (patch 7), C_4 crops (patch 8), C_3 grassland (patch 10) and parks and gardens (patch 11).

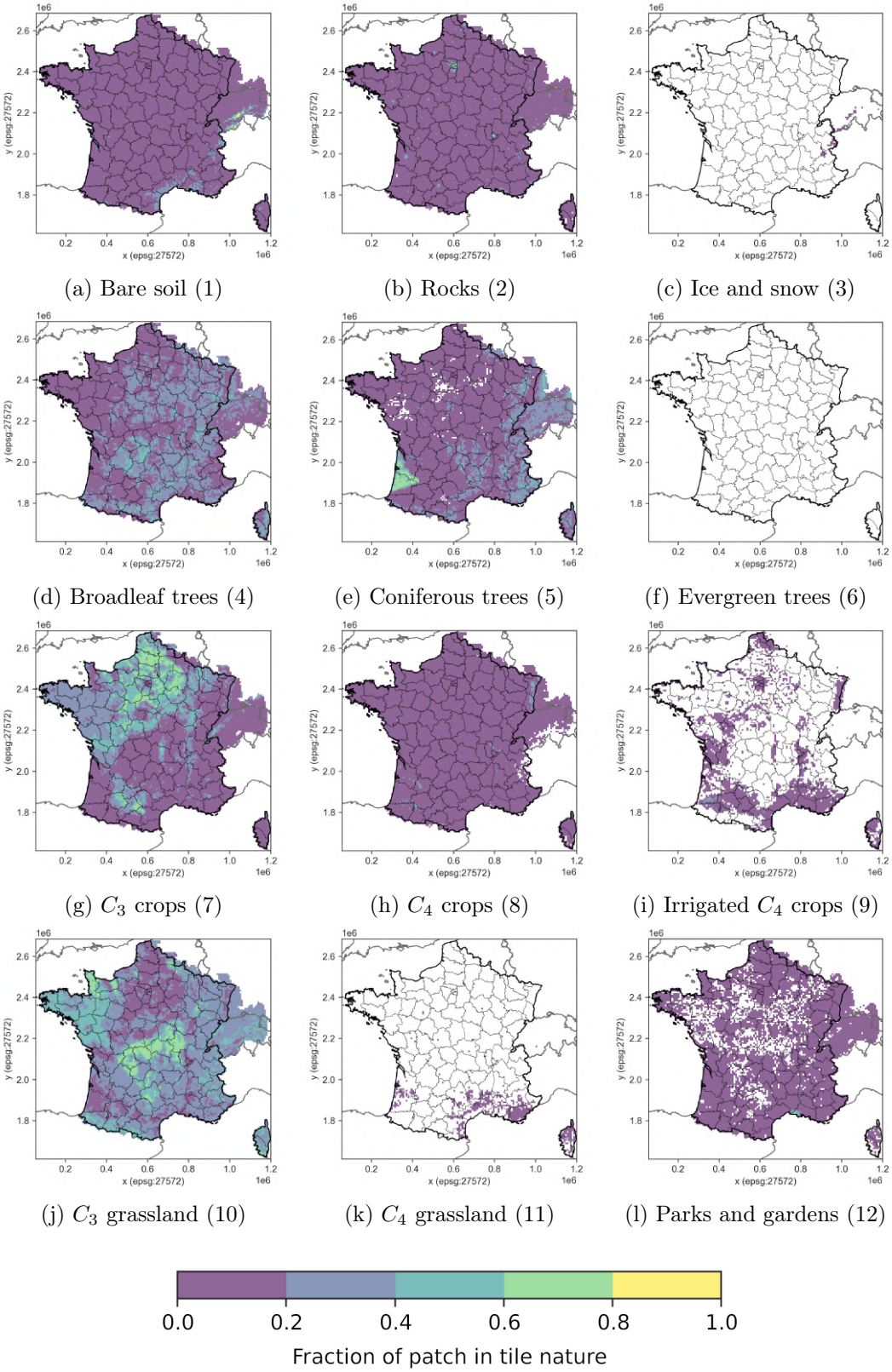
Patches have a major impact on soil hydrology because they allow modeling the absence or presence of different types of vegetation. Some patches model trees (patches 4, 5), while others model herbaceous vegetation (patches 7, 8, 9, 10 and 11). Within trees, deciduous broadleaf (patch 4) and coniferous trees (patch 5) develop different water management strategies. For example, conifers limit transpiration through their needle leaves, which have a small surface area in contact with the atmosphere. Among the herbaceous plants, C_3 (patch 7 and 10) and C_4 plants (patch 8) are distinguished, the latter presenting a variation in the photosynthetic process, more adapted to tropical areas.

In ISBA-DIF, the vertical extent of patches depends on the rooting system of the corresponding vegetation: only tree types (broadleaf and coniferous) reach two meters deep. The separate runs are aggregated by averaging the output variables weighting each patch by its respective fraction in the grid cell.

Table 2.2: ISBA patches and properties

Number	Patch	Comment	Max depth in DIF
1	Bare soil		1 m
2	Rocks	Representation of urban areas	20 cm
3	Permanent ice and snow	Only in the Alpine area	20 cm
4	Deciduous broadleaf trees	*Oak, elm...*	2 m
5	Coniferous trees	*Pine, spruce...*	2 m
6	Evergreen broadleaf trees	Tropical vegetation	
7	C_3[1] crops	*Wheat, rapeseed...*	1 m
8	C_4[1] crops	*Maize...*	1 m
9	Irrigated (C_4) crops	Mainly in the South-West	1 m
10	C_3 grassland	*Meadows, pastures...*	1.5 m
11	C_4 grassland	Tropical vegetation	
12	wetlands, parks, gardens...		2 m

[1] C_3 and C_4 correspond to different photosynthetic biochemical pathways.



(a) Bare soil (1) (b) Rocks (2) (c) Ice and snow (3)

(d) Broadleaf trees (4) (e) Coniferous trees (5) (f) Evergreen trees (6)

(g) C_3 crops (7) (h) C_4 crops (8) (i) Irrigated C_4 crops (9)

(j) C_3 grassland (10) (k) C_4 grassland (11) (l) Parks and gardens (12)

Figure 2.3: Distribution of patches in ISBA grid cells

2.1.3 ISBA Outputs

During the modeling process, the state of the atmosphere-soil-vegetation system is represented by a set of descriptive variables that are ultimately returned as output. They consist of state variables and fluxes relative to:

- the **surface energy budget**: radiation, sensible and latent heat fluxes, evaporation and transpiration fluxes, surface albedo...
- the **vegetation activity**: Leaf Area Index (LAI), biomass, gross primary production, respiration...
- the **different surface water reservoirs**: soil liquid and solid water content, canopy interception reservoir, snow fraction covering the ground, and **outgoing water flows** such as drainage and runoff.

2.1.3.1 Volumetric water content to soil wetness index

Since this thesis focuses on clay shrinkage-induced subsidence, the most useful output is the Soil Liquid Water Content (WG). It is expressed in volumetric terms (volume of water per volume of soil), and is calculated separately for the different layers in DIF. Because grain size affects both porosity and capillary forces, WG is highly correlated with soil texture. This can be seen on Figure 2.4 (1) and (2).

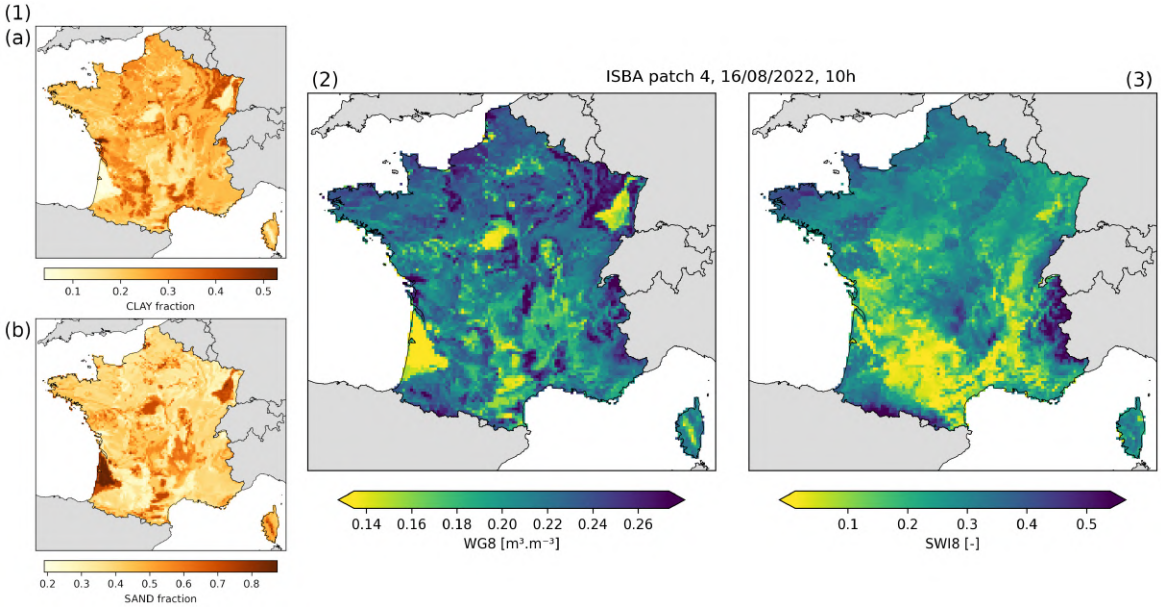


Figure 2.4: Normalization of (2) Soil Liquid Water Content (WG) relative to (1) texture through the (3) Soil Wetness Index (SWI). Example of the 16/08/2022 at 10h, for the deciduous broadleaf tree patch.

This effect is eliminated in ISBA by calculating a SWI (Figure 2.4 (3)). It consists of soil moisture normalized between wilting point (W_{wilt}) and field capacity (W_{fc}) as in Equation 2.2.

$$SWI = \frac{W - W_{wilt}}{W_{fc} - W_{wilt}} \tag{2.2}$$

The wilting point and field capacity are the soil moisture values associated to the matric potentials of -3 m and -150 m, that are associated to the states of unability of plant roots to absorb ground moisture and maximal retention after gravitational flow, respectively. This is detailed in section 1.2.2.2.

In ISBA-DIF, W_{fc} and W_{wilt} are derived from the Campbell [1974] equation of the SWCC. Since the SWCC changes with soil texture, multivariate regression was applied to parameters obtained for different soil classes, leading to the equations Equation 2.3 and Equation 2.4, currently in use for obtaining ISBA-DIF parameters. The variables X_{clay} and X_{sand} are the fractions of clay and sand.

$$W_{wilt} = 37.1342 \times 0.01 \times X_{clay}^{0.5} \tag{2.3}$$

$$\begin{aligned} W_{fc} = {} & 0.2298915119 - (0.4062575773 * X_{sand}) + (0.0874218705 * X_{clay}) + \\ & 0.2942558675 * X_{sand}^{1/3} + 0.0413771051 * X_{clay}^{1/3} \end{aligned} \tag{2.4}$$

2.1.4 Operational applications of ISBA

2.1.4.1 The SIM2 hydrometeorological chain

ISBA is used operationally by Météo-France for the management of water resources within the SAFRAN-ISBA-MODCOU (SIM) chain. It consists in its coupling with the SAFRAN meteorological analysis [Quintana-Seguí et al., 2008, Vidal et al., 2010] and with the MODCOU hydrogeological model, through the SURFEX modeling plateform. The latest version of SIM, called SIM2, is based on **SURFEX version 8.0** [Le Moigne et al., 2020], with the ISBA options **DIF** and **A-gs AST** for soil and vegetation, aggregated by patch.

2.1.4.2 SWI Uniforme

Another operational product based on ISBA exists for the purpose of providing expertise to the committee that decides whether French towns should be recognized in a state of natural disaster following a drought event, within the framework of the NatCat regime section 1.5.1. This tool, called SIM Uniforme, is based on a version of ISBA in **SURFEX version 8.0** with the **3L** soil scheme, the **A-gs AST** vegetation option and a **uniform** configuration: C_3 grassland patch everywhere and a soil texture consisting in 58% clay and 12% sand. The purpose of the uniform configuration is to provide an objective assessment of soil moisture throughout France to ensure equal treatment for all claimants. Since the only useful output of this simulation is the SWI, averaged over the entire depth, it is called SWI uniforme. It is available at: https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=301&id_rubrique=40 (last access 4 June 2024). The NatCat decision process rests on a SWI uniforme-based criterion, whose evolutions are detailed in section 1.5.1.1.

2.1.5 Uses of ISBA simulations

ISBA will later be used as basis of a drought index fitted to clay shrinkage induced subsidence. We developed a methodology relying on insurance claim data. The AST and NIT vegetation schemes will first be compared, to assess if modeling an interactive vegetation is useful in this study case. Then, the investigation will be done by patch, in contrast to the aggregated version. Thirdly, we will compare results to the operational ISBA versions. The best ISBA version will be used for computing the drought index over France over a historical period and in the future.

2.2 ATHENA insurance databases

The insurance data used throughout this work as a proxy for clay shrinkage-induced subsidence comes from the CCR ATHENA insurance databases.

Through its reinsurance activity, Caisse Centrale de Réassurance (CCR) has insight into the portfolios of its clients, called *cedants*, which represent 98% of insurance companies situated in France. ATHENA is the name of the CCR database that stores this information in two tables: a first table for insurance policies (contracts) and a second table for insurance claims. The two tables are linked by a common identification field. The first claims go back to 1989, and 2020 is the latest year for which we have a sufficiently complete dataset (it takes a long time to track claims given the way the system works). Therefore, we limited ourselves to the period 1989-2020 to analyze both claim and policy data.

The policy database contains annual information on insured properties and some of their characteristics. For example, some of these characteristics are the type of building (house, apartment), its insured value, or its use (professional or private). We count 176,419,070 entries considering only contracts taken out by house owners, with a spatial precision at the address, between 1989 and 2020. This number is much higher than the 19,230,227 houses counted for France by INSEE in 2017 [MTES, 2021]. This is because instances corresponding to insurance policies are duplicated by year in the ATHENA database.

On the other hand, the claims database stores claims compensated by insurers, for all the perils covered by the NatCat regime (section 1.5.1). It has 367,342 entries corresponding to claims resulting from clay shrinkage-induced subsidence. This number drops to 120,387 if only claims with a spatial precision at the address are considered.

Removing duplicates by year, we number 15,581,225 policies at unique locations in the database, between 2000 and 2022. They are distributed among the French departments as in Figure 2.5. Policies are distributed among departments in the same way than houses are (see Figure A.3), based on the 2017 INSEE census cited above. This shows that the policy database is overall representative of the French housing market. On the other hand, claims have a different spatial distribution: most of the damage occurred in the southern half of the country (departments of Haute-Garonne, Gers, Tarn, Bouches-du-Rhône, Gironde, Dordogne and Puy-de-Dôme).

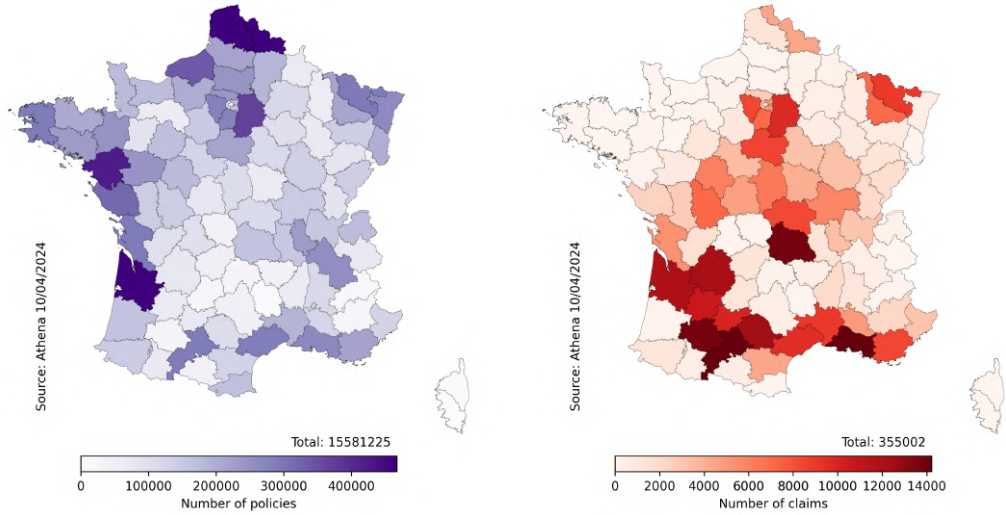


Figure 2.5: Number of policies (houses, address precision) at unique locations (duplicates by year were deleted) and all clay shrinkage induced subsidence claims, by department, between 2000 and 2020. The number of claims does not match the one cited above because of missing department identifiers.

2.2.1 Database completeness

Both policy and claim databases are subject to uncertainty because insurers sometimes submit incomplete information. Data may be underrepresented due to missing policies or claims, or overrepresented due to duplicates. To quantify this uncertainty, a *completeness coefficient* is calculated internally at CCR for each year and cedant. It is obtained by comparing individual policy and claim records with the total insured value of damages provided by cedant for each year. A score of 1 indicate a perfect match. If the completeness coefficient is inferior to 1, the data is underrepresented, and conversely, if superior to 1, the data overrepresented.

Applying a condition on completeness filters the data considerably: with a threshold of 10% (coefficient between 0.9 and 1.1), common in the insurance industry, only 14.7% of the 176,419,070 policies and 20.6% of the 120,387 claims with a precision at the address are kept. The distribution of policies and claims by year, after filtering for completeness (10% threshold for both policies and claims) and conserving only address data, is shown on Figure 2.6.

In addition to the difference in the order of magnitude of the number of policies and claims, we note that both are increasing over time. There is no data for 2017 (gray bar on Figure 2.6), which is explained by the lack of completeness of the data. The years 2016, 2019 and 2011 have a very high number of policies. In particular, 2018 stands out for its high number of claims relative to policies.

2.2.2 Uses of the insurance data

Data issued from the ATHENA database will be used twice in this work, with different processings. First, the yearly number of claims, aggregated by town without making any filtering on exhaustivity, will be used as a proxy for subsidence when developing the ISBA-

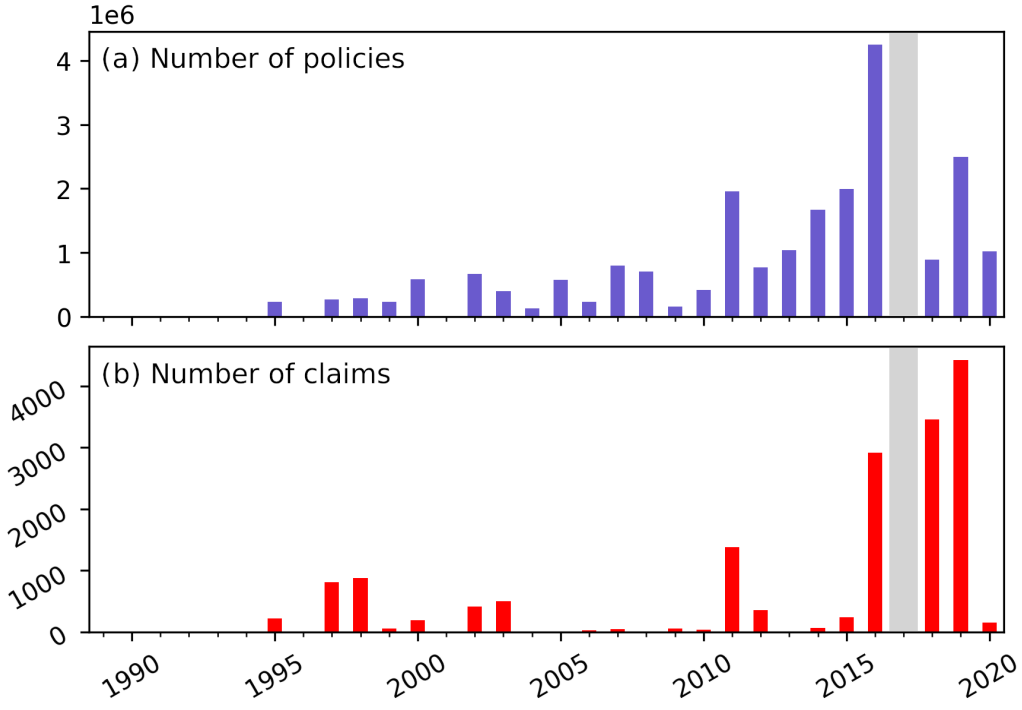


Figure 2.6: Number of policies and claims by year after filtering on completeness (10% threshold), spatial precision (address) and dropping duplicates

based drought index. Then, an extraction of policies merged with an extraction of claims, both filtered on exhaustivity, will form the body of a dataset that will be used to estimate to predictive capacity of the drought index combined with susceptibility maps in the last chapter.

2.3 Clay-shrinkage hazard maps

In its role as National Geological Survey, the BRGM has published several maps relative to expansive soils in France. They are useful to this work, as they describe the predisposing factor for clay shrinkage.

- **Susceptibility**, i.e. the potential of the soil to shrink and swell, is determined from three geological criteria: **lithology**, **geotechnical** properties and **mineralogy**.
- Combining it with insurance damage data enabled to obtain the **exposure** to clay shrink-swell map currently in operational use.

Clayey geological units were first identified from geological maps at 1:50,000 scale (approximate area of 20 km × 30 km and 100 m precision), and their expansive potential was assessed based on the three geological criteria cited above. The analysis was made department by department. All 96 metropolitan French departments were covered by the BRGM mapping program between 1998 and 2011. Reports for each department are listed in section A.4.

Each of the geological criteria is in classes expressing increasing susceptibility, that can take

3 to 4 different values. The spatial extents of classes, expressed as percentages of the area of France, are plotted in Figure 2.7. According to these maps, 61% of the area of France is covered by clayey geological units. Missing data for one of the criteria is compensated by the others when determining susceptibility or exposure. The susceptibility classes are scaled down in comparison to the three separate criteria because the calculation involves rounding down the sum of the latter.

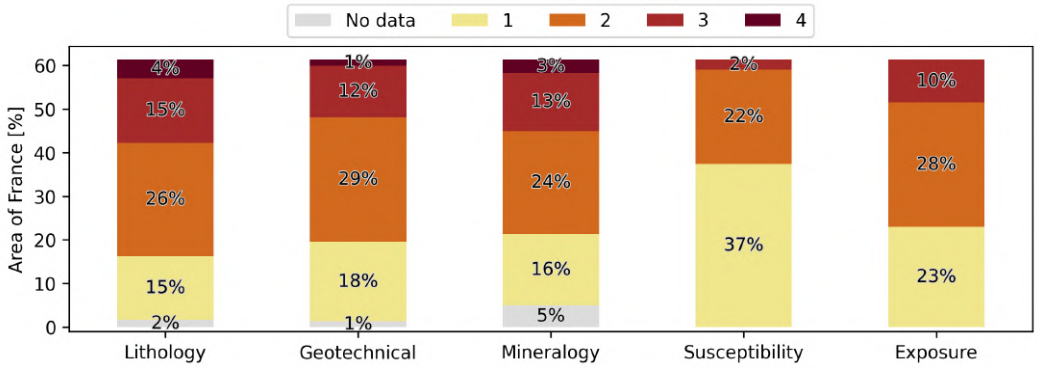


Figure 2.7: Spatial extent of classes for each expansive soil characteristic

2.3.1 Contributions to susceptibility

Once clayey units were identified from geological maps, three criteria were identified as relevant for assessing their potential to shrink and swell: how abundant the clayey zones are, and their structure (*lithology*), the unit's *geotechnical* properties, and its clay mineral content (*mineralogy*).

Although these three criteria were estimated independently, they are correlated. A chi-square test of independence allows us to reject the hypothesis of independent scores with a very high level of confidence (p-value « 0.001).

2.3.1.1 Lithology class

The lithology criterion, relative to the thickness and continuity of clayey geological units, is mapped over France in Figure 2.8, and its classes are described in Table 2.3. The characterization is based on both the geological maps at 1:50,000 scale and expert field knowledge.

Table 2.3: Description of lithology classes

Class	Description of unit
1	Non-clayey, with clay lenses
2	Non-dominant clay term
3	Mostly clayey, with non-clay layers or thickness $<$ 3m
4	Mainly clay or marl, continuous, thickness $>$ 3m

Data is missing for two departments: Gers (32) and Deux-Sèvres (79). They were among the first to be studied (reports were respectively published in 2001 and 1998), which could point

to the fact that the methodology was not yet perfected.

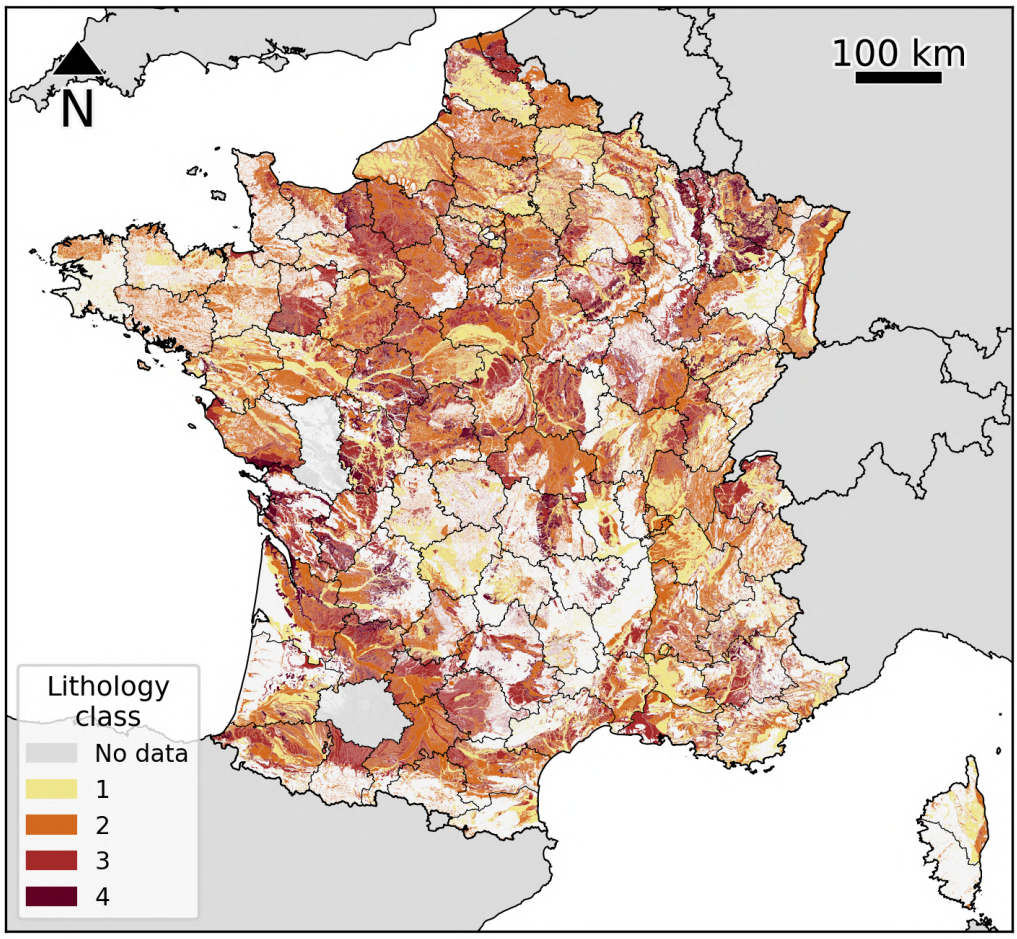


Figure 2.8: Lithology criterion over France. Classes are defined in Table 2.3.

2.3.1.2 Geotechnical class

The geotechnical criterion, describing the soil's mechanical behavior, is mapped over France in Figure 2.9. The characterization is based on results of geotechnical tests, either the Plasticity Index (PI) from the Atterberg limits test, the Methylene Blue Adsorption (MBA) coefficient, the Shrink-swell modulus (E_w) factor or the Swelling Index (C_s) from oedometric testing, obtained in the framework of construction projects. The correspondence between classes and range of values is specified in Table 2.4.

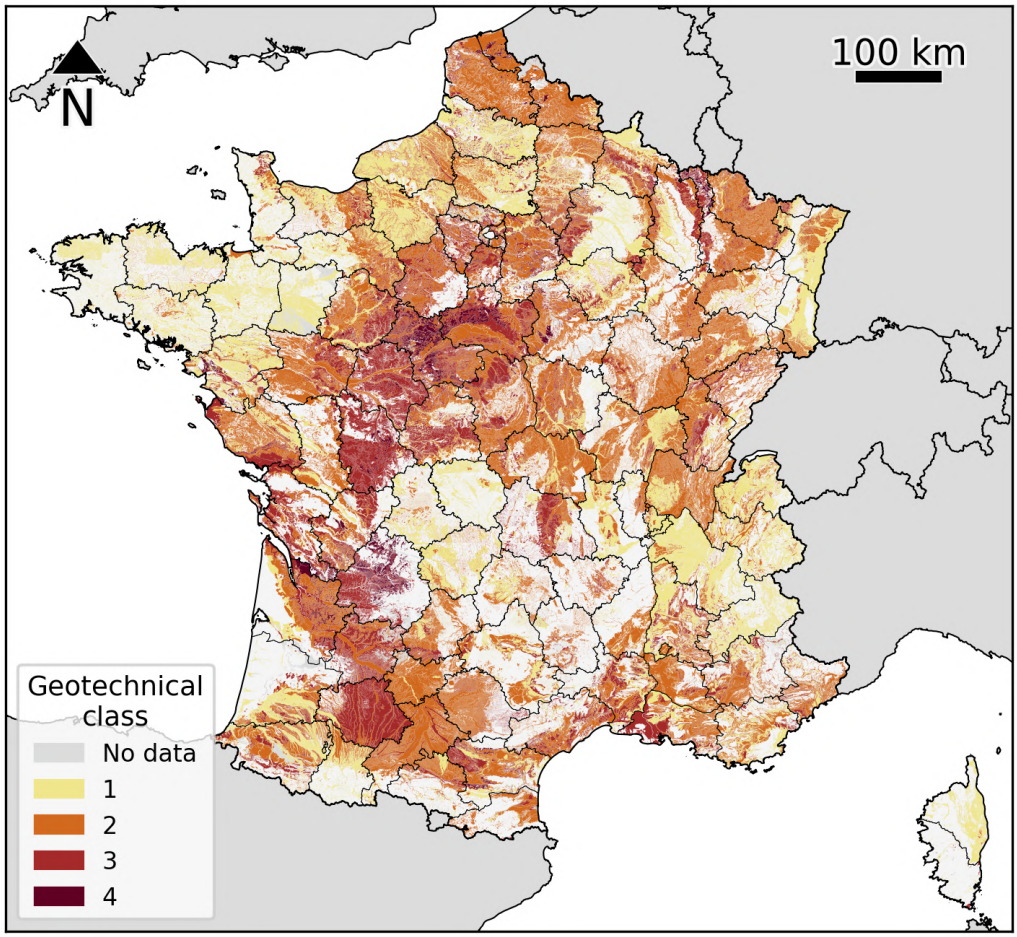


Figure 2.9: Geotechnical criterion over France. Classes are defined in Table 2.4.

Table 2.4: Description of geotechnical classes

Class	PI	MBA	E_w [%]	C_s
1	<15	<2.5	<40	<0.025
2	15-30	2.5-6	40-65	0.025-0.035
3	30-40	6-8	65-75	0.035-0.055
4	>40	>8	>75	>0.055

2.3.1.3 Mineralogy class

The mineralogy criterion, relative to the nature of soil minerals, is represented over France in Figure 2.10. The characterization is based on X-Ray Diffraction (XRD) analyses, yielding percents of expansive clay minerals, carried out during the map production phase and afterwards backed up by literature. Correspondence between classes and percentages is given in Table 2.5.

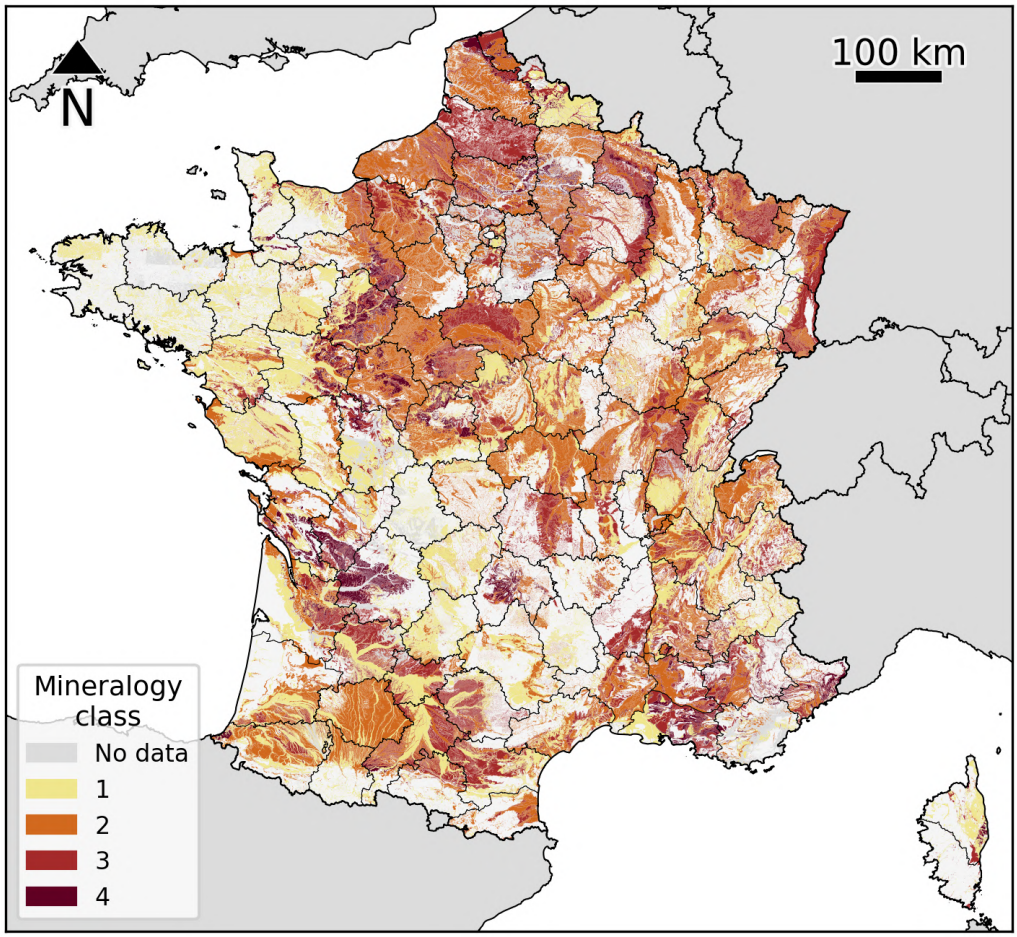


Figure 2.10: Mineralogy criterion over France. Classes are defined in Table 2.5.

Table 2.5: Description of mineralogy classes

Class	Average percent of expansive minerals
1	<25%
2	25-50%
3	50-80%
4	>80%

2.3.2 Susceptibility to clay shrink-swell

Susceptibility results from the average of the three criteria explained above, according to the rule shown in Table 2.6. The susceptibility of French soils is plotted in Figure 2.11. Class 1 of susceptibility is overrepresented compared to the other criteria because their sum is rounded down.

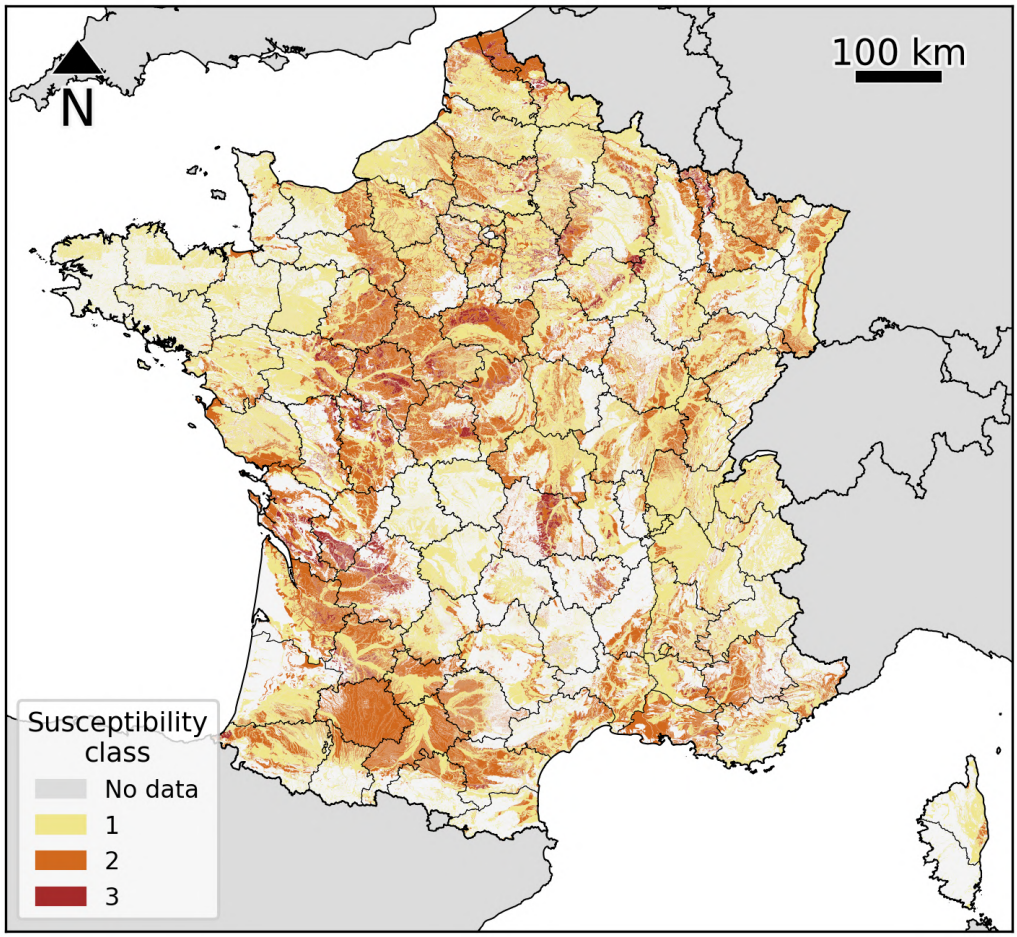


Figure 2.11: Susceptibility to clay shrink-swell over France. Classes are defined in Table 2.6.

Table 2.6: Estimation of susceptibility from the average of the 3 criteria

Class	Average of the 3 criteria
1	≤ 2
2	>2, ≤ 3
3	>3

2.3.3 Exposure to clay shrink-swell

For operational use, the susceptibility has been corrected with insurance claims data, resulting in the clay shrink-swell exposure map shown in Figure 2.12.

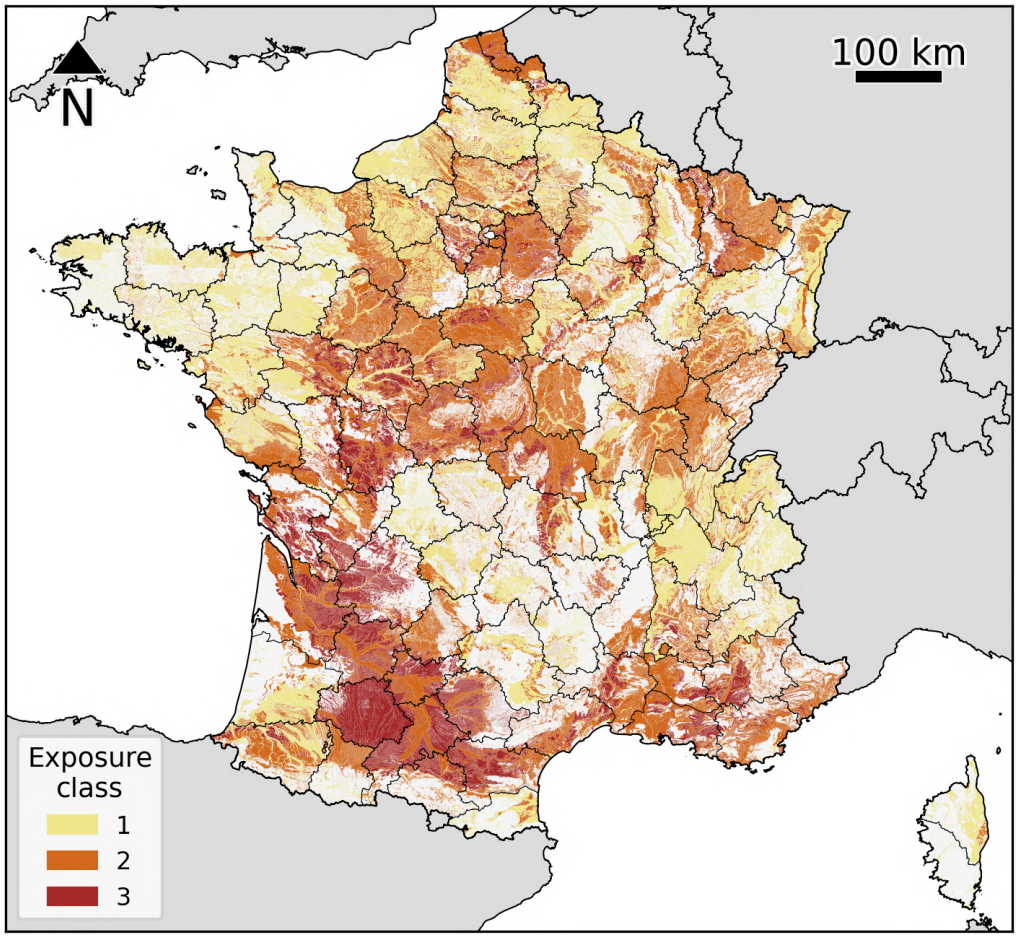


Figure 2.12: Exposure to clay shrink-swell for France. Classes are defined in Table 2.7.

Table 2.7: Definition of damage frequency classes

Class	Number of claims per km² of constructed area
1	< 2
2	2-10
3	> 10

The claims that were used for the exposure map are from the SILECC (Base de Données des Sinistres Indemnisés, Liés aux Evénements Cat-nat et Climatiques) database of the Mission Risques Naturels (MRN), an association of insurance companies. The database was formed by contributions of insurers and updated until 2017. The data allowed the calculation of a damage frequency per built area for each geological unit, from which classes were derived based on Table 2.7. Damage frequency classes were finally merged with double-weighted susceptibility classes, based on expert judgment.

2.3.4 Limitations

2.3.4.1 Superficial mapping

As expansive soil maps are based on geological maps, only outcropping units are characterized. Clay shrink-swell still can occur outside of exposure zones, if caused by clayey layers deep enough not to be identified, but shallow enough to be impacted by surface hydrological variations.

2.3.4.2 Lack of consistency between departments

The main limitation of the BRGM expansive soil maps is their lack of consistency, as they were achieved independently, department by department.

For example, the *j4-5W* geological unit called *Woëvre clay*, which overlaps 5 different departments in northeastern France, received different scores for each susceptibility criterion (3 and 4 for lithology, 2 and 3 for geotechnical properties, and 1, 2 and 3 for mineralogy). Although facies variations are expected at such large spatial scales, they should not follow administrative boundaries.

2.3.4.3 Scarcity of geotechnical and mineralogy data

Moreover, as geotechnical and mineralogical data are very limited, scores are sometimes based on a very small number of measurements (the mineralogy rating is often based on a single analysis by department). Individual criteria should therefore be interpreted with caution as the initial volume of data does not appear in the final score.

2.3.5 Uses of expansive soil maps

The three individual criteria for susceptibility, i.e. mineralogy, geotechnical properties and lithology, will be combined in the final chapter with the ISBA-based drought index in decision trees to assess their predictive capacity for insurance claims. Susceptibility and exposure are not included, as the former is redundant of the three criteria, and the latter already based on insurance data. The outlined limitations will not be compensated: they need to be borne in mind when drawing conclusions.

2.4 Instrumented site in Chaingy, Loiret (45)

The BRGM instrumented a house plot in Chaingy, in the Loiret department (45), France, with the purpose of better understanding how clay shrinkage occurs at a local scale.

The test site, described in Figure 2.13 and Figure 2.14, comprises a weather station, 10 FDR Thetaprobe soil moisture probes at a depth of 120 cm, and 2 extensometers.

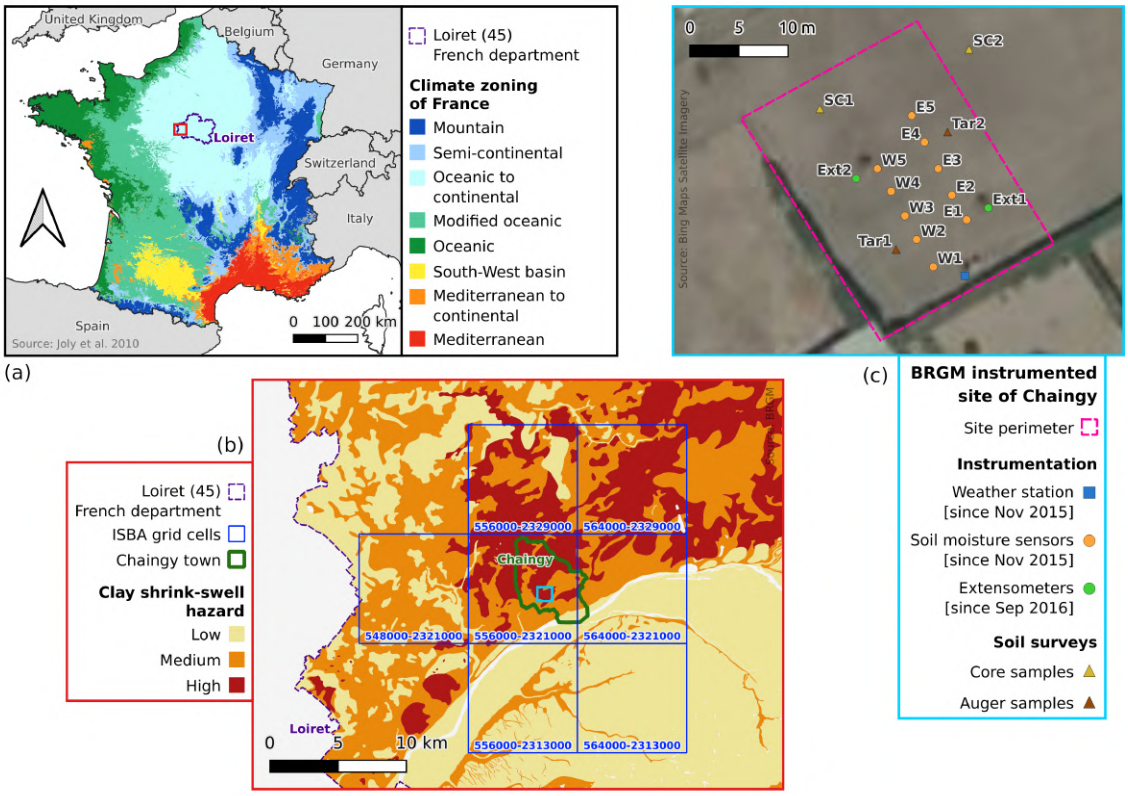


Figure 2.13: Context and disposition of the BRGM Chaingy instrumented site: (a) country scale, (b) regional scale, and (c) local scale

Figure 2.14: Photograph of the Chaingy instrumented site

The weather station records rainfall and air temperature. However, the rainfall record is not suitable for analysis due to numerous measurement interruptions. Indeed, all rainfall values

need to be summed for aggregating at a coarser time resolution, as opposed to soil moisture or displacement that can be averaged. The FDR probes measure volumetric water content, defined as the volume of water per volume of soil, through the soil's dielectric permittivity (see section 1.3.2.1). They all were calibrated in the same way, using a transfer function linking soil properties and water content supplied by the manufacturer. The probe precision is equal to 0.1 $m^3.m^{-3}$. Extensometers, detailed in section 1.3.2.2, estimate superficial vertical volume variations by measuring the distance between a plate buried at a depth of 30 cm and a 3-meter-deep anchor point in the bedrock. Their precision is better than 0.1 mm. The temporal resolution of the data obtained from the various instruments is 30 minutes.

The site is gently sloped from north-west to south-east, and is bordered on two sides by a hedge. Core samples represented on Figure 2.15, taken at locations SC1 and SC2 (see Figure 2.13 (C)), reveal the presence of 3 distinct layers: a backfilled topsoil (0-40 cm), a clay layer, either sandy as in SC1 or very plastic as in SC2 (40-180 cm), and a weathered limestone bedrock (below 180 cm). Given the agricultural context of the Chaingy municipality, this surface layer can be identified as former farmland reworked by ploughing. The geotechnical properties of the sandy clay were estimated from two auger samples taken at locations Tar1 and Tar2: the Plasticity Index (PI) ranges between 32 and 34, and the Methylene Blue Adsorption (MBA) ranges between 4.9 and 7.6. Furthermore, two oedometric tests were conducted on the sandy and plastic clays from SC1 and SC2 yielding Swelling Index (C_s) measurements of 0.001 and 0.039. These results collectively indicate a geotechnical class 3 characterization based on the BRGM criteria detailed in Table 2.4. More details on the instrumentation and the site context is available in Gourdier [2018].

(a) SC1

(b) SC2

Figure 2.15: Core samples made on the Chaingy instrumented site

The soil moisture probes have been operational since the end of 2015, one year prior to the

extensometers. Figure 2.16 presents the raw data acquired by all sensors. A considerable amount of data is missing in 2016 due to power issues.

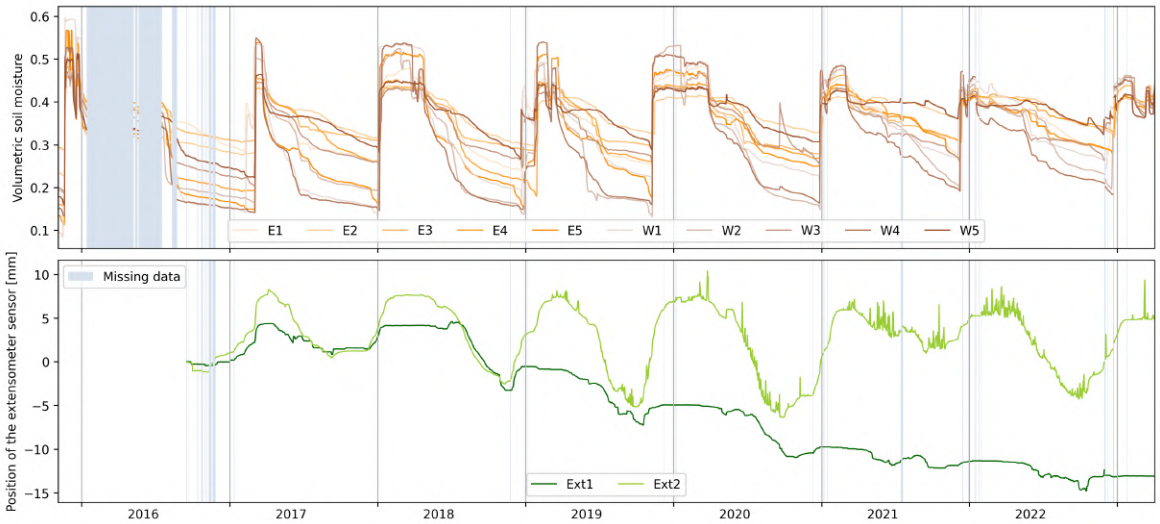


Figure 2.16: Volumetric soil moisture and displacement acquired on the Chaingy site

2.4.1 Soil moisture

All soil moisture sensors exhibit comparable patterns of annual periodicity, with a gradual decline from late spring to fall, and an abrupt rehydration occurring around the end of the calendar year. However, we observe different amplitudes: for example, W3 ranges from 0.15 to 0.55 m^3/m^3, while W5 ranges from 0.25 to 0.45 m^3/m^3. These differences can be attributed either to sensor calibration and thus local texture variations, or to local water content differences, which can be further amplified by the formation of desiccation cracks. Re-hydration is almost always synchronous for the ten probes. The temporal lag between the return of wet conditions in fall and the actual date of rehydration can be explained by both the important depth of the sensors (120 cm) and the low permeability of clayey soil. Although the distance from the hedge increases with the number of the probe, its effect on the soil moisture variations is not discernible.

2.4.2 Displacement

The signal of the two extensometers is drastically different. Ext2 is centered around 0 and has an annual periodicity with a constant amplitude range of 10 mm. In contrast, Ext1 gradually subsides and loses amplitude after 2018. The high-frequency peaks that appear in the Ext2 signal beginning in 2019 are a measurement artifact.

The fluctuations in Ext2 are indicative of soil swelling and shrinkage: they are positively correlated with soil moisture variations, although we note a difference. Swelling and shrinkage occur with the same kinetics, when soil desiccation is much slower than rehydration.

The subsiding signal recorded by Ext1 could result from irreversible shrinkage (also referred to as *inelastic subsidence* by Welch et al. [2024]), provided that the measuring system is

fully operational (in particular, that the anchor point is really stationary). However, one point challenges this hypothesis of irreversible shrinkage. Inelastic subsidence is anticipated to represent 10% of the shrinkage. Here, it is not the case as nearly all the volume decrease is permanent. We cannot identify any other natural soil-related phenomenon that would explain the observed subsidence.

2.4.3 Interdependence

The soil moisture signal recorded by the 10 sensors should be interpreted with caution for metrological reasons. Probes using FDR technology provide information on the volumetric water content, which is the ratio of the volume of water to the volume of the soil. However, when the soil swells or shrinks (as it is the case on our site), its volume changes. Consequently, in such conditions the volumetric moisture can vary for a constant gravimetric soil moisture value.

2.4.4 Uses of the Chaingy data

Data from the Chaingy site will provide a basis for comparison with ISBA. In particular, it is crucial to estimate uncertainties in soil moisture variations linked to the macroscopic scale of the model. This will be done by assessing variability observed at parcel level. This analysis will be done in chapter 4.

2.5 Programming environment

Since this thesis is essentially based on numerical modeling, we provide here details on the Python programming environment.

To manipulate large amounts of data, we used the packages *Pandas* [McKinney, 2010], and in particular *Geopandas* [Jordahl et al., 2020] for spatial data and *Xarray* [Hoyer and Hamman, 2017] for multidimensional data. Most of the figures were generated using the *Matplotlib* [Hunter, 2007] and *Seaborn* [Waskom, 2021] packages. Finally, the machine learning models used in chapter 6 were implemented using the *Scikit-learn* [Pedregosa et al., 2011] package.

Chapter 3

A custom drought index for clay shrinkage-induced subsidence

Contents

3.1 Introduction

The aim of this chapter is the development of a drought index to quantify the triggering conditions for clay shrinkage-induced subsidence.

The index is based on ISBA soil moisture simulations, that offer the advantages of flexible options (for testing different configurations) and of a wide spatial coverage (which will enable to make assessments at the national scale later on).

The evaluation criteria consists of a sample of subsidence claims aggregated at the town scale per year, from 2000 to 2018, from the ATHENA databases of CCR. The data is from 26 towns located in 5 different French departments (Haute-Garonne, Bouches-du-Rhône, Loiret, Moselle and Puy-de-Dôme), all of them highly concerned by the phenomenon. It is split in a validation set (20 towns, 4 in each department) and a calibration set (6 towns around Haute-Garonne). We normalize the number of claims by dividing them by the number of houses of each town in medium to high exposure zones, available from MTES [2021].

The methodology devised for this purpose consists in two main steps. First, annual drought characteristics are determined from daily (averaged from hourly) soil moisture simulations using different aggregation methods (number of days, minimum and integral below a soil moisture threshold value). Then, the resulting variable is statistically compared to the aggregated insurance claims by computing a Kendall tau (non-parametric) association metric. Soil moisture percentiles 1 to 50 are tested as drought threshold values. The statistical test is done separately for each department, merging the data of the 4 towns (76 values).

Several ISBA options (vegetation scheme, patch, soil layer depth) and index configurations (computation method, threshold value) were assessed in this manner, best correlations leading to maximal Kendall tau. The resulting index is called the Yearly Drought Magnitude Index (YDMI). In particular, results of the comparison of the ISBA A-gs options AST (prescribed LAI) and NIT (interactive LAI) were published in the Copernicus scientific journal *Natural Hazards and Earth System Sciences* [Barthelemy et al., 2024a]. They will be presented and commented in section 3.2. In addition to these results, an assessment separating patches and a comparison to operational ISBA versions will be presented in section 3.3 and section 3.4, respectively.

3.2 Comparison of AST and NIT ISBA schemes [Barthelemy et al., 2024a]

3.2.1 Summary

The methodology described above is applied to ISBA A-gs AST and NIT simulations with the aggregated patch configuration, testing different index computation methods. Although ISBA can extend from the surface to two meters deep, we only considered layers 5 to 8 (from 40 cm to 1 m) that are defined for the same patches, to have a vertical continuity. We have focused on the mid-depth layers rather than the most superficial ones, as the latter give mediocre results because they are too much influenced by short-term weather fluctuations. We obtained the following results, with a best configuration leading to an average Kendall tau of 0.42 (associated to *p-values* ≤ 0.001):

- The optimal drought threshold is equal to lowest soil moisture percentiles (1^{st} to 5^{th}): it implies that intense, summer drought events are the triggers for clay shrinkage.
- The best aggregation method consists in summing for each year soil moisture anomalies below the threshold. This metric combines both duration and severity characteristics of the drought event hence its highest performance.
- The deeper the layer, the higher the Kendall tau. We explain this by the fact that deeper layers are more isolated from the surface than superficial ones, and are consequently less influenced by weather fluctuations, reflecting only long term trends.
- Finally, correlation metrics are significantly improved with the NIT vegetation option rather than AST. It confirms that a detailed representation of plant metabolism is crucial to drought forecasting resorting to LSMs.

The resulting YDMI identified significant droughts in 2003 and 2018 for our sample of towns. These years were indeed associated to important numbers of claims. We comment on a bias in the insurance claims. As claims require the publication of a NatCat decree in order to be compensated, we identify situations in which the absence od the latter could explain inconsistencies between YDMI and damage. Local information that might influence both vertical displacement and building resistance, such as its characteristics or trees in their vicinity, is neglected through our wide study scale, which represents a major source of uncertainty.

3.2.2 Article

Nat. Hazards Earth Syst. Sci., 24, 999–1016, 2024
https://doi.org/10.5194/nhess-24-999-2024
© Author(s) 2024. This work is distributed under
the Creative Commons Attribution 4.0 License.

Natural Hazards and Earth System Sciences
Open Access
EGU

A new approach for drought index adjustment to clay-shrinkage-induced subsidence over France: advantages of the interactive leaf area index

Sophie Barthelemy[1,2,3], **Bertrand Bonan**[1], **Jean-Christophe Calvet**[1], **Gilles Grandjean**[2], **David Moncoulon**[3,†], **Dorothée Kapsambelis**[3], **and Séverine Bernardie**[2]

[1]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[2]Bureau de Recherches Géologiques et Minières (BRGM), Orléans, France
[3]Caisse Centrale de Réassurance (CCR), Department R&D Modeling Cat & Agriculture, Paris, France
†deceased, 18 July 2023

Correspondence: Jean-Christophe Calvet (jean-christophe.calvet@meteo.fr)

Received: 22 June 2023 – Discussion started: 27 July 2023
Revised: 25 January 2024 – Accepted: 2 February 2024 – Published: 21 March 2024

Abstract. Clay shrinkage, which consists of a reduction in the volume of clay soils during dry periods, can affect buildings and cause subsidence damage. In France, losses due to subsidence are estimated at more than EUR 16 billion for the period 1989–2021 (CCR, 2021) and are expected to increase under the effect of climate warming. This work aims to improve the current understanding of the conditions triggering subsidence by proposing an innovative drought index. We use a daily soil wetness index (SWI) to develop a new annual drought index that can be related to subsidence damage. The SWI is derived from simulations of soil moisture profiles from the interactions between soil–biosphere–atmosphere (ISBA) land surface model developed by Météo-France. The ability of the drought index to correlate with insurance claim data is assessed by calculating the Kendall rank correlation over 20 municipalities in France. The insurance data, aggregated by year and municipality, are provided by the Caisse Centrale de Réassurance (CCR). A total of 1200 configurations of the drought index are considered. They are generated by combining different calculation methods, ISBA simulation settings, soil model layers, and drought percentile thresholds. The analysis includes a comparison with the independent claim data of six additional municipalities and with a record of official "CatNat" (the French national natural disaster compensation scheme) decrees, useful for the analysis. The best results are obtained for drought magnitudes based on SWI values of the 0.8-to-1.0 m deep soil layer, an ISBA simulation with interactive leaf area index (LAI), and consideration of low drought SWI percentile thresholds. Comparison with claim data shows that drought magnitude is able to identify subsidence events while being spatially consistent. This drought magnitude index provides more insight into subsidence triggers while benefiting from advanced land surface modeling schemes (interactive LAI, multilayer soil). This work paves the way for more reliable damage estimates.

1 Introduction

Clay shrink–swell is the change in volume of clayey soils in response to changes in soil moisture. This phenomenon is related to the high affinity of certain clay minerals commonly found in soils for water molecules.

These minerals, structured in aggregates, swell and shrink under wet and dry conditions, respectively (Bronswijk, 1989). In temperate European regions such as in France, soils are in a hydrated state for most of the year, which means that exceptional movements occur during droughts (Cojean, 2007). On the other hand, surface soil moisture variations result from soil–atmosphere interactions. Precipitation brings in water, only some of which infiltrates into the soil, while evaporation causes water loss. Vegetation affects both water inputs, by intercepting precipitation and preventing some of

it from reaching the soil surface, and water outputs, through root water uptake and transpiration (Tessier et al., 2007).

Shrinking ground movements caused by droughts can affect buildings through their foundations, causing what is known as subsidence damage (Doornkamp, 1993). The damage is physically caused by the inability of a building to accommodate a relative displacement of its various parts, leading to structural failure. This relative displacement observed during droughts is caused by the different behavior of exposed shrinking and sheltered soils. The latter are immobile due to their constant moisture content. Local factors that increase spatial differences in soil moisture, such as the presence of vegetation, significant exposure to solar radiation, or a deficient domestic drainage system, are known to exacerbate subsidence (Page, 1998).

In France, subsidence losses were reported after the 1976 summer drought. A few years later, hundreds of thousands of houses were affected by the 1989 drought, which led to the inclusion of this peril in the French national natural disaster compensation scheme (CatNat) (Salagnac, 2007). Today, subsidence is still covered under the CatNat regime, where it is the second most costly peril after floods. Between 1989 and 2021, losses amounted to more than EUR 16 billion (CCR, 2021). Historically, the southwest of France has been the most exposed to subsidence. However, a significant number of events have been reported in the northeast since 2016 (CCR, 2022), showing a geographical expansion of the importance of subsidence in France. Since then, the cost of subsidence costs has averaged EUR 1 billion per year and even exceeded EUR 2 billion in 2022, which was marked by an exceptional drought (Toreti et al., 2022).

Several studies on subsidence damage (e.g., Crilly, 2001) have shown that single-family houses are the most vulnerable type of building because they tend to have shallow foundations. In 2019, approximately 19 million single-family houses were identified in France, more than half of which are located in zones with a medium to high exposure to clay shrinkage (MTES, 2021a). In addition, research led by Météo-France, the French meteorological service, within the CLIMSEC project (https://www.umr-cnrm.fr/spip.php?article605, last access: 11 March 2024) on the impact of climate change on soil water resources provided insight into climate-change-induced drought trends for the country (Vidal et al., 2010, 2012). It was shown that the duration and spatial extent of drought events affecting soils are bound to increase as a result of rising air temperatures and changes in the precipitation regime.

In this context, both policymakers and the insurance industry need more accurate predictions of subsidence losses. Research is needed to develop a better understanding of this phenomenon and of its financial consequences.

Subsidence depends on predisposing factors, such as the type of clayey soil or land use, and on triggering factors. The first studies to investigate the dependence of clay-shrinkage-induced subsidence on climate (Harrison et al., 2012; Corti et al., 2009, 2011) confirmed the ability of drought to trigger subsidence. These analyses were based on precipitation and air temperature data and recommended the use of land surface models (LSMs) as they provide more detailed information on droughts. In the wake of these findings, several works focusing on subsidence in France have used the soil wetness index (SWI) outputs of the interactions between soil–biosphere–atmosphere (ISBA) land surface model (LSM; Noilhan and Planton, 1989; Noilhan and Mahfouf, 1996) developed by Météo-France, in a simplified configuration (monthly averages, homogeneous vegetation and soil). This product, called "SWI uniforme", is also used for drought monitoring in the CatNat regime. However, the criterion based on this SWI uniforme product for the promulgation of decrees has changed several times, making it impossible to link the history of decrees to quantitative drought characteristics. Using SWI uniforme, a number of studies have assessed the impact of climate change on subsidence costs (Gourdier and Plat, 2018; CCR and Météo-France, 2018; Covéa and RiskWeatherTech, 2022). All of these papers concluded that subsidence-related costs would increase, but no consensus was reached on the magnitude of the increase. Other studies attempted to improve the predictive power of subsidence damage models using advanced machine learning techniques (Heranval et al., 2022; Ecoto and Chambaz, 2022; Charpentier et al., 2022).

Motivated by the above context, a new drought index is proposed in this study, which is adapted to clay-shrinkage-induced subsidence damage. The novelty of the present work lies in the analysis of simulations using a version of the ISBA LSM with an extended vegetation representation. The index is based on soil moisture simulations from a version of ISBA capable of simulating the details of the soil moisture profile together with an interactive vegetation leaf area index (LAI) and is adjusted by comparison with insurance data. The objective is to determine (1) which soil layers should be considered for subsidence damage prediction and (2) the extent to which interactive LAI is needed.

In Sect. 2, all data sets and the methodology for computing the drought index are presented. In Sect. 3, we detail the obtained optimal drought index and its relationship to the subsidence damage data. Comparison with an independent data set and with a data set of CatNat decrees is included. All of these results are discussed in Sect. 4 before concluding in Sect. 5.

2 Data and methods

The objective of this work is to create a new soil drought index that is adapted to induced subsidence losses. We here describe how the indices are computed from "LAI_clim" and "LAI_model" soil moisture simulations and how they are evaluated through a pairwise ranking with the insurance damage data. A comparison with independent data is in-

cluded in the analysis. The methodology is summarized as a flowchart in Fig. 1.

2.1 Soil moisture simulations

Soil moisture is simulated by ISBA within version 8.1 of the SURFEX (Surface Externalisée) modeling platform developed by Météo-France for numerical weather prediction and climate modeling (Masson et al., 2013; Le Moigne et al., 2020). ISBA calculates surface water and energy budgets in response to an atmospheric forcing. Soil moisture at a given time results from the balance between water inflows, from precipitation, and outflows, through runoff, infiltration, and evaporation. ISBA covers the entire French metropolitan area, on an 8 km meshed grid. It is used operationally for monitoring water resources in the SAFRAN-ISBA-MODCOU (SIM) hydrological suite (Habets et al., 2008), together with a meteorological analysis and a hydrogeological model. ISBA is used here "offline"; i.e., there is no feedback from the surface to the atmosphere. The SAFRAN reanalysis (Durand et al., 1993; Quintana-Seguí et al., 2008) is used as the atmospheric forcing.

A multilayer version of ISBA, ISBA-DIF (Boone et al., 2000; Decharme et al., 2011), with a diffusive scheme is used. The approach is to divide the soil column into layers whose respective thicknesses increase with depth to better represent hydrological processes, since important water and temperature gradients at the surface require a finer mesh. The number of layers considered is proportional to the rooting depth of the vegetation, reaching a maximum of 10 layers (adding up to 2 m depth) in the case of trees. It is important to analyze layers at different depths because moisture variations depend on both the distance to the surface and the vegetation roots.

As explained in Decharme et al. (2011), soil water-holding capacity and soil hydraulic parameters such as porosity, matric potential at saturation, and saturated hydraulic conductivity are related to soil texture in the model through empirical pedotransfer functions. The latter are described in Noilhan and Lacarrère (1995) and derived from Clapp and Hornberger (1978). In ISBA, texture is considered to be homogeneous throughout the whole soil column for each grid point. This texture is represented by clay, sand, and silt contents derived at the kilometer resolution from the Harmonized World Soil Database (HWSD) version 1.2 (Nachtergaele et al., 2012). Because soil texture is averaged within a model grid cell, this approach provides a large-scale representation but is not representative of what may be found on a house lot. Inferring the hydromechanical behavior of clayey soils from this information alone would be problematic. This is not the purpose of the drought index developed in this study.

To account for spatial variability within a grid cell, ISBA first runs separately for 12 generic plant functional types called patches (bare soil, rock, permanent snow, deciduous broadleaf trees, evergreen broadleaf trees, coniferous trees, C_3 crops, C_4 crops, irrigated C_4 crops, grassland, tropical grassland, and wetlands). The runs are then aggregated by averaging the output variables, weighting each patch by its respective fraction in the grid cell. The maximum soil depth of a patch depends on the rooting depth of the corresponding vegetation type and varies from patch to patch. The geographic distribution of the patches as well as the land surface parameters is obtained from the ECOCLIMAP-II database at kilometer resolution (Faroux et al., 2013). The present work focuses on layers simulated for all patches, as we aim to assess differences induced only by depth. These conditions were met for four soil layers: 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1.0 m, soil layers 5, 6, 7, and 8, respectively. Accordingly, we based our analysis on these four model layers. In this study, we are limited to a maximum depth of 1 m due to the patch continuity requirement. One of the direct consequences of climate warming is the spread of deep soil desiccation under severe and recurrent drought conditions. This can reach soil layers deeper than 1 m depending on the close environment configuration. Knowing that the amplitude of soil moisture variations decreases with depth (Ravina, 1983), a drought that reaches deep soil layers will intensively dry out shallow layers. Theoretically, such an event can be detected with surface layer information alone. However, it should be noted that drying occurs with a time lag that increases with depth. In this analysis, we have verified that even for the deepest model layers, the drought observed in a given year never overlaps with the following year, which would distort the index calculations. This assumption may be questioned under future climatic conditions, considering that the frequency and intensity of droughts in France will increase.

Because vegetation plays an important role in near-surface soil moisture variations, the ISBA model includes a representation of photosynthesis using a CO_2-responsive stomatal conductance scheme, ISBA-A-gs (Calvet et al., 1998). This scheme has been improved over time to account for specific plant responses to drought (Calvet, 2000; Calvet et al., 2004).

In the operational SIM product (SURFEX v8.0), the leaf area index (LAI), which quantifies the leaf surface available for exchange with the atmosphere, is prescribed from ancillary data. For each grid point, the LAI changes every 10 d, and the cycle is repeated every year, based on a climatology established from SPOT-VGT satellite data for the period from 1999 to 2005. A first simulation with this configuration is tested in this study and is called LAI_clim.

In the ISBA-A-gs configuration, the LAI is not prescribed but simulated from the modeled leaf biomass, taking into account the mass-based leaf nitrogen concentration (Calvet and Soussana, 2001). Phenology is driven by photosynthesis. Since the modeled photosynthesis depends on soil moisture, leaf temperature, solar radiation, and air humidity, all environmental conditions can affect the simulated LAI. Based on this concept, the LAI is updated daily. The modeled LAI has been validated at global (Gibelin et al., 2006), continental

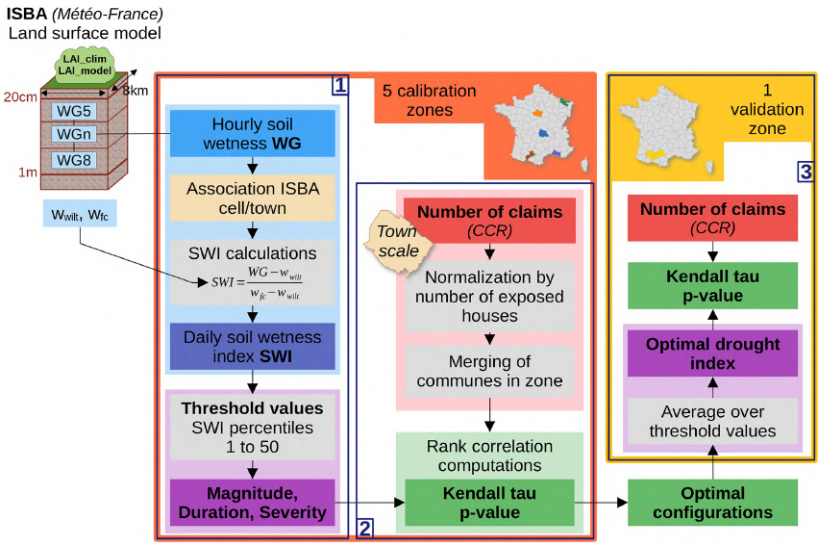


Figure 1. Flowchart of the methodology deployed in this study with focus on the three main steps.

(Szczypta et al., 2014), and regional (Brut et al., 2009) scales. In this work, a second simulation is tested with this configuration, which we call LAI_model.

Each simulation provides hourly volumetric soil moisture data for four 0.2 m thick soil layers from 0.2 to 1.0 m depth. Since we are focusing on longer-term droughts, the diurnal periodicity of the data is not required, and the hourly soil moisture values are averaged to daily values. In order to consider a single definition of drought, independent of the fact that water-holding capacity varies with soil texture, we convert volumetric soil moisture to soil wetness indices (SWIs). While volumetric soil moisture is expressed in $m^3\ m^{-3}$, SWI is unitless. SWI is derived by rescaling volumetric soil moisture between wilting point w_{wilt} and field capacity w_{fc} as illustrated in Fig. S1 (see the Supplement). These two quantities are derived from the model soil texture through soil pedotransfer functions (Boone et al., 1999; Decharme et al., 2013). We refer to the SWI of layers 5, 6, 7, and 8 as SWI5, SWI6, SWI7, and SWI8, respectively.

2.2 Natural disaster (CatNat) decree database

In France, financial losses to insured property following a natural catastrophe may be covered by a national compensation scheme, the "régime d'indemnisation des catastrophes naturelles" (CatNat regime). Compensation under this scheme is, with a few exceptions, subject to the official recognition of the state of natural disaster through the publication of a decree, an "arrêté de reconnaissance de l'état de catastrophe naturelle" (CatNat decree), for a given year and municipality. Details of each CatNat decree (accepted and rejected requests) are available as individual files on the CCR website (https://catastrophes-naturelles.ccr.fr/les-arretes, last access: 11 March 2024). For this analysis, we retrieved and merged these data. This information is relevant for our research because subsidence damage has only been compensated by this scheme since 1989.

Since 2000, the publication of a CatNat decree has been conditional on meeting a geotechnical criterion and a meteorological criterion. The geotechnical condition requires the municipality to have at least 3 % of its surface mapped as shrinkable clay, based on the Bureau de Recherches Géologiques et Minières (French Geological Survey, BRGM) exposure to clay shrink–swell map. The meteorological condition aims to identify exceptional periods of drought. Its current version requires the 3-month average of SWI uniforme to be the lowest or second lowest of a 50-year reference associated with the same month. This criterion has evolved several times since the creation of the system as the phenomenon has been studied. The condition established in 2000 is a "winter" criterion, based on the rainfall of the winter preceding the drought event. Later, the timing of the successive drought events of 2003, 2011 and 2018 led to the creation of "summer", "spring", and "fall" criteria that were added to the existing criterion.

2.3 Normalized insurance data

The insurance data used in this study are provided by the Caisse Centrale de Réassurance (CCR), the French public reinsurance company that manages a large database of its clients' contributions. It consists of the number of subsidence claims aggregated per year and per municipality, from 2000 to 2018. By "municipality", we mean a commune, the small-

est administrative unit in France. Subsidence claims correspond to accepted "legitimate claims", after a national recognition step of CatNat requests. The representativeness of the data varies over the years and the municipalities considered, depending on the proportion sent by insurers. Total and average annual costs were also provided by CCR, but we focused on the number of claims, which gave more satisfactory results. The data are aggregated at the municipal level for confidentiality reasons. In addition, this is an official and frequently used zoning classification, so there are a lot of additional data available.

For normalization purposes, the large number of subsidence claims is divided by the number of single-family houses in the municipality located in areas prone to subsidence. This information comes from a data set developed by the French Ministry of Ecological Transition (MTES/CGDD/SDES) (MTES, 2021b). It provides the number of houses per municipality in different exposure classes (no exposure, low, and medium to high). The zoning is based on a map of exposure to clay shrink–swell developed by the BRGM (BRGM, 2024), coupling a national susceptibility map (based on geotechnical data and 1 : 50 000 geological maps) and insurance statistics. We call the new variable obtained after normalization the "normalized number of claims".

In this study, we use data from 26 municipalities, the locations of which are shown in Fig. 2. Although the spatial resolution of the insurance data leads us to an analysis at the municipal level, we generalize the results by repeating the process for areas exposed to different climates. The study is limited to a selection of municipalities with a significant history of clay shrink–swell losses, since only the dependence on climate is studied (see Sect. 3).

To investigate the influence of this national recognition stage on the end-of-chain insurance data, we also used the history of accepted and refused national recognition requests for the towns forming our sample. These data were obtained by merging individual decree files downloaded from the CCR website (https://catastrophes-naturelles.ccr.fr/les-arretes, last access: 11 March 2024).

2.4 Drought indices

Providing a universal and quantitative definition of drought is challenging because drought can affect different levels of the hydrological cycle: precipitation, soil moisture, and streamflow (Dracup et al., 1980). Numerous indices characterizing different types of drought are available in the literature. For example, the Standardized Precipitation Index (SPI) (McKee et al., 1993) quantifies meteorological droughts by fitting accumulated precipitation at different timescales to a given distribution. The same methodology has already been adapted for agricultural droughts, where calculations are based on the SWI, creating a standardized soil wetness index (SSWI) (Vidal et al., 2010). As Charpentier et al. (2022) recall from the conclusions of Soubeyroux et al. (2012), the SPI and the SSWI provide complementary information: for example, the 2003 drought over France, which was caused more by a heatwave than a lack of precipitation, is only detected by the SSWI.

Given the nature of the clay shrink–swell phenomenon, we focus here on droughts affecting soil moisture by analyzing the SWI outputs of the ISBA model. LSMs are capable of accounting for the atmospheric forcing, soil and vegetation characteristics, and initial conditions. They are reliable tools for studying soil moisture variations.

Subsidence damage affects several regions located throughout France and is therefore subject to different climates (see Fig. 2 for a climatic zoning of the country). We consider the tipping point for subsidence not as an absolute value of soil moisture but rather as a measure of the deviation from the mean. This definition is consistent with the physical mechanism of subsidence damage, which is related to a difference in soil moisture content relative to its average value (Sect. 1).

In this study, we define droughts as periods in which the value of the daily SWI is below a threshold value. The latter corresponds to a certain percentile of the empirical SWI distribution calculated over a given ISBA grid point. Based on in situ observations in the United States and Canada, Ford et al. (2016) showed that the record length required to obtain stable daily soil moisture distributions ranges from 3 to 15 years. Assuming that these results are applicable to France, our study period of 19 years is sufficient to obtain reliable SWI percentiles. The use of percentiles also avoids the need to standardize SWI values, as the shape of their distributions is quite variable depending on the local climate (D'Odorico et al., 2000; Vidal et al., 2010). In this study we deliberately make no assumption about the triggering drought frequency and therefore consider thresholds ranging from the first percentile to the median.

Once drought periods have been identified using the daily SWI percentile criterion, several options are considered to quantify the drought events. For this purpose, the characteristics of drought events are described in Vidal et al. (2010): severity, duration, and magnitude. We apply this concept to our definition of drought:

- Severity is the maximum deviation of the SWI from the threshold.
- Duration is the number of days below the threshold.
- Magnitude is the sum of the daily SWI deficit values (equivalent to an integral) below the threshold.

All three indices are considered. Given the temporal resolution of the insurance data, we assume only one major drought event per year, as we are working on an annual scale. Therefore, a single index value is calculated each year.

For each municipality, the three indices are computed for two model configurations (LAI_clim and LAI_model), four

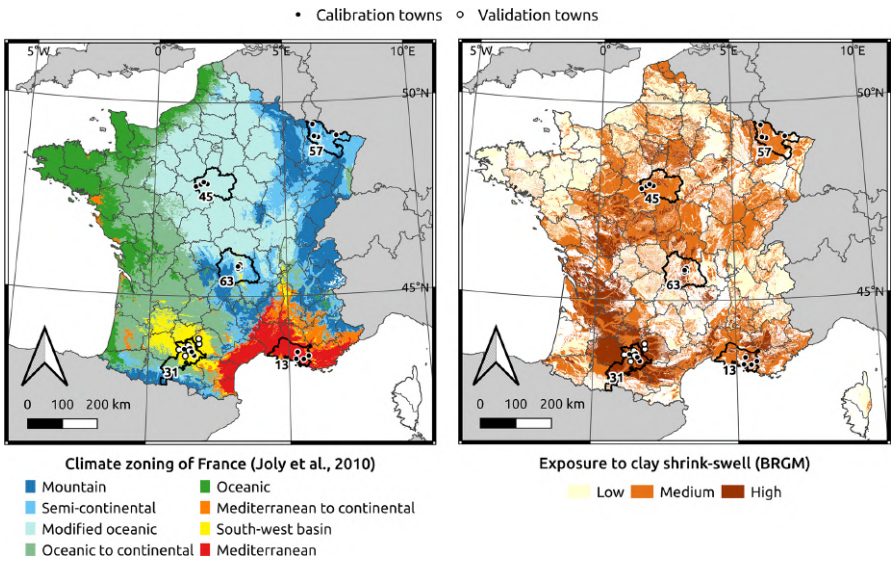


Figure 2. Location of the 26 study municipalities, along with a climate (Joly et al., 2010) and clay shrink–swell exposure map (BRGM, 2024). The black markers depict the municipalities used for calibrating the indices, and the white ones are used for validation. Bold lines indicate the five French departments (13, 31, 45, 57, 63) in which are located the calibration subsets (Bouches-du-Rhône, Haute-Garonne, Loiret, Moselle, Puy-de-Dôme).

model soil layers 5 to 8, and 50 SWI percentile values (from 1 to 50). This corresponds to 1200 configurations.

It should be noted that although the conversion from volumetric soil moisture to SWI is useful for unifying the information on a national scale (Fig. S1), it does not affect the magnitude calculations. In fact, the calculations are based on thresholds defined by percentiles of daily soil moisture, the distribution of which is unchanged by the linear transformation that is the conversion to SWI.

2.5 Pairwise ranking of years

To assess the ability of the drought indices to represent subsidence damage for a given year, we compute the Kendall rank correlation coefficient (Wilks, 2006) between the two variables, over the entire study period. The Kendall rank correlation test compares the respective ranks of paired data by counting concordant and discordant pairs. It produces a coefficient ranging from 0 to 1, the Kendall τ, and a p value (the null hypothesis being that the variables are not correlated). This test has the advantage of being nonparametric and robust to extreme values.

To maximize the number of observations used to compute the Kendall τ, we merge data from multiple municipalities to form regional subsets. An association is performed to pair the municipal insurance data with the soil moisture available on the ISBA grid. We compare the soil and vegetation characteristics of all the ISBA cells that are less than 2.5 km from the municipality boundary, ultimately selecting the cell with the highest grass and clay content (most representative of a damaged building environment) as the most representative cell. We choose to base the selection on soil and vegetation rather than distance because we believe that outside of mountainous areas, the atmospheric forcing does not change drastically for two neighboring cells.

For each regional subset, the Kendall τ is computed for all the 1200 possible drought index configurations. The highest Kendall τ values are used to identify optimal configurations. Correlation calculations are performed sequentially for regional subsets and are combined by calculating a mean Kendall τ.

2.6 Validation

The best drought index configuration is then applied to independent data to verify that we avoid overfitting. For this purpose, the drought index is computed for all the municipalities not used in the calibration (forming a validation set) and is confronted with the corresponding normalized number of claims. For further analysis, the optimal drought indices of both calibration and validation sets are sorted into classes and the distributions of the normalized damage counts are compared. Four drought index classes are defined from the calibration set: a first class that includes all zero values (no drought) and three other classes that divide the rest of the population into equal groups. The class delineation is applied to the validation set.

Finally, both the drought index and the normalized insurance data are compared to the set of historical CatNat decrees.

For each subset, we assess the ability of the optimal drought index to predict claims by counting contingencies and deriving scores (Wilks, 2006). First, we transform quantitative optimal drought magnitudes and normalized claim counts into Boolean occurrences of predictions and observations, respectively. The correspondences for each subset are reported in a contingency table. We refer to the four metrics in the contingency table as true positives, true negatives, false positives, and false negatives (TP, TN, FP, and FN, respectively). The performance of the index in identifying damage patterns is then evaluated for each set by calculating scores. From the variety of scores available in the literature, we selected the proportion correct (PC), bias (B), false-alarm ratio (FAR), probability of detection (POD), and probability of false detection (POFD). Each of these metrics is described in detail in Table 1.

2.7 Study area

The calibration set consists of five subsets of four municipalities distributed throughout France, in different climatic settings (continental, modified oceanic, Mediterranean, and southwest basin, based on the typology proposed by Joly et al., 2010). The validation set merges six municipalities in the southwest basin climate setting, from the same classification. We choose municipalities corresponding to different urban contexts (dense urban to rural) to have a representative sample as Corti et al. (2011) showed that urban centers are not susceptible to subsidence, while discontinuous urban areas are half as vulnerable as rural areas.

The locations of each set, as well as the climate (Joly et al., 2010) and exposure to clay shrink–swell maps (BRGM, 2024), are detailed in Fig. 2. Each calibration subset is referred to as the corresponding French department (administrative subdivision): Bouches-du-Rhône (13), Haute-Garonne (31), Loiret (45), Moselle (57), and Puy-de-Dôme (63).

3 Results

3.1 Assessment of the drought index configurations

For the calibration set consisting of the five regional subsets, the 1200 drought index configurations are evaluated by measuring the Kendall rank correlation with the normalized number of claims. Figure 3 shows the Kendall τ coefficients averaged over the calibration subsets and by 5th-percentile threshold groups for each drought index type, simulation configuration, and model soil layer. Each group, of equal size, consists of 25 values (5 calibration subsets $\times$ 5 thresholds). The maximum p value obtained per group is plotted in the Supplement (Fig. S2).

Figure 3 shows that the average Kendall τ calculated over the groups ranges between 0.13 and 0.42. Similarly, the maximum p value observed per group ranges between 0.0001 and 1. For both LAI_model and LAI_clim simulations, and especially for the duration index, the results tend to get worse as the thresholds get higher. We also note that the performance of the severity index decreases significantly for the most superficial 0.2–0.4 m model soil layer. The most stable index is the magnitude index. For extreme thresholds (1st to 5th percentiles), there is little difference between the three indices, and the average Kendall τ increases with the depth of the soil layer. Finally, the highest correlation coefficients are obtained with the LAI_model simulation, extreme thresholds, and layers below 0.6 m (model soil layers 7 and 8). We also observe with the LAI_model configuration that the p value of each group is always below 0.001 for model soil layer 8 (0.8–1.0 m); cf. Fig. S2 in the Supplement.

3.2 Optimal drought index

We define an optimal drought index from the LAI_model simulation, the SWI of the deepest model soil layer (layer 8, from 0.8 to 1.0 m), and averaged magnitudes computed with thresholds corresponding to the 1st to 5th percentiles. This index is hereafter referred to as the optimal drought magnitude.

Figure 4 plots the average optimal drought magnitude of the five calibration subsets against the corresponding average normalized number of claims over the entire study period. The error bars indicate the amplitude of the values (minimum and maximum of the variable in the subset). The same data are also presented as a scatterplot in the Supplement (Fig. S3).

The CatNat decrees issued over the period are also shown in Fig. 4, as gray bars whose shading changes with the number of decrees per year and subset. Over the whole period, we count 20, 25, 6, 12, and 16 decrees promulgated for subsets 13, 31, 45, 57, and 63, respectively.

From Fig. 4, we can see that the optimal drought index identifies 2003 and 2018 as the most important drought years for almost all subsets. This pattern appears in the normalized number of claims with some gaps. Discrepancies are also identified for years with zero drought magnitude when claims were reported, with the example of 2005 for Department 13. Some local drought events are detected by both magnitude and insurance data, such as 2011 for Department 31. The amplitude of the magnitudes indicates that when a drought is detected, every municipality in the calibration subset is affected. This means that the index is spatially consistent. Overall, there is a good correspondence between zero magnitudes and the absence of damage. Furthermore, we find a very strong association between normalized damage counts and CatNat decrees.

The number of rejected claims is also shown in Fig. 4. Situations in the calibration and validation sets can be identified

Table 1. Scores for assessing the performance of the optimal drought magnitude, based on the fractions of true positives, true negatives, false positives, and false negatives (TP, TN, FP, and FN, respectively).

Name	Equation	Optimal value	Definition
Proportion correct	PC $= (TP + TN) / (TP + TN + FP + FN)$	1	Ability to predict correctly (accuracy)
Bias	B $= (TP + FP) / (TP + FN)$	1 (unbiased)	Tendency to see more (>1) or fewer (<1) predictions than observations
False-alarm ratio	FAR $= FP / (TP + FP)$	0	Fraction of predictions that turn out to be wrong
Probability of detection	POD $= TP / (TP + FN)$	1	Fraction of occurrences of the event for which there was a prediction
Probability of false detection	POFD $= FP / (FP + TN)$	0	Fraction of wrong predictions when the event did not occur

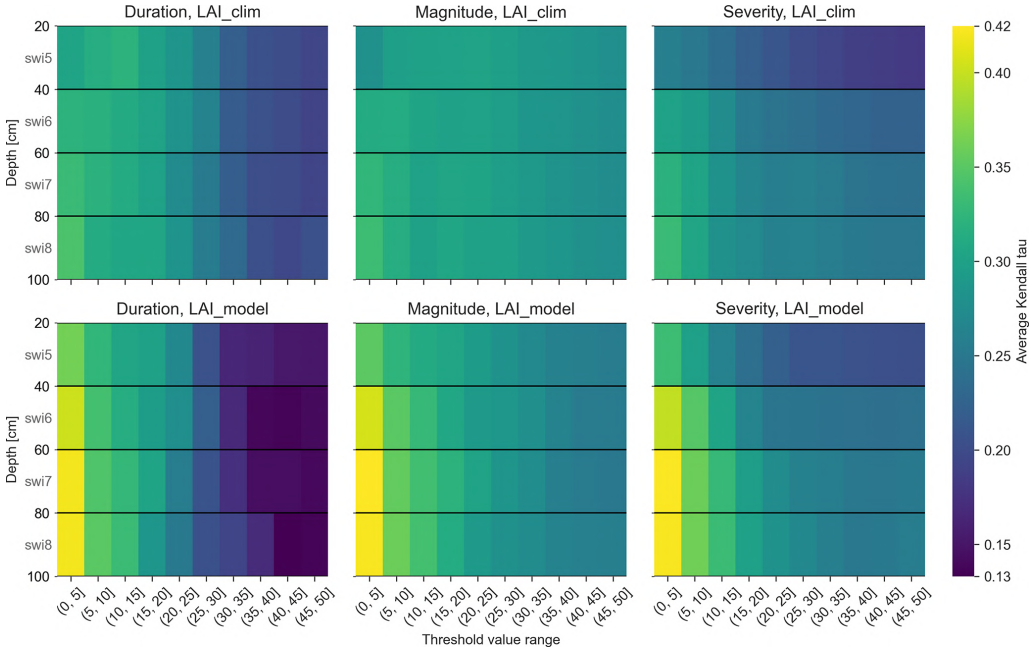


Figure 3. Average Kendall τ per group, computing the rank correlation between drought index and normalized number of claims, for all calibration subsets, separating index type (duration, magnitude, severity), model simulation (LAI_clim, LAI_model), model layer (SWI5 to SWI8, i.e., 0.2–0.4 to 0.8–1.0 m), and threshold value range (from 1st to 50th percentiles in groups of five).

where rejected claims could be the source of inconsistencies between the drought index and claims. In particular, all the inconsistencies noted between index and claims (positive index and no claims) observed in 2003 and 2018 can be explained by this factor (three inconsistencies in 2003 and one in 2018).

The predicted counts and scores calculated for all subsets are shown in Table 2. For all calibration subsets, the PC and POD are above 0.5 and the FAR and POFD are below 0.5 (except for the high FAR of 0.57 for zone 57). On the other hand, the obtained bias ranges between 0.64 and 1.75, with the subset having the smaller bias (closest to 1) being subset 45 ($B = 1.18$). The scores are averaged when the calibration subsets are merged, compensating for differences.

3.3 Validation

For the six municipalities of the validation set, we compute optimal drought magnitudes (LAI_model, layer 8, average

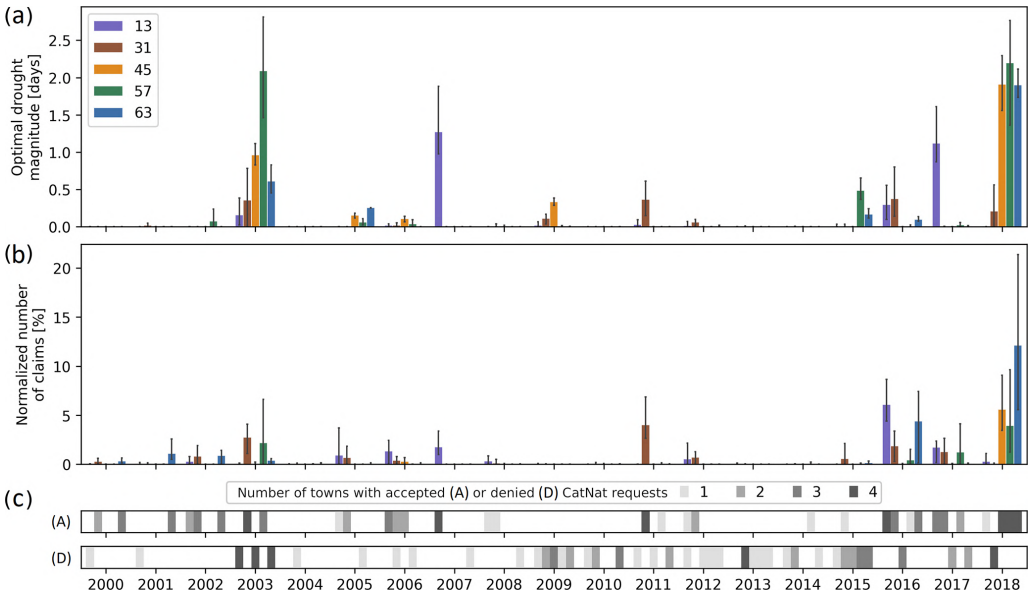


Figure 4. Optimal drought magnitude **(a)** and normalized number of claims **(b)** averaged by calibration subset and numbers of towns per subset **(c)** with accepted (A) or denied (D) CatNat requests. The error bars indicate the amplitude of values.

Table 2. Table of contingencies (TP, TN, FP, FN) and scores (PC, *B*, FAR, POD, POFD) for calibration and validation subsets. The highest and lowest values are in bold and in italics, respectively.

Subset		Number	TP (%)	TN (%)	FP (%)	FN (%)	PC	*B*	FAR	POD	POFD
Calibration	13	76	34	33	*8*	25	0.67	0.71	*0.19*	0.58	*0.19*
	31	76	**40**	*14*	9	**37**	*0.54*	*0.64*	*0.19*	*0.52*	0.39
	45	76	21	**58**	13	8	**0.79**	1.18	0.38	0.73	*0.19*
	57	76	*20*	47	**26**	*7*	0.67	**1.75**	**0.57**	**0.75**	**0.36**
	63	76	27	38	13	22	0.64	0.81	0.33	0.54	0.26
	All	380	28	38	14	20	0.66	0.88	0.33	0.59	0.27
Validation	All	114	29	28	16	27	0.57	0.80	0.35	*0.52*	0.36

of the 1st to 5th percentiles). The Kendall rank correlation between the optimal drought magnitude and the normalized number of claims yields a τ value of 0.23 and a p value of less than 0.001, indicating a statistically significant correlation. Figure 5 plots the average optimal drought magnitude and the normalized number of claims for this set over the 19-year study period. CatNat decrees are shown as gray bars in the background of the second graph. These results are also presented in a scatterplot in the Supplement (Fig. S4).

In Fig. 5, we see that the magnitude index is able to identify droughts in more than half of the municipalities for the years 2003, 2009, 2011, 2012, 2014, 2016, and 2018. However, not all of these years correspond to years with significant damage (2002, 2003, 2011, 2012, 2016, 2017), and some years even remain undetected. As with the calibration set, we observe a very strong correspondence between normalized damage numbers and CatNat decrees. The scores obtained for the validation set, displayed in Table 2, are quite similar to the ones obtained for the merged calibration subsets.

Then, the optimal magnitudes of calibration and validation sets are sorted into four classes. Statistics describing the normalized number of claims per magnitude class for both sets are available in Table 3. Figure 6 shows the distribution of the same variable per class, separating the two sets.

From Table 3 and Fig. 6, we can see that the entire population is evenly distributed among the classes for the two sets, except for classes 2 and 3 of the validation set, which contain more and fewer elements than class 1, respectively. For the calibration set, the value of the damage quartiles increases with the magnitude class, indicating a positive association. The same is observed for the validation set for the lower and middle quartiles. We observe that the distributions of the normalized number of claims are similar for both sets in classes

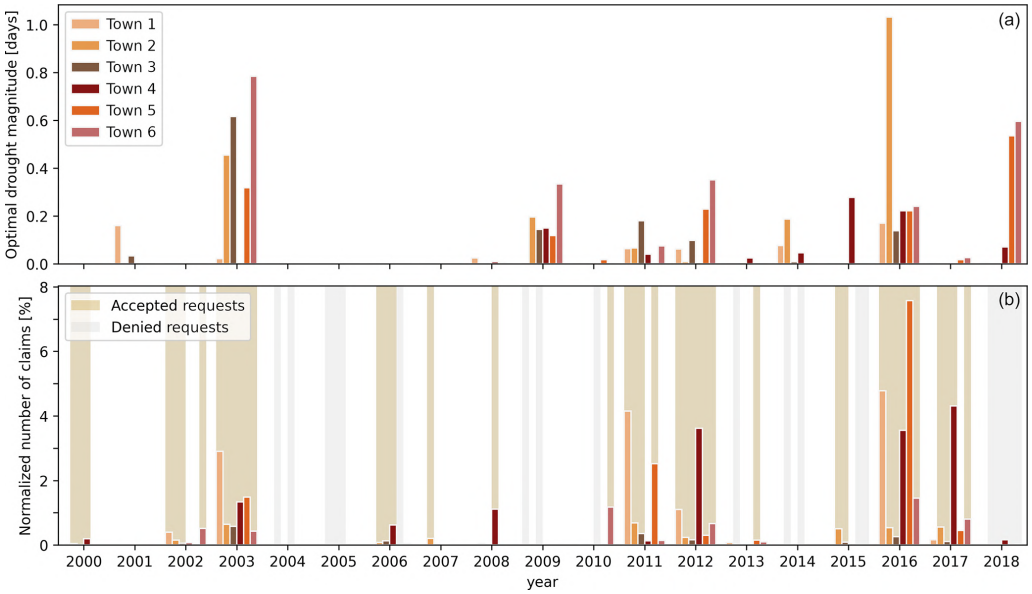


Figure 5. Optimal drought magnitude **(a)** and normalized number of claims and CatNat decrees **(b)** for the six municipalities of the validation set.

Table 3. Statistical distribution of normalized number of claims (in %) by magnitude class of calibration and validation sets. Class limits, numbers of observations, and quartiles (Q) are indicated.

Magnitude class	Range	Calibration				Validation			
		Number	Q 25 %	Q 50 %	Q 75 %	Number	Q 25 %	Q 50 %	Q 75 %
0	0	220 (58 %)	0.0	0.0	0.0	63 (55 %)	0.0	0.0	0.1
1	(0, 0.03]	54 (14 %)	0.0	0.0	0.1	18 (16 %)	0.0	0.1	1.0
2	(0.03, 0.26]	53 (14 %)	0.0	0.1	0.7	23 (20 %)	0.0	0.2	0.9
3	(0.26, 2.81]	53 (14 %)	0.1	1.0	3.8	10 (9 %)	0.0	0.5	0.6

0 and 2. Class 1 of the validation set is characterized by a high upper quartile and class 3 by a low one, in contrast to the wide range noted in class 3 for the calibration set.

4 Discussion

In the previous section, we presented an optimal drought index that we calibrated over five regional subsets and compared to independent data. The findings are discussed in the following section, along with perspectives.

4.1 How does the drought magnitude compare to existing drought indices?

This paper presents a new drought index, the yearly drought magnitude, which quantifies the yearly summed dry anomaly of an LSM-derived daily SWI. The reason for choosing a drought index based on LSM output is based on the nature of the phenomenon under study. As mentioned above, clay shrinkage is a physical response of soils to drying. Therefore, the most relevant way to monitor it is to focus on soil moisture variations. Meteorological variables alone are not able to provide a sufficient description of the hydric state of soils due to the multiple interactions between atmosphere, soil, and vegetation at the surface, hence the difference between the SPI and SSWI noted by Soubeyroux et al. (2012) and Charpentier et al. (2022). As illustrated by the indices listed by WMO and GWP (2016), the only way to assess soil moisture evolution is to weight contributions from simple water balance models to more complex LSMs.

Although both are based on an ISBA-modeled SWI, our approach has a major difference from the one developed by Vidal et al. (2010). In the latter, the magnitude index is calculated from an SSWI consisting of a monthly SWI standardized at different timescales by fitting to a given distribution over the period 1958–2008. This method was not applicable to our work because the insurance data lead us to focus on the years 2000–2018, a period which is too short for a

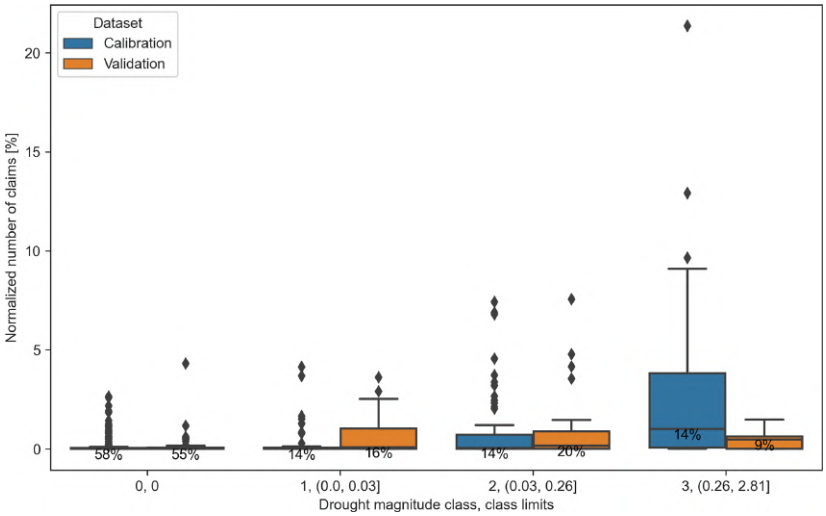


Figure 6. Distributions of the normalized number of claims per magnitude class, separating the calibration and validation sets.

proper distributional analysis of the monthly SWI. However, the standardization step is still included in the index calculation, since the thresholds are based on SWI percentiles.

4.2 Why is an interactive LAI needed?

The evaluation of 1200 possible index configurations showed the best performance of the magnitudes computed from the LAI_model simulation, model soil layer 8 (0.8–1.0 m), and extreme drought thresholds (1st to 5th percentiles).

The only difference between LAI_model and LAI_clim simulations is that the LAI varies over time in response to water and energy budgets for LAI_model, while the LAI cycle is the same each year, based on a climatology for LAI_clim. The interannual variability in the LAI depends on meteorological conditions and the effect of soil water deficit on photosynthesis and plant growth. At the same time, the simulated soil moisture depends on the amount of water extracted from the soil by plant transpiration, and thus the temporal evolution of soil moisture is linked to the LAI variable. The interactive leaf area index scheme of LAI_model is an improvement over the fixed pattern of LAI_clim because it includes vegetation feedback in response to soil drought. Drought conditions limit plant growth, LAI values, plant transpiration, and root water extraction from the soil. Figure 7 shows time series of the LAI, SWI5 (0.2–0.4 m), and SWI8 (0.8–1.0 m). Time series are shown for LAI_clim and LAI_model simulations at a single ISBA grid point located in the calibration subset corresponding to Department 31. It can be observed that larger LAI values in a simulation tend to trigger smaller SWI8 values at the end of the summer and during fall. For example, the lower soil moisture content of LAI_clim in 2003 with respect to LAI_model corresponds to higher LAI values during the warm season. This effect is much more visible for SWI8 than for SWI5. As explained by Ravina (1983), the hydraulic conductivity of the topsoil layer decreases with drying to the point where moisture in the deeper layers can remain practically unchanged. Soil moisture variations in deep layers are therefore more dependent on water uptake by roots than on diffusion processes. This explains the large impact of vegetation transpiration and the stronger correlation with the LAI.

Figure 3 shows that all annual drought indices correlate better with the number of insurance claims for deep soil layers. The best score values are obtained for SWI8. Figure 7 shows that the main difference between the two SWI variables is the presence of a high-frequency component in the SWI5 time series. The proximity of this most superficial layer to the surface, where meteorological exchanges take place, explains these short-term fluctuations, which consequently fade with depth. In other words, the SWI of a deep soil layer, which indirectly filters out high-frequency variations, reflects longer-term soil moisture trends compared to surface soil moisture. This explains the better correlation of the SWI8-based drought indices with damage data, considering that subsidence is a long-term, gradual phenomenon. For all simulations, layers, and index types, better results are observed for low thresholds. Consequently, a daily SWI below the 5th percentile of the empirical distribution is relevant for the definition of drought when monitoring subsidence.

4.3 Is the optimal drought magnitude a reliable proxy for subsidence damage?

Figure 3 shows that magnitude is a more robust drought index for representing subsidence damage than duration and sever-

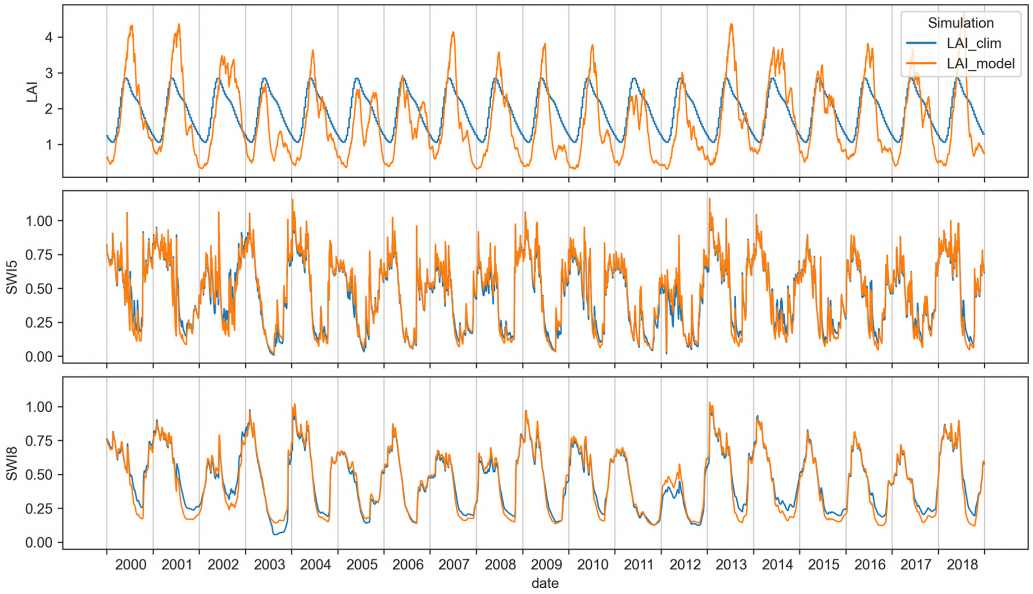


Figure 7. Time series of the LAI and SWI5 and SWI8 (corresponding to soil layers 0.2–0.4 and 0.8–1.0 m, respectively), differentiating between simulations of LAI_clim and LAI_model, for a single ISBA grid point located in the calibration subset attached to Department 31.

ity. In particular, a deterioration in the performance of the duration and severity indices is observed for increasing percentile drought thresholds and decreasing soil layer depths. The duration index loses much information for large percentile drought thresholds, hence the lower correlation values. As with severity, this index is inherently sensitive to extreme values, which is not well suited to the frequent soil moisture variations in near-surface layers that are not representative of long-term trends (as seen in Fig. 7). Nevertheless, we note very similar Kendall τ values for the three indices at percentile drought thresholds of less than the 15th percentile. Finally, magnitude is chosen as the best index because it combines both duration and severity, compensating for their respective drawbacks. Also, as shown in Fig. 4, magnitudes are spatially consistent: similar drought trends are observed for points within the same regional calibration subset. The average Kendall τ of 0.42 between magnitudes and the standardized number of claims for the calibration set shows a strong positive association, with a high statistical significance (p value ≤ 0.001).

The optimal magnitude index identifies significant drought events in 2003 and 2018 for more than half of the study areas (see Fig. 4). These two years experienced notable summer drought events, which have been widely documented in the literature (see for example Buras et al., 2020). The index has the advantages of identifying not only the occurrence of such droughts but also their absence and of being spatially consistent.

Nevertheless, we observe discrepancies: drought years with no damage recorded and, conversely, damage recorded in the absence of drought. The functioning of the French CatNat system offers possible explanations for both situations. As explained above, the coverage of subsidence damage by this system requires the publication of a decree for a given year and municipality, the issuance of which is conditional upon meeting geotechnical and meteorological criteria. The meteorological criterion is currently calculated on a seasonal basis and requires a frequency threshold to be exceeded. Losses are only documented in years with decrees, hence their strong co-occurrence noted in Figs. 4 and 5. The absence of decrees provides an explanation for the inconsistencies between positive magnitudes and zero damage. A drought event can occur without leading to a decree if the seasonal criteria are not met. On the other hand, it is possible to have low drought magnitudes and a CatNat decree in the same year. Low threshold magnitudes quantify the driest moments of the year (summer to fall droughts), while a single season only needs to be drier than usual to satisfy the CatNat meteorological criterion. For such years, the presence of claims can be explained by delays in reporting, as it is difficult for both homeowners and insurers to accurately time the occurrence of subsidence damage. This would explain why, for certain years and municipalities, claims are reported without significant magnitudes. The link to the CatNat regime is not the only factor explaining the inconsistencies between magnitudes and claims: several sources of uncertainty have not been considered in this study. These are discussed in the next section.

The analysis of the scores provides more insight, subsetwise. The high PC (>0.5) obtained for all zones shows that

the index makes a majority of correct predictions, despite the limitations mentioned above. The same conclusions can be drawn from the POD (>0.5): more than half of the loss events were detected for each subset. On the other hand, the important range of bias variation was unexpected: using thresholds defined from a given SWI frequency, the occurrence of subsidence damage is either under-predicted (subset 31) or over-predicted (subset 57). We can link this finding to the record of CatNat decrees: for the same number of years and cities, 25 decrees were issued for subset 31 and 12 for subset 57. Over-prediction is more likely for a small number of observations, and vice versa. This is not the only explanation, as the less biased subset (45, $B = 1.18$) is associated with the smallest number of decrees (6). Finally, the FAR and POFD characterize the tendency of the index to make false predictions. For all subsets, the POFD is below 0.4: fewer than 4 out of 10 predictions turn out to be wrong in the absence of subsidence claims. The FAR is also below 0.4, except for in subset 57: fewer than 4 positive claim predictions out of 10 turn out to be wrong. In both cases, a minority of predictions are wrong. The high FAR obtained for subset 57 (0.57) is related to the high bias (1.75) noted above. The scores computed over the entire calibration data set compensate for the differences between the subsets; these results, which are more robust, were expected, given the largest number of conditions.

The validation step also provides useful information to analyze how the optimal magnitude index fits the subsidence hazard. The lower Kendall τ of 0.23 indicates a weaker association of the index with subsidence damage compared to the calibration set while maintaining high statistical significance (p value less than 0.001). Once again, we can explain the disparities between drought magnitude and damage by the absence of CatNat decrees and the delays in reporting. Nevertheless, the magnitude index manages to identify most of the years with significant damage. The results obtained for the validation set from the contingency table are quite similar to, if not slightly less effective than, those obtained for the entire calibration set. The satisfactory performance of the index computed from unseen data proves that we successfully avoid overfitting, i.e., the over-adaptation of the model to specific patterns of a given data set, in our case the calibration set. In addition, the magnitude classification reveals underlying trends in both the calibration and the validation sets (Fig. 6). The similar distributions of both sets across classes indicate that the validation set, although smaller, is representative. However, these distributions are uneven, with the first class, corresponding to no damage, containing more than half of the magnitudes (Table 3). This is not surprising, since we are focusing on extreme and thus rare events. This is problematic for the validation set, as there are only a limited number of data left to divide into classes. For example, the last class in the validation set contains 9 % of the magnitudes, or 10 values, which is too small for a proper analysis. While the data in the validation set are representative of the proportion of null events, they are insufficient to provide reliable damage distributions associated with positive magnitudes. This provides an explanation for the incoherent decreasing upper-quartile observation from magnitude class 1 to magnitude class 3 in Table 3. The zoning between classes can be refined using a larger data set. While the magnitude drought index is able to identify the occurrence of drought events relevant for subsidence damage monitoring, the index alone is not able to provide accurate quantitative damage estimates at this stage. The validation step could be improved using a larger data set.

4.4 What are the sources of uncertainty?

The non-linearity of the relationship between optimal drought magnitudes and losses can be explained by several sources of uncertainty. Subsidence losses are the result of a combination of factors, not all of which have been considered at the same level of detail in this paper.

4.4.1 Representativeness of the insurance database

The number of subsidence claims used in this paper is a sample from a database compiled by CCR from insurance company submissions. It is important to keep in mind that the representativeness of these data varies across years and cities, depending on what was submitted. This is truer the smaller the city (see the number of single-family houses with moderate to high exposure to clay shrinkage per city in the Supplement).

4.4.2 Circularity

As already explained, knowledge of the history of the CatNat decrees issued is crucial for the interpretation of a subsidence damage record. However, the meteorological criterion that determines the issuance of these decrees is based on the ISBA model, just like the magnitude index we developed, although in a different configuration. Therefore, there is an indirect link between the two variables we are correlating, which induces a certain circularity in the approach. This weakness is recognized, but no valuable alternative can be considered: insurance claims are the only available evidence of subsidence, and LSMs are the tools that currently provide the most consistent estimates of soil moisture at the national scale. Nevertheless, we can consider the dependency between magnitudes and claims low, since the meteorological criterion of the CatNat decree is based on a version of the ISBA model (SWI uniforme) that is very different from the one used in this study and that has been modified several times. A possible alternative to insurance claims as a proxy for subsidence is the direct use of remotely sensed ground motion. In particular, satellite-borne interferometric synthetic aperture radar (InSAR) data can be used to infer vertical motion after appropriate processing, as done by Burnol et al. (2021). For example, the European Ground Mo-

tion Service (EGMS; Crosetto et al., 2021) has provided vertical displacements over Europe with high spatial and temporal resolution, based on the Copernicus Sentinel-1 satellites, since 2018. The main advantage of this technique is its large spatial coverage. However, the interpretation of such data is not trivial. In the case of clay shrink–swell, the vertical displacements are non-linear (seasonal periodicity), of small amplitude (a few to tens of millimeters), and spatially heterogeneous due both to the natural irregularity of clay soils and to the contrasting responses of reflectors (less movement is expected for tall buildings on pile foundations, as explained by Tzampoglou et al., 2022). It can therefore be challenging to separate a signature expansive soil signal from other phenomena such as subsidence induced by water pumping (Meisina et al., 2006). We recognize the potential of these data, but the EGMS data set was not available at the time of the study. In addition, it begins in 2018, which barely overlaps with our study period, which spans 2000 to 2018.

4.4.3 Differences in temporal and spatial resolution

Uncertainty is also introduced by the different temporal and spatial resolution of the two main data sets compared in this study, i.e., soil moisture and the number of claims.

The annual temporal resolution of this study is forced by the insurance claim data set, while the soil moisture simulations are more precise (daily). The coarse resolution of the first data set is explained by the difficulty for insurers of precisely dating the occurrence of subsidence damage due to the slow nature of the phenomenon. For this reason, this work could not be carried out at a finer temporal scale. However, there is an advantage to using daily instead of annual soil moisture data: the SWI distribution is expected to be more robust over a 19-year period due to the large number of observations. This is particularly important in the context of a changing climate. The hypothesis of a stable daily soil moisture distribution remains a source of uncertainty inherent to this work.

The two data sets also present a difference in spatial resolution: the number of claims is available at the municipal scale, while soil moisture is modeled at an 8 km $\times$ 8 km scale. These two grids are of the same order of magnitude. The overlap allows them to be paired using a criterion other than distance alone.

4.4.4 Cumulative effect of subsidence

After a particularly hot summer, the soil under perennial plants is likely to be so dry that a single wet season may not be able to compensate for the water deficit. This leads to a cumulative effect of droughts and thus soil movement (Page, 1998). For simplicity, the drought index described in this study is calculated from the SWI of a single year only. The cumulative effect is therefore neglected and is a source of uncertainty. The good agreement obtained here between drought magnitudes and normalized claims indicates that the conditions of a single year are a satisfactory enough predictor of subsidence occurrence. Taking into account the cumulative effect would improve the agreement with the numbers of claims. This step could be implemented in a damage model by, for example, weighting magnitudes by their history.

4.4.5 Zoning of clay shrink–swell exposure

Clay shrinkage depends on triggering factors, such as drought, and predisposing factors, the most important of which are the presence of clay, its nature, and its ability to shrink and swell. In order to minimize the influence of these factors, we focused our analysis on the housing stock located in moderate- to high-hazard zones of municipalities with a significant history of subsidence damage. The blending of moderate- and high-exposure zones in the housing data set is an initial source of uncertainty, as the same drought is expected to trigger different responses in soils of different susceptibility. Despite the selection of the housing data set, the uncertainty arising from the exposure to clay shrink–swell exposure map itself must be considered. The exposure map, made at a regional scale of 1 : 50 000, does not integrate very local clayey soil occurrences. The latter are frequent in relation to clay formation and depositional geological processes (BRGM, 2024). The lack of precision of the clay maps here affects the number of houses in different hazard zones used in the normalization step. The associated uncertainty is transferred to the value of the normalized number of claims. At this stage, we are not trying to make precise damage predictions but rather only identify drought years. Therefore, this source of uncertainty is not the most dominant. Moreover, this study does not take into account the intensity of the deformation induced by the shrink–swell mechanism, which is a significant component explaining the deformations of the structures induced by variations in water content.

4.4.6 Local factors

Local factors other than the sporadic presence of swelling clay play a significant role in the subsidence phenomenon:

- The presence of vegetation (especially deciduous trees) around the building is critical, as roots are known to locally intensify soil drying during droughts (Freeman et al., 1992; Hawkins, 2013; Page, 1998).

- The characteristics of the building itself (e.g., building type, number of floors, presence of a basement) determine its vulnerability to subsidence damage. This is important in understanding the causes of a subsidence claim (Page, 1998).

- Various elements of the building environment, such as its orientation, the slope of the land, or the presence of pavement, can limit or enhance soil moisture vari-

ations and thus seasonal ground movements (Cooling and Ward, 1948; Page, 1998).

None of these factors were considered in this work because this information was not available for integrated municipal-scale data. The specific effect of trees on soil moisture fluctuations could be investigated by modifying the ISBA model configuration. For all the other local factors, evaluation of their influence on subsidence is only possible through local analysis using instrumented sites.

4.5 What are the possible applications for this work, in and outside of France?

The main result of this research is a new index, the yearly drought magnitude, specifically adapted to the problem of clay-shrinkage-induced subsidence. It is computed from the SWI output of the ISBA LSM and is calibrated and validated with insurance data. As shown above, this index is relevant for the identification of subsidence-induced drought events. It faces several limitations due to the various sources of uncertainty.

The research was carried out on a sample of 20 cities, but the index can be calculated for the whole country. A first possible application is the use of the magnitude of drought itself to monitor conditions likely to cause subsidence over time, as is currently done by French institutions. Another possible application, this time for the insurance industry, would be to use the index as an improved predictor in subsidence loss models, such as the one developed by Charpentier et al. (2022).

As several regions worldwide face similar clay-shrinkage-induced subsidence problems (see for example MacQueen et al., 2023, and Mostafiz et al., 2021, for the UK and the USA, respectively), the methodology can be adapted to other countries. The SURFEX modeling platform, in which ISBA is implemented, can be applied anywhere, provided that high-resolution atmospheric forcing is available.

5 Conclusion

In this paper, we propose a new annual drought index, the optimal drought magnitude, tailored to the subsidence hazard. Our work can be summarized as follows.

- The proposed optimal drought magnitude is the mean of five annual integrals of daily soil wetness index (SWI) values under several thresholds (1st to 5th percentiles of the empirical SWI distribution). It is based on the eighth model layer (depth of 0.8 to 1.0 m) of the ISBA land surface model in a configuration that allows interactive LAI simulation. In this configuration, the average Kendall τ between the drought index and the normalized number of claims is equal to 0.42 with a high statistical significance (p value ≤ 0.001). The validation step indicates that we avoid overfitting.

- The optimal drought magnitude identifies events that are likely to generate subsidence claims. It is spatially consistent. Differences with subsidence damage claims can be explained by the way claims are collected and by the lack of information on local conditions.

- The optimal drought magnitude benefits from recent advances in land surface modeling (multilayer, interactive LAI). It could be used as a predictor in subsidence loss models to provide more accurate cost estimates. As this index is based on ISBA simulations, future subsidence damage risks could be assessed by forcing ISBA through downscaled climate model simulations.

Future developments will focus on reducing uncertainties by working at finer spatial scales and by investigating the specific effects of different vegetation types.

Code availability. The analysis was carried out with Python codes that can be made available upon request. SURFEX can be downloaded freely at http://www.umr-cnrm.fr/surfex/data/OPEN-SURFEX/open_surfex_v8_1_20210914.tar.gz (CNRM, 2016). It is provided under a CECILL-C license (French equivalent to the LGPL license).

Data availability. The data presented in the figures are available online at https://doi.org/10.6084/m9.figshare.23559507 (Barthelemy, 2023). Geographic distribution of exposed single-family houses can be downloaded freely at https://www.statistiques.developpement-durable.gouv.fr/nouveau-zonage-dexposition-au-retrait-gonflement-des-argiles-plus-de-104-millions-de-maisons (MTES, 2021b).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/nhess-24-999-2024-supplement.

Author contributions. SBa and JCC designed the experiments. SBa performed the investigation, did the formal analysis, and wrote the paper. BB processed the LAI_model data set. All co-authors participated in the interpretation of the results and the revision of the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Special issue statement. This article is part of the special issue "Drought, society, and ecosystems (NHESS/BG/GC/HESS inter-journal SI)". It is a result of the Panta Rhei Drought in the Anthropocene workshop 2022, Uppsala, Sweden, 29–30 August 2022.

Acknowledgements. The authors would like to thank the operational services of Météo-France (DCSC), especially Jean-Marie Willemet, for providing the LAI_clim soil moisture data set, as well as Jean-Philippe Naulin (CCR) and the three anonymous reviewers for helpful comments. This paper is dedicated to the memory of David Moncoulon.

Financial support. The PhD thesis work of Sophie Barthelemy was co-funded by BRGM (Bureau de Recherches Géologiques et Minières), CCR (Caisse Centrale de Réassurance), and Météo-France.

Review statement. This paper was edited by Floris van Ogtrop and reviewed by three anonymous referees.

References

Barthelemy, S.: A new drought index fitted to clay shrinkage induced subsidence over France: benefits of interactive leaf area index, Barthelemy et al. 2023 – Figure data, Figshare [data set], https://doi.org/10.6084/m9.figshare.23559507.v1, 2023.

Boone, A., Calvet, J. C., and Noilhan, J.: Inclusion of a third soil layer in a land surface scheme using the force-restore method, J. Appl. Meteorol., 38, 1611–1630, https://doi.org/10.1175/1520-0450(1999)038<1611:IOATSL>2.0.CO;2, 1999.

Boone, A., Masson, V., Meyers, T., and Noilhan, J.: The Influence of the Inclusion of Soil Freezing on Simulations by a Soil–Vegetation–Atmosphere Transfer Scheme, J. Appl. Meteorol., 39, 1544–1569, https://doi.org/10.1175/1520-0450(2000)039<1544:TIOTIO>2.0.CO;2, 2000.

BRGM: M'informer sur le retrait-gonflement des argiles – Exposition du territoire au phénomène, https://www.georisques.gouv.fr/articles-risques/retrait-gonflement-des-argiles/exposition-du-territoire-au-phenomene, last access: 11 March 2024.

Bronswijk, J. J. B.: Prediction of actual cracking and subsidence in clay soils, Soil Sci., 148, 87–93, https://doi.org/10.1097/00010694-198908000-00002, 1989.

Brut, A., Rüdiger, C., Lafont, S., Roujean, J.-L., Calvet, J.-C., Jarlan, L., Gibelin, A.-L., Albergel, C., Le Moigne, P., Soussana, J.-F., Klumpp, K., Guyon, D., Wigneron, J.-P., and Ceschia, E.: Modelling LAI at a regional scale with ISBA-A-gs: comparison with satellite-derived LAI over southwestern France, Biogeosciences, 6, 1389–1404, https://doi.org/10.5194/bg-6-1389-2009, 2009.

Buras, A., Rammig, A., and Zang, C. S.: Quantifying impacts of the 2018 drought on European ecosystems in comparison to 2003, Biogeosciences, 17, 1655–1672, https://doi.org/10.5194/bg-17-1655-2020, 2020.

Burnol, A., Foumelis, M., Gourdier, S., Deparis, J., and Raucoules, D.: Monitoring of expansive clays over drought-rewetting cycles using satellite remote sensing, Atmosphere, 12, 1262, https://doi.org/10.3390/atmos12101262, 2021.

Calvet, J.-C.: Investigating soil and atmospheric plant water stress using physiological and micrometeorological data, Agric. For. Meteorol., 103, 229–247, https://doi.org/10.1016/S0168-1923(00)00130-1, 2000.

Calvet, J.-C. and Soussana, J.-F.: Modelling CO_2-enrichment effects using an interactive vegetation SVAT scheme, Agric. For. Meteorol., 108, 129–152, https://doi.org/10.1016/S0168-1923(01)00235-0, 2001.

Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A., and Wigneron, J.-P.: An interactive vegetation SVAT model tested against data from six contrasting sites, Agric. For. Meteorol., 92, 73–95, https://doi.org/10.1016/S0168-1923(98)00091-4, 1998.

Calvet, J.-C., Rivalland, V., Picon-Cochard, C., and Guehl, J.-M.: Modelling forest transpiration and CO_2 fluxes – Response to soil moisture stress, Agric. For. Meteorol., 124, 143–156, https://doi.org/10.1016/j.agrformet.2004.01.007, 2004.

CCR: Les Catastrophes naturelles en France, Bilan 1982–2021, https://catastrophes-naturelles.ccr.fr/documents/148935/368920/BILAN+Cat+Nat+1982-2021.pdf/cb6ffbc8-f096-e8a6-5f71-3fe68ee8d78e?t=1675351436030 (last access 11 March 2023), 2021.

CCR: Rapport au Ministre de l'économie, des finances, et de la souveraineté industrielle et numérique sur le régime d'indemnisation des catastrophes naturelles, https://www.ccr.fr/documents/35794/1130659/Rapport+au+ministre+-+Decembre+2022.pdf/c6edbb8e-e5ba-e5d6-2d30-d2750c229086?t=1675073436815 (last access: 11 March 2024), 2022.

CCR and Météo-France: Conséquences du changement climatique sur le coût des catastrophes naturelles en France à horizon 2050, 32 pp., https://www.ccr.fr/documents/35794/35836/Etude+Climatique+2018+version+complete.pdf/6a7b6120-7050-ff2e-4aa9-89e80c1e30f2?t=1536662736000 (last access: 11 March 2024), 2018.

Charpentier, A., James, M., and Ali, H.: Predicting drought and subsidence risks in France, Nat. Hazards Earth Syst. Sci., 22, 2401–2418, https://doi.org/10.5194/nhess-22-2401-2022, 2022.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water Resour. Res., 14, 601–604, https://doi.org/10.1029/WR014i004p00601, 1978.

CNRM: SURFEX code, CNRM [code], http://www.umr-cnrm.fr/surfex/data/OPEN-SURFEX/open_surfex_v8_1_20210914.tar.gz (last access: 11 March 2024), 2016.

Cojean, R.: Introduction, Rev. Française Géotechnique, 120–121, 7–9, https://doi.org/10.1051/geotech/2007120007, 2007.

Cooling, L. F. and Ward, W. H.: Some examples of foundation movements due to causes other than structural loads, in: Proceedings 2nd international conference on soil mechanics and foundation engineering, Rotterdam, the Netherlands, 21 to 30 June 1948, 162–167, https://www.issmge.org/publications/publication/some-examples-of-foundation-movements-due-to-causes-other-than-structural-loads (last access: 11 March 2024), 1948.

Corti, T., Muccione, V., Köllner-Heck, P., Bresch, D., and Seneviratne, S. I.: Simulating past droughts and associated building

damages in France, Hydrol. Earth Syst. Sci., 13, 1739–1747, https://doi.org/10.5194/hess-13-1739-2009, 2009.

Corti, T., Wüest, M., Bresch, D., and Seneviratne, S. I.: Drought-induced building damages from simulations at regional scale, Nat. Hazards Earth Syst. Sci., 11, 3335–3342, https://doi.org/10.5194/nhess-11-3335-2011, 2011.

Covéa and RiskWeatherTech: Changement climatique et Assurance: Quelles conséquences sur la sinistralité à horizon 2050?, https://www.covea.com/sites/default/files/2022-02/202202_Livre_Blanc_Cov%C3%A9a_Risques_Climatiques.pdf (last access: 11 March 2024), 2022.

Crilly, M.: Analysis of a database of subsidence damage, Struct. Surv., 19, 7–14, https://doi.org/10.1108/02630800110384185, 2001.

Crosetto, M., Solari, L., Balasis-Levinsen, J., Bateson, L., Casagli, N., Frei, M., Oyen, A., Moldestad, D. A., and Mróz, M.: Deformation Monitoring at European Scale: The Copernicus Ground Motion Service, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B3-2021, 141–146, https://doi.org/10.5194/isprs-archives-XLIII-B3-2021-141-2021, 2021.

Decharme, B., Boone, A., Delire, C., and Noilhan, J.: Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions, J. Geophys. Res.-Atmos., 116, D20126, https://doi.org/10.1029/2011JD016002, 2011.

Decharme, B., Martin, E., and Faroux, S.: Reconciling soil thermal and hydrological lower boundary conditions in land surface models, J. Geophys. Res.-Atmos., 118, 7819–7834, https://doi.org/10.1002/jgrd.50631, 2013.

D'Odorico, P., Ridolfi, L., Porporato, A., and Rodriguez-Iturbe, I.: Preferential states of seasonal soil moisture: The impact of climate fluctuations, Water Resour. Res., 36, 2209–2219, https://doi.org/10.1029/2000WR900103, 2000.

Doornkamp, J. C.: Clay Shrinkage Induced Subsidence, Geogr. J., 159, 196–202, https://doi.org/10.2307/3451410, 1993.

Dracup, J. A., Lee, K. S., and Paulson, E. G.: On the definition of droughts, Water Resour. Res., 16, 297–302, https://doi.org/10.1029/WR016i002p00297, 1980.

Durand, Y., Brun, E., Merindol, L., Guyomarc'h, G., Lesaffre, B., and Martin, E.: A meteorological estimation of relevant parameters for snow models, Ann. Glaciol., 18, 65–71, https://doi.org/10.3189/S0260305500011277, 1993.

Ecoto, G. and Chambaz, A.: Forecasting the cost of drought events in France by super learning, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2022-541, 2022.

Faroux, S., Kaptué Tchuenté, A. T., Roujean, J.-L., Masson, V., Martin, E., and Le Moigne, P.: ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, Geosci. Model Dev., 6, 563–582, https://doi.org/10.5194/gmd-6-563-2013, 2013.

Ford, T. W., Wang, Q., and Quiring, S. M.: The observation record length necessary to generate robust soil moisture percentiles, J. Appl. Meteorol. Climatol., 55, 2131–2149, https://doi.org/10.1175/JAMC-D-16-0143.1, 2016.

Freeman, T. J., Burford, D., and Crilly, M. S.: Seasonal foundation movements in London Clay, in: Proceedings of the 4th International Conference on Ground Movements and Structures, Cardiff, UK, 8–11 July 1991, 485–501, 1992.

Gibelin, A.-L., Calvet, J.-C., Roujean, J.-L., Jarlan, L., and Los, S. O.: Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products, J. Geophys. Res.-Atmos., 111, D18102, https://doi.org/10.1029/2005JD006691, 2006.

Gourdier, S. and Plat, E.: Impact du changement climatique sur la sinistralité due au retrait-gonflement des argiles, in: Journées Nationales de Géotechnique et Géologie de l'Ingénieur (JNGG), https://hal-brgm.archives-ouvertes.fr/hal-01768395/document (last access: 11 March 2024), 2018.

Habets, F., Boone, A., Champeaux, J.-L., Etchevers, P., Franchistéguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., Noilhan, J., Seguí, P. Q., Rousset-Regimbeau, F., and Viennot, P.: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, J. Geophys. Res.-Atmos., 113, D06113, https://doi.org/10.1029/2007JD008548, 2008.

Harrison, A. M., Plim, J. F. M., Harrison, M., Jones, L. D., and Culshaw, M. G.: The relationship between shrink-swell occurrence and climate in south-east England, Proc. Geol. Assoc., 123, 556–575, https://doi.org/10.1016/j.pgeola.2012.05.002, 2012.

Hawkins, A. B.: Some engineering geological effects of drought: Examples from the UK, Bull. Eng. Geol. Environ., 72, 37–59, https://doi.org/10.1007/s10064-013-0458-7, 2013.

Heranval, A., Lopez, O., and Thomas, M.: Application of machine learning methods to predict drought cost in France, Eur. Actuar. J., 13, 731–753, https://doi.org/10.1007/s13385-022-00327-z, 2022.

Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., and Wavresky, P.: Les types de climats en France, une construction spatiale, CyberGeo, 2010, 2–22, https://doi.org/10.4000/cybergeo.23155, 2010.

Le Moigne, P., Besson, F., Martin, E., Boé, J., Boone, A., Decharme, B., Etchevers, P., Faroux, S., Habets, F., Lafaysse, M., Leroux, D., and Rousset-Regimbeau, F.: The latest improvements with SURFEX v8.0 of the Safran–Isba–Modcou hydrometeorological model for France, Geosci. Model Dev., 13, 3925–3946, https://doi.org/10.5194/gmd-13-3925-2020, 2020.

MacQueen, M., Lawson, M., and Ding, W.-N.: The 2018–2019 UK residential dwellings clay shrinkage subsidence event, Int. J. Build. Pathol. Adapt., https://doi.org/10.1108/IJBPA-06-2022-0091, 2023.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev., 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, 2013.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, in: Eight Conference on Applied Climatology, Anaheim, USA, 17–22 January 1993.

Meisina, C., Zucca, F., Fossati, D., Ceriani, M., and Allievi, J.: Ground deformation monitoring by using the Permanent Scatterers Technique: The example of the Oltrepo Pavese (Lombardia, Italy), Eng. Geol., 88, 240–259, https://doi.org/10.1016/j.enggeo.2006.09.010, 2006.

Mostafiz, R. Bin, Friedland, C. J., Rohli, R. V., Bushra, N., and Held, C. L.: Property Risk Assessment for Expansive Soils in Louisiana, Front. Built Environ., 7, 1–10, https://doi.org/10.3389/fbuil.2021.754761, 2021.

MTES: Cartographie de l'exposition des maisons individuelles au retrait-gonflement des argiles, https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2021-06/note_methode_croisement_retrait_gonflement_argiles_juin2021v3.pdf (last access: 11 March 2024), 2021a.

MTES: Nouveau zonage d'exposition au retrait-gonflement des argiles: plus de 10,4 millions de maisons individuelles potentiellement très exposées, https://www.statistiques.developpement-durable.gouv.fr/nouveau-zonage-dexposition-au-retrait-gonflement-des-argiles-plus-de-104-millions-de-maisons (last access: 19 March 2024), 2021b.

Nachtergaele, F., Velthuizen, H., Verelst, L., and Wiberg, D.: Harmonized World Soil Database Version 1.2 [data set], https://esdac.jrc.ec.europa.eu/ESDB_Archive/Soil_Data/Docs_GlobalData/Harmonized_World_Soi_Database_v1.2.pdf (last access: 11 March 2024), 2012.

Noilhan, J. and Lacarrère, P.: GCM Grid-Scale Evaporation from Mesoscale Modeling, J. Climate, 8, 206–223, https://doi.org/10.1175/1520-0442(1995)008<0206:GGSEFM>2.0.CO;2, 1995.

Noilhan, J. and Mahfouf, J. F.: The ISBA land surface parameterisation scheme, Glob. Planet. Change, 13, 145–159, https://doi.org/10.1016/0921-8181(95)00043-7, 1996.

Noilhan, J. and Planton, S.: A Simple Parameterization of Land Surface Processes for Meteorological Models, Mon. Weather Rev., 117, 536–549, https://doi.org/10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2, 1989.

Page, R. C. J.: Reducing the cost of subsidence damage despite global warming, Struct. Surv., 16, 67–75, https://doi.org/10.1108/02630809810219641, 1998.

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France, J. Appl. Meteorol. Climatol., 47, 92–107, https://doi.org/10.1175/2007JAMC1636.1, 2008.

Ravina, I.: The influence of vegetation on moisture and volume changes, Géotechnique, 33, 151–157, https://doi.org/10.1680/geot.1983.33.2.151, 1983.

Salagnac, J.-L.: Lessons from the 2003 heat wave: A French perspective, Build. Res. Inf., 35, 450–457, https://doi.org/10.1080/09613210601056554, 2007.

Soubeyroux, J.-M., Kitova, N., Blanchard, M., Vidal, J.-P., Martin, É., and Dandin, P.: Sécheresses des sols en France et changement climatique: Résultats et applications du projet ClimSec, La Météorologie, 8, 21–30, https://doi.org/10.4267/2042/47512, 2012.

Szczypta, C., Calvet, J.-C., Maignan, F., Dorigo, W., Baret, F., and Ciais, P.: Suitability of modelled and remotely sensed essential climate variables for monitoring Euro-Mediterranean droughts, Geosci. Model Dev., 7, 931–946, https://doi.org/10.5194/gmd-7-931-2014, 2014.

Tessier, D., Coquet, Y., Lefèvre, Y., and Bréda, N.: Rôle de la végétation dans les processus de propagation de la sécheresse dans les sols argileux, Rev. Française Géotechnique, 120–121, 35–43, https://doi.org/10.1051/geotech/2007120035, 2007.

Toreti, A., Bavera, D., Acosta Navarro, J., Cammalleri, C., de Jager, A., Di Ciollo, C. ., Hrast Essenfelder, A., Maetens, W., Masante, D., Magni, D., Mazzeschi, M., and Spinoni, J.: Drought in Europe – August 2022, 19 pp., https://doi.org/10.2760/264241, 2022.

Tzampoglou, P., Loukidis, D., and Koulermou, N.: Seasonal Ground Movement Due to Swelling/Shrinkage of Nicosia Marl, Remote Sens., 14, 1440, https://doi.org/10.3390/rs14061440, 2022.

Vidal, J.-P., Martin, E., Franchistéguy, L., Habets, F., Soubeyroux, J.-M., Blanchard, M., and Baillon, M.: Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite, Hydrol. Earth Syst. Sci., 14, 459–478, https://doi.org/10.5194/hess-14-459-2010, 2010.

Vidal, J.-P., Martin, E., Kitova, N., Najac, J., and Soubeyroux, J.-M.: Evolution of spatio-temporal drought characteristics: validation, projections and effect of adaptation scenarios, Hydrol. Earth Syst. Sci., 16, 2935–2955, https://doi.org/10.5194/hess-16-2935-2012, 2012.

Wilks, D. S.: Statistical methods in the atmospheric sciences, 2nd edn., Int. Geophys. Series, Academic Press, Elsevier, 91, 627 pp., ISBN 13: 978-0-12-751966-1, 2006.

WMO and GWP: Handbook of Drought Indicators and Indices, in: Integrated Drought Management Programme (IDMP), Integrated Drought Management Tools and Guidelines Series 2, Geneva, edited by: Svoboda, M. and Fuchs, B. A., https://www.droughtmanagement.info/literature/GWP_Handbook_of_Drought_Indicators_and_Indices_2016.pdf (last access: 13 July 2023), 2016.

3.3 Comparison of ISBA patches

After obtaining these initial results, the experiment was repeated on the ISBA A-gs NIT simulation separately for each patch. We restricted ourselves to the patches corresponding to bare soil (1), rocks (2), deciduous broadleaf trees (4), C_3 crops (7), C_4 crops (8) and C_3 grassland (10) as they are the only ones existing throughout the whole country (see Figure 2.3), and therefore are potentially viable for operational use. The resulting average Kendall tau between YDMI and claims is plotted in Figure 3.1.

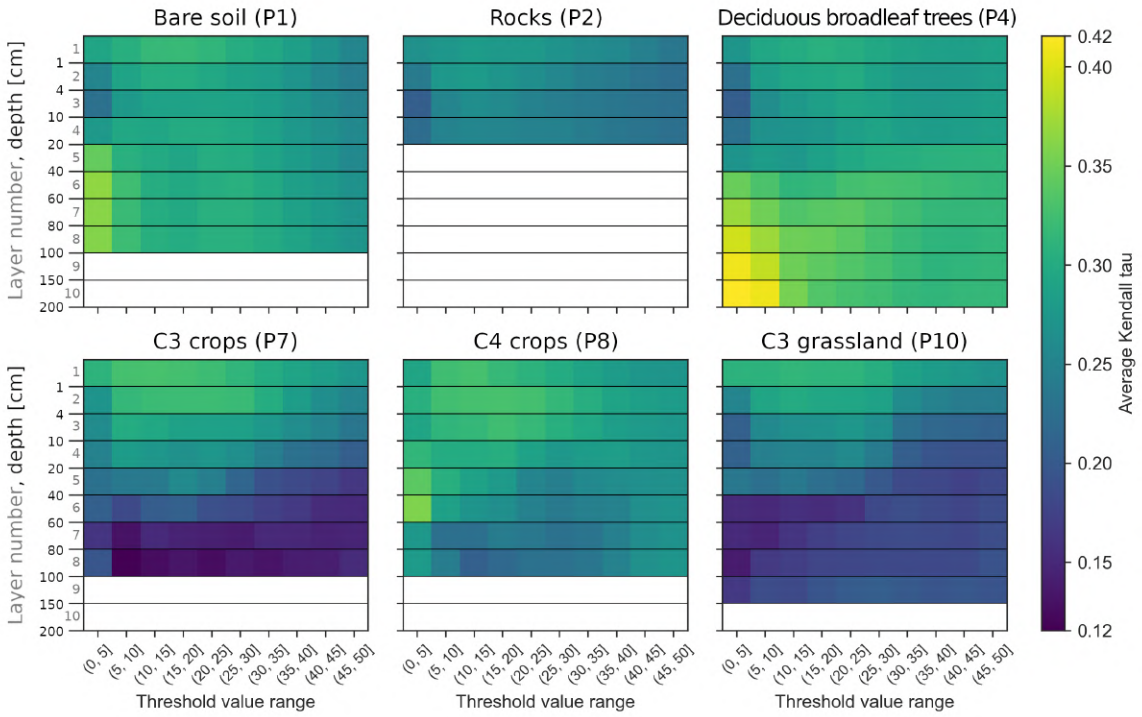


Figure 3.1: Average Kendall tau obtained between number of claims and YDMI for the different ISBA patches

A highest average Kendall tau of 0.42 is obtained when the YDMI is based on the deciduous broadleaf tree patch, deepest layers and lowest threshold values. Figure 3.2 provides an example of LAI and SWI variations for the 8^{th} model layer for each patch, to improve our understanding of correlation results. The rocks patch is absent from the comparison as it does not exist at such depths.

The first plot shows that the LAI has an annual periodicity, with a maximum reached in summer. These variations are due to the vegetation cycle. For example, in the case of trees, LAI decreases and increases reflect to the senescence and onset of leaves at the beginning and end of winter. Decreases in midsummer (observed for instance in 2005 for all patches) correspond to vegetation wilting. Slight variations in the pattern can be observed depending on the patch: for example, C_3 crops (patch 7) reach their peak earlier in the year than other types, and wilt more often in the summer heat. The year 2003, notable for its drought [Salagnac, 2007], stands out as wilting affects all patches and is not followed by regrowth.

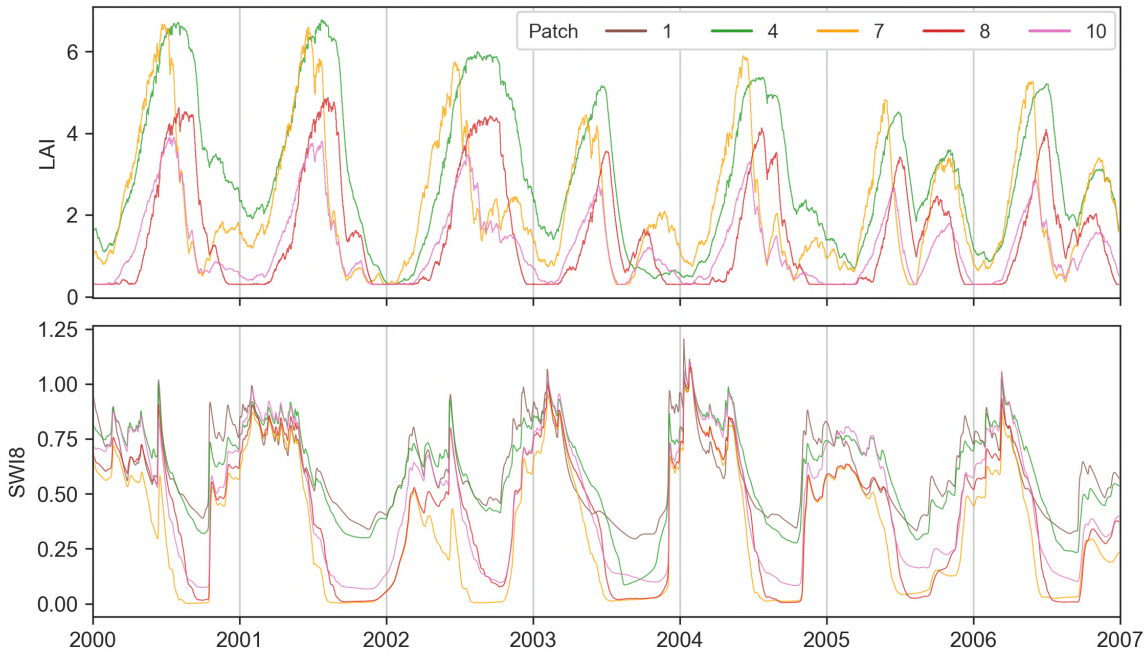


Figure 3.2: LAI (top) and SWI (bottom) of layer 8 (80 cm to 1 m) for different patches, between 2000 and 2006, for the ISBA point located at lon=1.49, lat=43.64 (Haute-Garonne). Patches 1, 4, 7, 8 and 10 correspond to bare soil, deciduous broadleaf trees, C_3 crops, C_4 crops and C_3 grassland, respectively.

The second plot separates herbaceous vegetation (crops and grassland) from bare soil (Patch 1) and trees (Patch 4) as the formers often reach lower moisture contents in the summer. This is explained by the increased evaporative effect of vegetation: it is less pronounced for trees, which are show a better adaptation to drought through their protection mechanisms. C_3 crops are particularly affected by drought, due to their early LAI peaks: the wilting point (SWI = 0) conditions prevail for a long time each year, hence the differences in YDMI and thus in Kendall tau.

The surroundings of a suburban, detached house frequently consist of a garden, with a short cut lawn and trees. Trees in the vicinity of the building often coincide with damage occurrence (see section 1.2.2.3), which can explain why they obtained the highest correlation score.

3.4 Comparison of operational ISBA versions

The two optimal configurations obtained previously (aggregated and by patch) were eventually compared, applying the same method, to the operational products SIM2 and SWI uniforme. The characteristics of these two ISBA simulations are reminded in Table 3.1. SIM2 and SWI uniforme are based on the moisture content averaged over all layers. The SIM2 outputs by layers correspond to the AST simulations of Barthelemy et al. [2024a].

Resulting average Kendall tau are presented in Figure 3.3. We conserved the same colorbar range than previous figures for consistency.

Table 3.1: Characteristics of ISBA operational simulations

Name	Soil scheme	Layers	Texture	Vegetation Scheme	Vegetation
SWI Uniforme	3L	3	58% clay 12% sand	A-gs AST	C_3 grassland
SIM2	DIF	10	HWSD	A-gs AST	Aggregated by patch

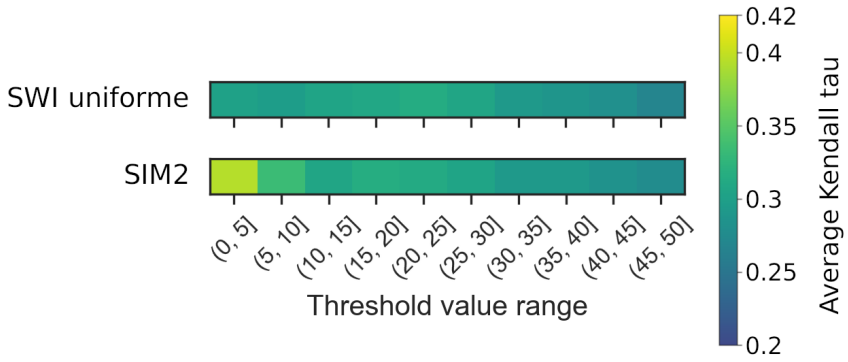


Figure 3.3: Average Kendall tau obtained between number of claims and YDMI for different ISBA versions

The Kendall tau is lower than those obtained previously for the NIT vegetation scheme, aggregated or for the tree patch, which was equal to 0.42. For comparable threshold values, SWI uniforme obtains the lowest scores, which can be attributed to the simplistic 3L soil scheme. This result clearly demonstrates that there is a difference between operational soil moisture products and those developed in research. Updates to the ISBA version could have the potential to enhance the reliability of operational applications.

3.5 Optimal index configuration

While similar results are obtained, basing an optimal index on patches rather than aggregated simulations is more suitable for the clay shrinkage-induced subsidence problem, which primarily affects detached houses in a suburban context. Furthermore, basing the optimal YDMI on layer 8 is more consistent with previous results than basing it on layer 10. As shown in Figure B.1, the difference between YDMI based on layer 8 and layer 10 is rather small.

We define an **optimal YDMI**, based on the **deciduous broadleaf tree patch** (patch 4), the ISBA **layer 8** (between 80 cm and 1 m), and drought threshold value between the 1^{st} **and** 3^{rd} **percentiles** of daily soil moisture.

3.6 Conclusion

In summary, the **YDMI** allows quantification of the conditions that trigger clay shrinkage on an annual basis. The methodology is based on statistical comparison with insurance claims data and was proved robust.

The best results (Kendall tau of 0.42 with high statistical significance) are obtained when the YDMI is based on the 8^{th} model layer of ISBA (depth between 80 and 100 cm), the deciduous broadleaf tree patch, and a drought definition corresponding to a daily SWI below the 1^{st} to 5^{th} distribution percentiles. Deep model layers such as this one are less influenced by meteorological variability than shallow ones and therefore allow focus on long-term trends, in agreement with the low kinetics of clay shrinkage. The tree patch is adapted as vegetation growing near houses has been shown to be an aggravating factor. Finally, the drought definition shows that the most intense soil moisture conditions encountered in a year are the most relevant for understanding subsidence induced by clay shrinkage.

In addition, we have shown that an interactive representation of vegetation in ISBA significantly improves the results. Also, the research version of ISBA gives better results than the operational versions, which can be explained by its more advanced options.

In the following chapter 4, ISBA simulations and associated YDMI will be compared with data from the instrumented site of BRGM in the Chaingy municipality. Gaining insight into how ISBA compares to field data will make it possible to assess the uncertainties inherent to modeling.

Chapter 4

Comparison to data from the Chaingy instrumented site

Contents

4.1 Introduction

The work presented in chapter 3 consisted in extracting the most relevant drought information from ISBA soil moisture simulations for understanding the occurrence of clay shrinkage induced subsidence damage. It has led to the creation of the YDMI.

In the continuity of these results, this chapter explores uncertainties linked to ISBA by comparing the optimal simulations by patch (simulation V8.1_DIF_NIT_P10) and derived YDMI to the data acquired on the Chaingy instrumented site (presented in section 2.4). In particular, differences induced by the contrasting spatial scales will be investigated.

4.2 Compared data

As detailed in section 2.4, the instrumented site in the Chaingy municipality includes 10 Frequency Domain Reflectometry (FDR) soil moisture probes. We identify 7 ISBA grid cells, of resolution 8 km × 8 km, (shown in Figure 2.13 (B)) whose centroids are less than 8 km away from the instrumented site. These grid cells are used for comparison. Only the years 2017 to 2022 are included, due to the large amount of missing data in 2016.

Soil Texture

The soil texture of the sandy clay layer at the Chaingy site based on the geotechnical investigations is compared with that of the 7 ISBA cells in Figure 4.1. The Chaingy sandy clay is

classified as a sandy clay loam. A single ISBA grid cell belongs to the same category, while the rest ranges from sandy loam to clay loam.

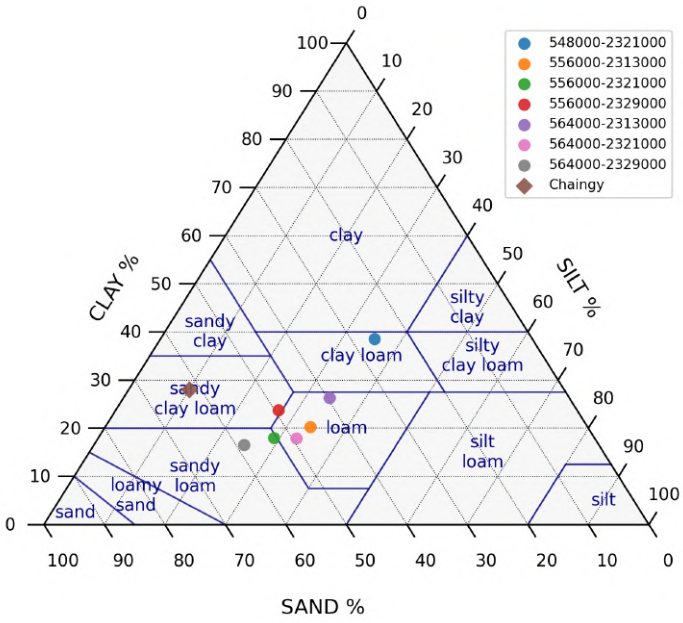


Figure 4.1: Soil texture of the Chaingy sandy clay (brown diamond) and the 7 ISBA grid cells located less than 8 km away from the site (colored circles)

Vegetation

The Chaingy site is covered by a lawn (Figure 2.14). As the 7 ISBA cells include the six main patches defined over France, we will consider the C_3 grassland patch (number 10) for comparison.

Depth

The soil moisture probes of the Chaingy site are planted 120 cm deep. They will be compared to the ISBA layer 9, extending from 100 to 150 cm.

4.3 Soil moisture variations

Soil moisture values of both the Chaingy site and the ISBA V8.1_DIF_NIT_P10 simulation were averaged by day, their initial time steps being 30 minutes and 1 hour, respectively.

As absolute soil moisture content values are texture-dependent, they require a normalization in order to be compared. To do so, we rescaled volumetric water content W between minimum and maximum (Equation 4.1), as done by Etchevers et al. [2022] for similar purposes. The resulting variable is called W_{norm}.

$$W_{norm} = \frac{W - W_{min}}{W_{max} - W_{min}} \tag{4.1}$$

W_{norm} variations between 2017 and 2022 on the Chaingy site and for the 7 ISBA grid cells are plotted in Figure 4.2. Their comparison provides valuable insight.

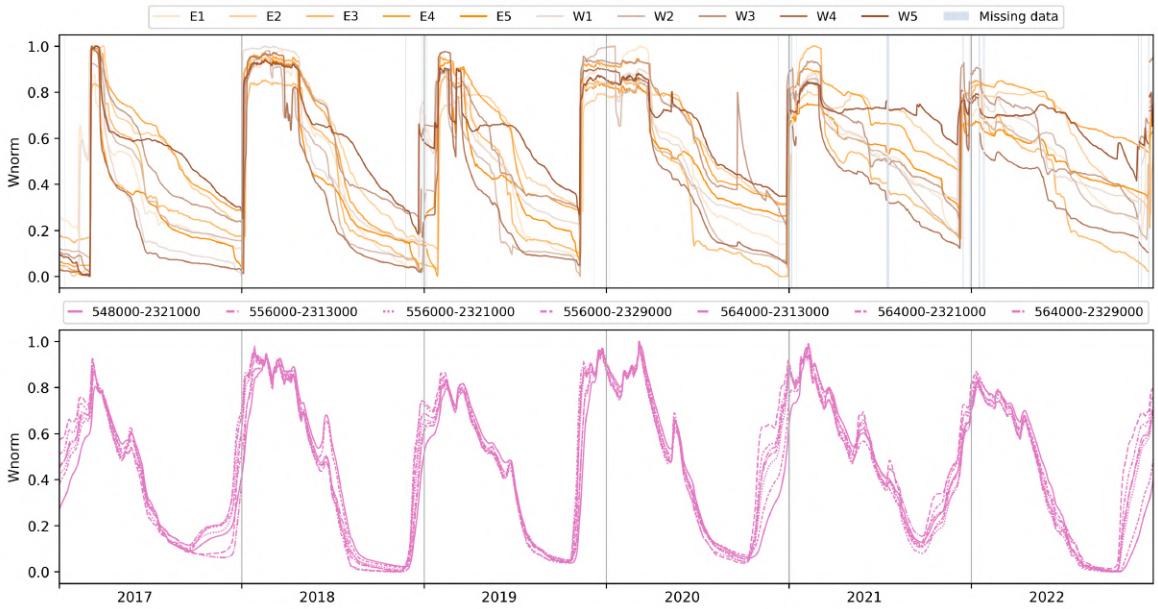


Figure 4.2: W_{norm} for the Chaingy site (top) and the 7 ISBA simulations (C_3 grassland, 100 to 150 cm) (bottom) from 2017 to 2022

First, we note a greater dispersion between curves associated to the Chaingy site than between neighboring ISBA grid cells, despite the fact that the Chaingy probes are separated by no more than 10 meters, whereas the grid cells are separated by 10 kilometers or more. This observation indicates that in reality, volumetric soil moisture exhibits a considerable variability at the local scale, which is not accounted for by ISBA. This discrepancy cannot be attributed to heterogeneous soil textures, as the curves were normalized beforehand. Also, soil rehydration measured at the Chaingy site always occurs at the same time for all probes except in 2017 and 2019. This can be attributed to local heterogeneities varying over time, for example the opening of shrinkage cracks that facilitate water flows.

The difference in drying and rehydration kinetics measured on the Chaingy site was raised previously in section 2.4. We also report differences between model and observation: drying is more gradual and rehydration more abrupt with a greater time lag on the field than in the model. These differences could be explained by the definition of the soil hydraulic properties in ISBA. They are approximated solely on the basis of texture (see section 2.1.1.2) and considered homogeneous along the vertical. This representation is not adapted to the Chaingy site, as the core samples revealed the presence of two different layers (backfilled topsoil and clay) above 120 cm. In line with this point, Sobaga et al. [2023] have shown that accounting for the vertical heterogeneity of the soil profile increased the agreement with their field lysimeter data.

Despite the differences enumerated above, the normalized soil moisture curves of the Chaingy site and ISBA grid cells exhibit analogous variations, and are consistent towards interannual trends. For example, both observation and simulation indicate that the spring of 2017 was

drier than that of 2018. The average Pearson correlation coefficient between the two is 0.74 is, with a root mean square error (RMSE) of 0.22 and a bias of -0.06 (slight tendency of ISBA to underestimate normalized soil moistures).

4.4 Yearly Drought Magnitude Index

The comparison of Chaingy and ISBA-modeled soil moisture was further refined by computing a Yearly Drought Magnitude Index. The normalization step has no impact on magnitude calculations, as drought periods are identified in relation to distribution quantiles and not soil moisture values.

4.4.1 Calendar year shift

As previously noted, the rehydration of soil at depth following a summer drought occurs later at the Chaingy site than in ISBA simulations. We attributed this difference to the modeling choice of vertically homogeneous soil hydraulic parameters in ISBA, which does not align with the stratification observed in reality.

The fact that the consequences of a drought can sometimes extend beyond the end of the calendar year, as was the case in 2016, highlights the need for an adaptation of the YDMI computation method. To quantify the drought episode in its entirety, calculations are conducted with a shift of 100 days in the calendar year. This new index is referred to as the 100-day shifted YDMI. The YDMI and the 100-day shifted YDMI, derived from ISBA simulations, are identical.

4.4.2 Comparison

The resulting 100-day shifted YDMI between 2017 and 2022 for the Chaingy site and ISBA simulations, along with the history of NatCat requests, are plotted in Figure 4.3.

The discrepancy between the YDMIs derived from the Chaingy probes and the ISBA simulation is immediately apparent. The probe data indicate that the two most significant drought years were 2018 and 2017, in contrast to 2022 and 2018 for the simulations. While the two graphs present different information, they both show consistency between probes or grid cells. This element demonstrates the robustness of the YDMI calculation method. The higher YDMI values obtained for the Chaingy site than for the ISBA simulation recall that the observed inter-annual variability is greater than that which was modeled.

The quasi-null YDMI for 2022 on the Chaingy site was an unexpected outcome, given the notable drought that has affected Europe this year [Toreti et al., 2022]. It is unclear whether this disparity is due to a local effect or an overestimation of soil moisture deficits by the ISBA model.

A review of the history of NatCat recognition requests (section 1.5.1) provides valuable insights to clarify this point. It is not relevant to focus solely on the town of Chaingy, as it submitted requests on an annual basis between 2017 and 2022. A more useful approach is be to count the number of requests issued by all the towns in the vicinity of Chaingy. This is illustrated in Figure 4.3 (C), which includes 19 towns within a 10 km radius. The data

indicates that there was a peak in recognition requests in both 2018 and 2022. If the requests were accepted in 2018 and not in 2022, it is because the recognition criteria based on the SWI uniform were met in the former case and not in the latter. Similar trends are observed when changing the radius to 20 km, indicating that this is not a mere sampling effect.

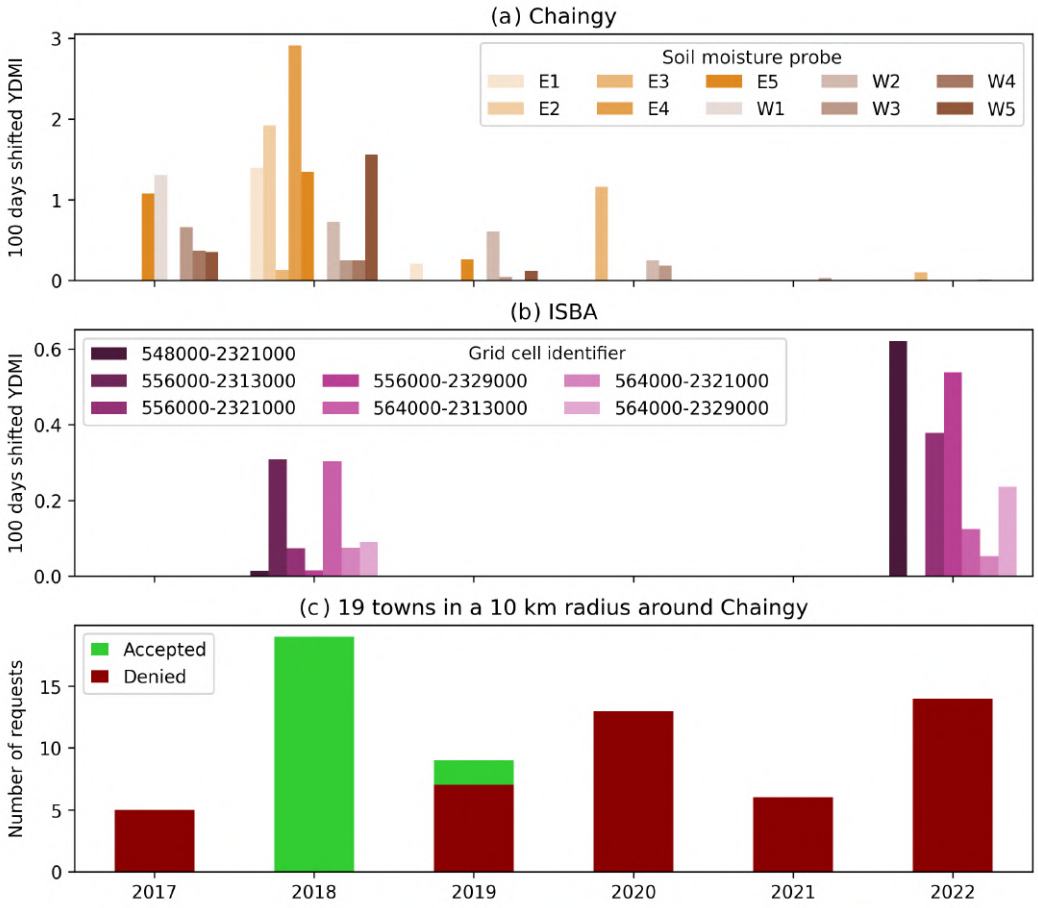


Figure 4.3: 100-day shifted YDMI for (a) the Chaingy site, (b) the 7 ISBA simulations (C_3 grassland, 100 to 150 cm), and (c) NatCat recognition requests for the 19 towns situated in a radius of 10 km around Chaingy, 2017-2022

The significant number of requests in 2018 suggests that the sample of towns experienced exceptional drought conditions that year. It is likely that this situation is underestimated by ISBA due to its coarse spatial scale. However, the data also indicates that 2022 was a year of extreme conditions, in line with the ISBA-derived YDMI. The only plausible explanation for the null YDMI at the Chaingy site this year is that the data is affected by very local factors.

4.5 Displacement

The analysis of displacements measured at the Chaingy site allows us to proceed with further analysis.

As previously described in section 2.4, the two extensometers, Ext1 and Ext2, are associated

with very different signals. Ext2 displays annual cycles of shrinkage and swelling, whereas Ext1 is subject to subsidence, the cause of which is currently unknown. It is more relevant to focus on the Ext2 acquired signal, given that its cyclical variations are similar to soil moisture's. The average Pearson correlation coefficient between displacement and soil moisture is 0.72. This relatively low value is partly due to the temporal lag between rehydration and swelling, because we are comparing measurements made at 120 cm and integrated over the entire depth. The displacement measured by Ext2 is recalled in Figure 4.4. For clarity, it has been centered by subtracting the mean, and smoothed by taking the median over a rolling window of 20 days.

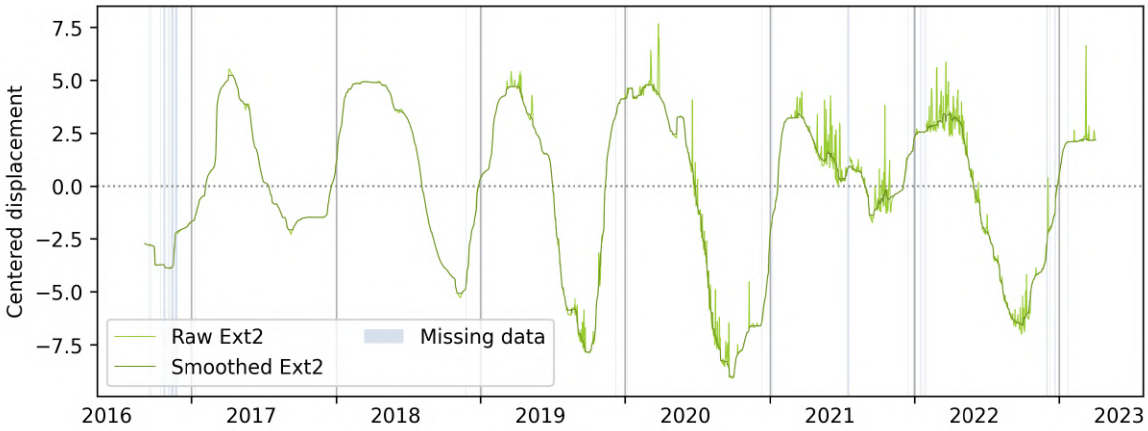


Figure 4.4: Centered displacement acquired by Ext2, raw and smoothed by taking the median over a rolling window of 20 days

Ext2 indicates that there was a significant shrinkage episode in 2020 that is not associated with drought, based on soil moisture probe-derived YDMI. One plausible explanation for this discrepancy could be that shrinkage leads to a decrease of the soil's volume, which in turn results in an increase of volumetric water content for a constant amount of water. This would indicate that soil moisture probes overestimate water content when the soil shrinks, as observed in 2020.

We have made a calculation to test this hypothesis. The shrinkage measured at the instrumented site is of the order of 0.2% of the soil thickness (shrinkage of about 5 mm for a thickness of 270 cm). Transposing this order of magnitude, we assume that a 1 m^3 cube would shrink by 2 mm. Such a decrease in volume is equivalent to multiplying the initial water content by 1.002 in the volumetric soil moisture equation. Since the water content measured by probes is between 0.15 and 0.55 $m^3.m^{-3}$, the variation is always inferior to the probe's precision of 0.1 $m^3.m^{-3}$. The shrinkage moisture range being from 0.15 to 0.35 $m^3.m^{-3}$, in the best case scenario the soil would have to shrink by 23 cm for the difference to be noticeable, which is far from the order of magnitude of the measured displacement. This demonstration shows that it is not possible to attribute the discrepancy between absence of drought and shrinkage to the probe's measurement of volumetric soil moisture.

However, all of the points made in this section cannot be generalized because they are based on a single measuring instrument.

It would have been interesting to compare the signal measured by each of the extensometers with a simulation of displacement based on soil moisture, made by models as described in section 1.3.3.2. However, this was not possible because such models require precise characterization of the soil, and the shallowest layer of our site was not geotechnically investigated.

4.6 Conclusion

In summary, ISBA provides reliable estimates of soil moisture. However, there are certain aspects of the simulation that do not fully match reality. The coarse resolution of 8 km results in an averaged signal, with no estimation of the variability within a grid cell. Data from the Chaingy instrumented site have shown that this variability can be important, even at a very local scale. It may be beneficial to implement a representation of soil volume variations in ISBA to more accurately characterize this variability. In particular, shrinkage can have a significant impact on moisture variations through the formation of desiccation cracks.

The objective of the following chapter 5, in the continuity of the approach, is to propose a retrospective and projections of clay shrinkage triggering conditions, for France, resorting to YDMIs.

Chapter 5

Drought index retrospective and projections

Contents

5.1 Introduction

Previously, chapter 3 presented the stages of development of a drought index adapted to clay-shrinkage induced subsidence, called the **Yearly Drought Magnitude Index (YDMI)**. The uncertainties arising from the fact that it is based on modeling at a coarse spatial scale were assessed in chapter 4, through a comparison with *in situ* data obtained at the Chaingy site instrumented by BRGM. As a reminder, the YDMI was developed based on a sample of claims from 25 highly exposed French towns, located in 5 different geographical areas of the country. Various configurations of ISBA and of the index were tested. We defined an **optimal YDMI configuration**, based on the vegetation patch modeling deciduous broadleaf trees (patch 4), the 8^{th} soil model layer (at a depth between 80 and 100 cm) and considering as drought thresholds the 1^{st} to 5^{th} daily SWI distribution percentiles.

As the method has already been presented and discussed, this chapter focuses on the calculation of YDMI for the entire country (with the optimal configuration). The index is computed for a historical period (2000-2022), forcing ISBA with the SAFRAN atmospheric reanalysis, and for a future period (2006-2065) forcing ISBA with the outputs of an ensemble of climate models under the scenarios RCP 4.5 and 8.5. This work provides a retrospective on past clay shrinkage triggering conditions, and gives insight on how they should evolve in the future.

The resulting historical and projected YDMI are the main topic of an article submitted to the Copernicus scientific journal *Hydrology and Earth System Sciences* [Barthelemy et al., 2024b], which is currently under review. The article will be summarized in section 5.2. In addition to these results, we will explore two points related to the coarse temporal and spatial scales of the YDMI. First, we will examine the temporality of past drought events in section 5.3.

Second, we will propose a statistical method to refine the spatial resolution of the YDMI in section 5.4.

5.2 Historical and projected YDMI [Barthelemy et al., 2024b]

5.2.1 Summary

The article submitted to *Hydrology and Earth System Sciences* focuses on YDMI calculations for France, both retrospective and projected.

YDMI were computed in their optimal configuration (deciduous broadleaf tree patch, 8^{th} model layer, and threshold value between the 1^{st} and the 5^{th} daily SWI distribution percentiles), for a historical period and for a future period. For the historical simulation, covering the years 2000 to 2022, ISBA was forced by the SAFRAN meteorological analysis. For the projected simulation, between 2006 and 2065, ISBA was forced by an ensemble of six different couples of a Global Climate Model (GCM) and a Regional Climate Model (RCM), under greenhouse gas emission scenarios RCP 4.5 and 8.5. The six GCM-RCM pairs were previously selected in the framework of the EXPLORE-2 project [OFB, 2024] on the basis of several criteria. These criteria, detailed in Robin et al. [2023], were availability for several RCPs, realistic behavior over Europe, absence of known errors, consistency of physical processes between GCM and RCM, and conservation of the dispersion modeled by EURO-CORDEX. In addition, the CO_2 feedback on vegetation is disabled in the projected ISBA simulations, in order to reduce uncertainties, in contrast to the historical simulation.

In both ISBA simulations, mountainous areas defined by a mean grid cell elevation superior to 1100 m were filtered out. We found that winter freezing of soil water at depth, which occurs above this altitude, biases the YDMI calculations. In addition, it is not relevant to model soil moisture in a mountainous context, because deep soils are only found in valleys whose elevation is lower than the mean grid cell elevation, making the modeling inaccurate.

The historical analysis highlights the years 2003, 2018, 2019, 2020 and 2022 in which significant widespread droughts occurred, based on the YDMI. The 2022 event was particularly intense. This assessment is consistent with the annual reports of the European Drought Observatory (EDO) [Micale et al., 2015, Masante et al., 2018, 2019, Barbosa et al., 2020, Toreti et al., 2022]. The projected YDMIs show an upward trend in the future, considering all models together. It is more pronounced for RCP 8.5 than for RCP 4.5. For several models, this increasing trend was characterized with high statistical significance by the Mann-Kendall trend test. Spatialization of the trends identifies southwestern and northeastern France as future hotspots.

A key point of the study is a discrepancy that there is between the historical and projected YDMI. Extreme years of the former, in particular 2022, are not modeled by the latter. Apart from the differences in vegetation modeling in ISBA, this can also be explained by the difficulty of climate models to represent extreme temperatures in Western Europe, as described by Vautard et al. [2023].

5.2.2 Article

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

Analyzing past and future droughts that induce clay shrinkage in France using an index based on water budget simulation for trees

Sophie Barthelemy[1,2,3], Bertrand Bonan[1], Miquel Tomas-Burguera[1,a], Gilles Grandjean[2], Séverine Bernardie[2], Jean-Philippe Naulin[3], Patrick le Moigne[1], Aaron Boone[1], and Jean-Christophe Calvet[1]

[1]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

[2]Bureau de Recherches Géologiques et Minières (BRGM), Orléans, France

[3]Caisse Centrale de Réassurance (CCR), Dpt R&D Modeling Cat & Agriculture, Paris, France

[a]Now at University of the Balearic Islands, Palma, Balearic Islands, Spain, and Natural Hazards and Emergencies Observatory of the Balearic Islands, Inca, Balearic Islands, Spain

Correspondence to: Jean-Christophe Calvet (jean-christophe.calvet@meteo.fr)

Abstract. Clay shrinkage is the retraction of clayey soils under dry conditions, caused by the loss of adsorbed water molecules from clay minerals. This phenomenon called clay-shrinkage induced subsidence can cause permanent damage to buildings if the drying extends below the foundations. In France, soils with these characteristics are widespread, affecting 48 % of the mainland territory (MTES, 2021), resulting in damage amounting to 20.8 billion euros since 1989 (CCR, 2023b). The causes of clay shrinkage are not yet fully understood, particularly at large spatial scales that are critical for land management. In a previous study (Barthelemy et al., 2023), a drought index designed specifically for clay shrinkage was created. It is a yearly index called the year drought magnitude. This index measures the daily soil moisture anomaly over the course of a year, based on the Interactions between Soil, Biosphere, and Atmosphere (ISBA) land surface model of Météo-France. Its properties have been fine-tuned by comparing it to a sample of insurance data. As a continuation of this work, our aim is to analyze past and future soil moisture drought events that may cause subsidence by calculating yearly drought magnitudes for France. Prior to this, we refined the ISBA configuration by focusing solely on tree vegetation types. Historical and projected simulations were conducted with the main difference being the meteorological forcing provided to ISBA. The historical simulation covered the years 2000-2022 and used the SAFRAN atmospheric reanalysis, while the projected simulation covered the years 2006-2065 and used an ensemble of climate models under Representative Concentration Pathway (RCP) 4.5 and RCP 8.5. The historical simulation revealed significant soil moisture droughts in France in 2003, 2018, 2019, 2020, and 2022. In 2022, there were notably high index values throughout the country. The projected simulation indicated that drought conditions are expected to worsen in the future, particularly under RCP 8.5 compared to RCP 4.5. The scenarios diverged significantly after 2046, and both the north and south of the country were equally affected. Differences between historical and projected year drought magnitudes were observed: projections are more pessimistic on average and more optimistic regarding extreme events. This discrepancy can be explained either by differences in climate forcing or by differences in the vegetation response of the land surface scheme.

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

1 Introduction

Soil moisture refers to the water stored in the unsaturated zone of the soil. Despite representing only 0.05% of total freshwater (Gleick et al. 1993), this reservoir plays a crucial role in climate-determining physical processes by interacting with adjacent land surface components. It is linked to the atmosphere in terms of both water and energy balance (Seneviratne et al., 2010). Water is exchanged through precipitation (positive fluxes) and evapotranspiration (negative fluxes). However, not all precipitation infiltrates into the soil. Some is intercepted by vegetation and evaporates before reaching the ground, while some runs off into surface reservoirs such as rivers and lakes. Evapotranspiration, which is the sum of bare soil evaporation, interception, and plant transpiration, requires a significant amount of energy to go from liquid to vapor phase and therefore has a significant impact on the energy balance. Vegetation should not be overlooked as it contributes the most to total land evapotranspiration on the global scale (Seneviratne et al., 2010). After entering the soil, some water moves to groundwater reservoirs through gravity (drainage). The soil retains moisture through capillary forces exerted by its granular structure. The water-holding capacity of the soil and the intensity of the forces depend on the characteristics of the pore network and the texture of the soil. Additionally, soil moisture distribution and evolution are heterogeneous within the soil volume. This distribution is influenced by local soil texture variations, as well as distance from the surface and the presence of vegetation roots.

In situ observations of soil moisture are scarce. Modeling soil moisture evolution over time is challenging due to the complexity of the system. Therefore, it is recommended to use land surface models (LSMs) instead of simpler alternatives (Berg and Sheffield, 2018). Soil moisture modeling is crucial for water resource management and weather forecasting because it affects water and energy balances (Dirmeyer et al., 2021). Soil moisture can affect the carbon cycle, as seen in Europe during the summer of 2003 when vegetation suffered from extreme drought and became a carbon source (Ciais et al., 2005). Soil moisture variability can also impact human activities that depend on it. Episodes of low moisture, also known as agricultural droughts, can have catastrophic consequences for agriculture and forestry when they occur over extended periods of time (Moravec et al., 2021).

Dry soils can cause shrinkage due to soil particles losing their adsorbed water molecules. This volume change is specific to clay minerals such as smectite, vermiculite, montmorillonite, which are formed by weathering and found in surface deposits that are inherently heterogeneous. Soil shrinkage can cause irreversible damage to buildings when differencial displacements occur between dry clay soil and the soil beneath the foundation, resulting in cracks in walls as evidence of deformation. This phenomenon is known as clay shrinkage induced subsidence. It is a global issue, as described by Barthelemy et al. (2023), and has been extensively studied (Meisina et al., 2006; Vincent et al., 2009; Soubeyroux et al., 2011; Hawkins, 2013; Gourdier and Plat, 2018; Mostafiz et al., 2021; Tzampoglou et al., 2022). In France, this phenomenon affects 48% of the territory to a moderate or strong degree (MTES, 2021). The “CatNat” (NatCat) scheme is a system under which the French state compensates for damages caused by natural disasters, including droughts that cause clay shrinkage. According to CCR

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

(2023b), this hazard costed 2.9 billion euros in compensations for 2022, while a total of 20.8 billion euros has been spent since 1989. The cost of damages has reached an unprecedented level since 2017.

Increasing atmospheric CO_2 concentrations are already causing significant climate changes globally (Lashof and Ahuja, 1990; Solomon et al., 2009; Hansen et al., 2013). An analysis of past conditions in Europe indicates changes in both precipitation and temperature patterns. Northern Europe is experiencing more precipitation, while southern Europe is becoming drier (Gudmundsson and Seneviratne, 2015). An increase in temperature has led to more dry and hot events in recent years compared to the second half of the 20th century (Manning et al., 2019). Climate projections also suggest drier meteorological conditions in Europe in the future (Spinoni et al., 2018). Some studies specifically examine changes in soil moisture trends. According to Samaniego et al. (2018), extreme agricultural drought events are expected to become more frequent, in line with precipitation and temperature projections. Despite its location between the wet north and the dry south, France is projected to experience drying trends in terms of both climatic conditions (Spinoni et al., 2018) and soil moisture (Vidal et al., 2012; Dayon et al., 2018; Soubeyroux et al., 2023). The most severe scenarios are expected to have larger impacts (Dayon et al., 2018).

Several estimates of the impact of climate change on the costs associated with subsidence due to clay shrinkage have been made for France (Gourdier and Plat (2018), Covéa and RiskWeatherTech (2022), and CCR (2023a)). These estimates are based on damage models, which are statistical models used operationally by the insurance industry to estimate losses following a disaster. In cases of subsidence due to clay shrinkage, inputs such as precipitation, temperature, and soil moisture may serve as proxies for weather conditions. Inputs such as clay content may serve as proxies for soil properties. All estimates agree that costs increase as desiccation increases in response to climate warming. The accuracy of these predictions depends on the quality of the inputs.

Clay shrinkage is a significant phenomenon that remains poorly understood at large spatial scales due to the complex soil moisture dynamics and the heterogeneity of clayey soils. Research is currently underway to better understand the drivers of clay shrinkage in order to assist public authorities and the insurance industry. In a previous study (Barthelemy et al., 2023), a yearly drought magnitude index (YDMI) was developed specifically for this hazard, based on soil moisture simulations from the Interactions between Soil, Biosphere, and Atmosphere (ISBA) land surface model. The YDMI quantifies daily soil moisture anomalies summed over a year based on a soil layer at a depth of 0.80 to 1 m. Its characteristics were adjusted by comparison with a sample of insurance claims data.

The objective of this study is to present a retrospective analysis of droughts in France since 2000 and to provide insight into climate warming-induced trends through YDMI. The YDMI is first computed for a historical period (2000-2022) by forcing ISBA with the SAFRAN atmospheric reanalysis. YDMI trends through 2065 in response to global warming are then assessed by forcing ISBA with the outputs of six different climate models under RCP 4.5 and RCP 8.5.

Section 2 presents the data used in this study, including the ISBA land surface model and historical and projected forcing, as well as the methods employed for index calculation and result analysis. The results are presented and described in Section 3, followed by a discussion and comparison with other works in Section 4. Finally, conclusions are drawn in Section 5.

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

2 Data and methods

This section describes the various surface simulations used and the concept of drought magnitude. The present analysis is summarized as a flowchart in Fig. 1.

2.1 The ISBA land surface model

In this study, soil moisture is simulated by ISBA within version 8.1 of the SURFEX (SURFace Externalisée) modeling platform for numerical weather prediction and climate modeling (Masson et al., 2013; Le Moigne et al., 2020). The ISBA model calculates surface water and energy budgets in response to an atmospheric forcing. Soil moisture at a specific time is determined by the balance between water inflows from precipitation and outflows through runoff, infiltration, and evapotranspiration (Noilhan and Planton, 1986; Noilhan and Mahfouf, 1996). The ISBA model covers the entire French metropolitan area on an 8 km grid, totaling 8925 grid points. The SAFRAN-ISBA-MODCOU (SIM) hydrological suite (Habets et al., 2008) is used operationally to monitor water resources, along with a meteorological analysis and a hydrogeological model. In this case, ISBA is used offline, meaning there is no feedback from the surface to the atmosphere. SAFRAN is a mesoscale analysis of near-surface atmospheric variables, including air temperature, air humidity, wind speed, solid and liquid precipitation, and incident solar and infrared radiation fields. It combines surface observations and atmospheric model simulations (Quintana-Seguí et al., 2008; Vidal et al., 2010).

In this study, we employ a multilayer version of ISBA, ISBA-DIF (Boone et al., 2000; Decharme et al., 2011), which uses a diffusive scheme. The soil column is divided into layers, with increasing thicknesses at greater depths, to better represent hydrological processes and the plant root water uptake. This is necessary due to significant water and temperature gradients at the surface, which require a finer mesh. The number of layers considered is proportional to the rooting depth of the vegetation, with a maximum of 10 layers (adding up to 2m depth) for trees. The definition of multiple layers at different depths is relevant because moisture variations depend on both the distance to the surface and the plant root density.

The ISBA model takes into account the influence of soil texture on moisture variations, as described by Decharme et al. (2011), using empirical pedotransfer functions. These functions establish a connection between parameters such as porosity, matric potential at saturation, and saturated hydraulic conductivity with sand and clay fractions. Specifically, the relationships derived from Clapp and Hornberger (1978) by Noilhan and Lacarrère (1995) are used. In this study, soil texture is (1) represented by the clay, sand, and silt contents derived from the Harmonized World Soil Database (HWSD) version 1.2 (Nachtergaele et al., 2012) at a kilometer resolution, and (2) assumed to be uniform throughout the soil column at each grid point.

To account for variability in land cover within a grid cell, ISBA runs separately for 12 patches. These patches include bare soil, rock, permanent snow, and 9 generic plant functional types: deciduous broadleaf trees, evergreen broadleaf trees, coniferous trees, C_3 crops, C_4 crops, irrigated C_4 crops, grassland, tropical grassland, and wetlands (C_3 and C_4 corresponding to different photosynthetic pathways). The separate runs can be aggregated by averaging the output variables and weighting each patch by its respective fraction in the grid cell. The depth of soil in a patch is determined by the rooting depth of the

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

corresponding vegetation type and may vary from patch to patch. The geographic distribution of the patches and land surface parameters are derived from the ECOCLIMAP-II database at kilometer resolution (Faroux et al., 2013).

The ISBA model includes a representation of photosynthesis using a CO_2-responsive stomatal conductance scheme, ISBA-A-gs (Calvet et al., 1998), because vegetation plays a critical role in near-surface soil moisture variations. This scheme has been improved over time to account for specific plant responses to drought (Calvet, 2000; Calvet et al., 2004). For each plant functional type, the ISBA-A-gs configuration simulates Leaf Area Index (LAI) from modeled leaf biomass, considering the mass-based leaf nitrogen concentration (Calvet and Soussana, 2001). Phenology is driven by photosynthesis, which is affected by soil moisture, leaf temperature, solar radiation, and air humidity. Therefore, all environmental conditions can influence the simulated LAI. The LAI is updated daily. The simulated LAI has been validated at global (Gibelin et al., 2006), continental (Szczypta et al., 2014), and regional (Brut et al., 2009) scales.

2.2 Historical and projected simulations

For this study, two different runs of ISBA were conducted, differing mainly in atmospheric forcing: (1) historical simulation, (2) projected simulation.

The historical simulation used the SAFRAN reanalysis to force ISBA with atmospheric data from 2000-2022. The atmospheric variables are available hourly on a regular 8 km grid. In this article, we refer to SAFRAN as a reference for the past climate.

The projected simulation experiment was conducted by forcing ISBA with projected atmospheric forcing developed in the framework of the EXPLORE-2 project (Explore2 - des futurs de l'eau, 2024). The dataset was generated in several steps. Projected greenhouse gas (GHG) concentrations were fed into a selection of global climate models (GCMs) which were coupled to regional climate models (RCMs) through dynamical downscaling, yielding atmospheric variables over a future period. The data were corrected for bias by comparing it with the SAFRAN historical reference reanalysis, using the ADAMONT method (Verfaillie et al., 2017). The GHG trajectories selected were Representative Concentration Pathways (RCPs) 2.6, 4.5, and 8.5 (Moss et al., 2010), and the RCM selection was based on the EURO-CORDEX works (Jacob et al., 2014, 2020; Kotlarski et al., 2014). These simulations cover the period from 2006 to 2100. For this study, we have chosen to use RCP4.5 and RCP8.5 scenarios applied to a set of 6 GCM-RCM combinations (detailed in Table 1), which we will refer to as models. This amounts to a total of 12 simulations, providing a range of estimates. In Explore2, GCM-RCM selection is based on several criteria: availability for several RCPs, realistic behavior over Europe, absence of known errors, consistency of physical processes between GCM and RCM, and conservation of the dispersion modeled by EURO-CORDEX (Robin et al., 2023). As we deal with rising atmospheric concentrations of CO_2, it is important to consider the conflicting effects of CO_2 on vegetation growth. On one hand, increased CO_2 content stimulates photosynthetic activity by increasing the availability of its main substrate, resulting in increased leaf area and plant transpiration. On the other hand, an increase in CO_2 also causes stomata closure, reducing leaf conductance to atmospheric gases and limiting transpiration. Climate change has been shown to have a significant impact on vegetation, as demonstrated by Laanaia et al. (2016) who predicted earlier leaf onset and peak LAI in the future. To reduce uncertainties, we disabled the effect of CO_2 on vegetation in ISBA for the projected simulation

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

experiment. We only considered data up to 2065, even though the projected forcing was available up to 2100, in order to limit uncertainties. For this experiment, we conducted 12 ISBA runs spanning from 2006 to 2065, divided into three time horizons: near future (2006-2025), future (2026-2045), and distant future (2046-2065).

Both experiments provide hourly volumetric soil moisture data for all patches and layers. Volumetric soil moisture is defined as the water content of the soil, expressed in units of volume of water per unit of volume of soil (m^3m^{-3}). Given our focus on long-term drought events, the data's hourly periodicity is unnecessary. Therefore, we can average the hourly soil moisture values to obtain daily values. Local factors such as the presence of trees around the buildings can play a significant role in the subsidence phenomenon during droughts (Freeman et al., 1992; Hawkins, 2013; Page, 1998). The impact of the local land cover variability can be reduced by selecting a single vegetation patch. To conduct a nationwide analysis, we require a patch that exists throughout the entire territory. This limits our options to bare soil, rocks, deciduous broadleaf trees, C_3 crops, C_4 crops, and grassland patches. The patch selection was made by comparing YDMI simulations to insurance subsidence claims using the method developed by Barthelemy et al. (2023), which resulted in choosing the deciduous broadleaf trees patch. The analysis also revealed the highest correlations for deep model layers, leading us to focus on the moisture variations of the model soil layer 8 (at a depth between 0.8 and 1.0 m), as described in Barthelemy et al. (2023). More information on the methodology and results can be found in the Supplement (see Fig. S1).

Therefore, this work will analyze ISBA outputs for the single deciduous broadleaf trees patch configuration, specifically for the 8th model soil layer located at a depth between 0.8 m and 1.0 m. For the present analysis, we have chosen to exclude mountainous areas. The reliability of ISBA simulations in mountainous areas may be lower due to the coarse spatial resolution. Weather conditions can change rapidly in rugged terrain, such as from one valley to another. Additionally, there is a major drawback to applying our method in a mountainous context. The YDMI is calculated from the liquid water content of the soil, which decreases in winter as soil temperatures fall below the freezing level. Liquid soil moisture distributions are biased by this phenomenon, which also affects the index computations. In mainland France, at a depth between 0.8 and 1.0 m, this phenomenon is only present in mountainous areas. These areas are identified by considering a threshold of 1100 m for the grid cell average altitude. We selected this value after analyzing ISBA soil ice patterns. A total of 567 grid cells (6.4%) meeting this condition were filtered out of the 8925 grid points.

2.3 The YDMI

The YDMI is used to characterize drought patterns based on soil moisture variations (Barthelemy et al., 2023). This index calculates the average integral of daily soil moisture in a year under threshold values ranging from the first to the fifth percentiles of the empirical distribution. Although originally designed to characterize clay shrinkage triggering conditions, it can be used more generally to identify long-term drought trends in soils. No fitting to a known statistical law is necessary and this avoids making assumptions about the distribution of soil moisture and its stability under a fluctuating climate.

The two sets of YDMI values are referred to as historical and projected magnitudes. The drought threshold values for the projected simulations are determined by the percentiles of the daily soil moisture distribution for the near future time horizon

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

of 2006-2025 only. The percentiles that define the thresholds are computed separately for each simulation (1 historical and 12 projected simulations), ensuring that droughts that occur with equal frequency are identified, even if the models are biased (Fig. S2). The statement that the longer the period, the more robust the distribution is not applicable in this context due to the hypothesis of a stable soil moisture distribution being irrelevant in the context of climate change. Additionally, basing the reference on the first 20 years allows for the identification of emerging trends during this period. Assuming a fixed definition of drought based on past values implies no adaptation and unchanged vulnerability of assets to drought conditions. This assumption is relevant in the context of clay shrinkage-induced subsidence, which mainly affects already-built homes.

The YDMI result is classified into four classes ranging from no drought (class 0) to extreme drought conditions (class 3). The methodology proposed by Barthelemy et al. (2023) for defining the classes involves sorting all positive YDMI values into three groups of equal size to form classes 1, 2, and 3 of increasing YDMI values, while all null values are assigned to class 0. The zoning of the classes is defined based on the historical YDMI set to ensure consistency between the two experiments.

2.4 Statistical analysis

The analysis focuses on the YDMI values obtained from historical and projected experiments. In section 3.1, we describe the spatial patterns of historical YDMI by commenting on individual maps for each year. Subsequently, we investigate differences in historical and projected YDMI in section 3.2. First, we provide a preliminary assessment by comparing their distributions, while separating them by RCP scenario, model, and time horizon. To investigate trends for each model and scenario, we performed a non-parametric Mann-Kendall test (Wilks, 2011). The null hypothesis associated with this test is the absence of trend, and the resulting p-value indicates the probability of obtaining such an outcome if the null hypothesis were correct. These results provide insight into the statistical significance of patterns in our data. After conducting this model-wise assessment, we compare in each point the third quartile (75th percentile) of YDMI between historical and projected sets, separated by RCP and time horizon. This provides spatialized and robust information on future drought trends. Finally, we assess trends in drought spatial extent by comparing the distribution of model grid cells in each YDMI class annually for each set, RCP scenario, and time horizon. Due to the expected opposite precipitation trends in northern (wetting) and southern (drying) Europe with global warming (Spinoni et al., 2018), we conducted an analysis at the national scale and for the two French regions, Grand Est and Occitanie, located in the North and South of the country, respectively.

2.5 NatCat regime

The national natural disaster compensation scheme in France is called “CatNat” (NatCat). It covers major natural disasters, including subsidence caused by clay shrinkage. Homeowners can only be compensated if their municipality is recognized as being in a state of natural disaster with respect to a specific hazard. Municipalities are the smallest French administrative units (there were 34,955 municipalities as of 01/01/2022). This recognition is done through the publication of an official decree for a given year. The criteria for recognition are based on exceptional events, with a return period of more than 10 years, that cause claims to be selected.. On of these criteria relies on ISBA simulations based on a simplified version of the model (Barthelemy

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

et al., 2023). The CCR website has the details of each decree, including the municipalities that have made the demand and the final decision. These details are available online (https://catastrophes-naturelles.ccr.fr/les-arretes). This paper compares historical YDMI with NatCat recognition requests. The latter are used as a proxy for drought-induced clay shrinkage.

3 Results

3.1 Historical drought magnitudes

Figure 2 shows the YDMI values for the years 2000 to 2022, computed from the ISBA historical simulation. Similar maps, which plot YDMI classes and dominant drought months, are available as Figs. S3 and S4 in the Supplement. The years 2003, 2018, 2019, 2020, and 2022 stand out for their high YDMI values. The 2022 drought event is the most remarkable of the 23-year series, both in terms of its geographic extent and intensity. The southwest region was the most affected. The year 2003 ranked second in terms of impact, affecting the northeast to southwest zones of the country and Brittany. In contrast, the events of 2018 and 2020 were more localized, mainly affecting the central to northeastern regions and, in 2020, the Rhône corridor. In 2019, a central to southwestern area was affected, except for the Cévennes. In addition to these large, widespread droughts, there were also several more localized events, such as the one in Brittany in 2006 or the one in the southeast and Corsica in 2017. This historical analysis allows us to identify 2003, 2018, 2019, 2020, and 2022 as significant drought years and provides a reference for assessing future droughts.

3.2 Projected drought magnitudes

Figure 3 presents the projected distributions of YDMI for different RCP scenarios, models, and time horizons. The gray box encompasses all models. The distribution of historical YDMI is represented by the yellow box labeled 2000-2022. Table 2 provides the values of the third quartile (percentile 75). For the first time horizon (2006-2025) and both RCPs, there are minimal differences between models and historical YDMI values. The results show a divergence with increasing time horizons. Although the distributions vary considerably for each model, the combination of all models indicates an upward trend over time. This effect is more pronounced for RCP8.5 than for RCP4.5. Specifically, the third quartile of YDMI is multiplied by 2 (0.2 to 0.40) for RCP4.5 and by 3 (0.21 to 0.62) for RCP8.5 from the near future to the distant future, considering all models together.

Figure 4 displays the trends in YDMI identified by the Mann-Kendall test for each of the 12 projected simulations processed separately. The different models present contrasting results for both RCPs. Although some punctual decreasing trends are identified, more in RCP 4.5 than in RCP 8.5, patterns of increasing YDMI dominate with higher confidence levels. This is particularly true for the EC-RCA4 and EC-RACM models in RCP 4.5, and models EC_RCA4, MPI_CCLM, and MPI_REMO in RCP 8.5. For RCP4.5, most models identify an increase in the southwestern region of the country, although the affected ISBA grid points rarely overlap between models. For RCP8.5, four out of six models identify a more widespread increase in

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

the same area, with half of them having higher levels of confidence. In three of the models, the increase extends as far as the northeast, and for MPI-CCLM, it is maximal in the west.

Figure 5 displays the third quartile (75th percentile) of YDMI in each grid cell, combining all models and separating time horizons and RCP. Focusing on the third quartile allows for a more robust analysis of trends in the highest YDMI values of the dataset compared to the maximum. The historical subplot indicates that the southwest and center of France have been the most affected by high YDMI, as observed on the individual maps. There is an observed trend of increasing YDMI over time, as seen across all models. Projections for both scenarios are similar up to 2045, but diverge in the final time horizon of 2046-2065, with RCP8.5 showing a more significant increase. The two scenarios agree on an increase, but on different scales, in southwestern France bordering the Pyrenees, and in eastern France from the Jura to the Vosges. Although we did not anticipate any differences between the YDMI projected for the historical period (2000-2022) and the near-future period (2006-2025), we did observe higher YDMI values in the former.

Figure 6 displays the YDMI class fractional area for each YDMI class, separated by time horizon and RCP scenario, for France, and for the Grand Est and Occitanie regions. It is observed that the proportion of drought-free years tends to decrease with time for both scenarios for France, with the average percentage of grid cells per year in class 0 decreasing from 60% for 2000-2022 to 40% for 2046-2065 RCP8.5. Compared to class 0, the percentage of cells with nonzero YDMI values (classes 1, 2, and 3) increases, particularly for RCP 8.5. Class 3 is the only category defined solely by a lower bound criterion, and therefore it concentrates extreme events, resulting in a high number of outliers. The section 3.1 identifies five notable drought years for the period 2000-2022, which stand out as class 3 outliers: 2022 (with 90% of France in class 3), 2003 (60%), 2020 (50%), 2018 (35%), and 2019 (20%). It is worth noting that the historical year 2022 remains unmatched in terms of drought, all models and scenarios considered. The projected class 3 outliers are biased downward compared to the other historical class 3 outlier years (2003, 2020, 2018, 2019). In addition to the observed inconsistency, there are significant differences in the distributions of historical and near-future YDMI values. This finding was unexpected, given the large time overlap (2000-2022 vs. 2006-2025). A higher proportion of null YDMI and lower proportion of nonzero YDMI were found for the historical experiment compared to the projected experiment, which is related to the outlier difference mentioned above. Similar conclusions can be drawn for the Grand Est region (Fig. 6b) and Occitanie (Fig. 6c). During the historical period, Grand Est experienced fewer droughts than Occitanie. However, both regions are projected to experience a drying pattern. This indicates that the differences between historical and projected YDMI are greater for Grand Est than for Occitanie.

4 Discussion

The objective of this work is to describe drought trends in France, both in the past and future periods. In the previous section, we presented a retrospective analysis of droughts in France between 2000 and 2022, by calculating YDMI using ISBA and the SAFRAN reanalysis. Five years stand out in order of importance: 2022, 2003, 2020, 2018, and 2019. We then analyzed future YDMI trends until 2065 under RCP 4.5 and RCP 8.5. All models indicate that drought characteristics will worsen in the future,

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

particularly under RCP 8.5. We discovered unexpected differences between historical and near-future YDMI values. In this section, we will discuss these findings in relation to other studies.

4.1 Is historical YDMI consistent with external sources of information?

To evaluate the significance of the YDMI in monitoring clay shrinkage-induced subsidence, we compare the retrospective index with the history of issued NatCat recognition requests. Figure 7 compares the distribution of YDMI in France for each year, sorted into four classes, with the number of NatCat requests issued. It is important to note that drought is not the only factor causing clay shrinkage, as the presence of clay mineral is necessary. Therefore, we do not anticipate a perfect correlation between the index and the requests. However, it is the most significant factor that changes over time, making year-to-year comparisons relevant. It is important to note that the comparison is made using the total number of requests, not just the accepted ones that led to a decree, as one of the recognition criteria is based on ISBA. As explained in the previous section and illustrated in Fig. 7a and Fig. 7b, the YDMI identifies 2022, 2003, 2020, 2018, and 2019 as significant drought years in order of importance. Figure 7c shows that the highest number of requests are concentrated in these years. Specifically, in 2003 and 2022, there were more than 7500 requests, which is over 20% of all French municipalities. However, it is important to note that there is a bias in the NatCat dataset as the number of requests issued each year increases with time. For instance, there were no requests in only two years: 2000 and 2001. Similar drought conditions occurred later in the record (2002, 2004, 2007, 2008, 2014, 2021), but this time they were associated with requests. This phenomenon is widely recognized in the insurance industry. Although the conditions remain the same, the number of requests increases over time due to two reasons: (1) an increase in awareness of the issue and (2) some municipalities repeat the process each year after their initial request is denied.

The European Drought Observatory (EDO) is a service provided by the Joint Research Center (JRC) of the European Commission. Since 2011, it has been reporting on significant drought events in Europe. Each report provides detailed information on the characteristics of the drought event at different levels, including temperature, precipitation, soil moisture, and vegetation. The assessment of soil moisture is based on simulations performed by the JRC hydrological model LISFLOOD. To ensure objectivity, we compared our YDMI retrospective with independent reports. Météo-France also conducts similar studies, but they are based on ISBA simulations. For the period 2011-2022, we collected five reports on the years 2015, 2018, 2019, 2020, and 2022. The 2015 report by Micale et al. (2015) indicates that France, particularly its eastern region, experienced a significant soil moisture deficit during the summer, surpassing that of 2003. In 2018, France experienced a drought that began in September in the northeast and spread to the rest of the country by the end of the month (Masante et al., 2018). The 2019 report by Masante et al. (2019) describes an event that mainly affected central France. In 2020, the drought episode centered on France throughout the summer, with the northern half of the country being the most affected (Barbosa et al., 2020). The 2022 report by Toreti et al. (2022) describes significant soil moisture anomalies throughout the country, which is consistent with our findings. All these works highlight the same drought years, except for 2015. The location of the event also matches,

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

such as central France in 2020 or the entire country in 2022. However, YDMI values in 2015 do not correspond to the EDO report. The event may have gone undetected due to the drought not propagating in depth this year.

4.2 How can differences between historical and projected YDMI be explained?

Section 3.2 describes the differences between historical and 2006-2025 projected YDMI. These differences were unexpected, because of the time overlap and the weak divergence of CO_2 concentrations in the projected near future. The projected simulations are more pessimistic in the median and more optimistic at the extremes than the historical simulations. The gap in the data may be attributed to differences in climate forcing between reanalysis and climate models, specifically in temperature or precipitation fields. This explanation is plausible, as a recent study by Vautard et al. (2023) demonstrated that CMIP6 climate models had difficulty simulating temperature extremes in Western Europe. Underestimation of high temperatures results in underestimation of drought, as high temperatures drive evaporative fluxes. The reason for the difference may also be found deeper in the modeling process, specifically in the land surface model. The only variation between the simulations is the CO_2 feedback on vegetation, which is enabled for the historical simulation and disabled for the projected simulation. This could lead to a dissimilar vegetation response for the same CO_2 levels. Berg and Sheffield (2018) suggest that failing to account for the impact of CO_2 on plant water use could result in a significant bias in soil moisture fluctuations. Although we previously identified two opposing effects of CO_2 on plant transpiration (Sect. 2.2), reduced stomatal conductance may dominate, resulting in limited transpiration and thus limited soil desiccation.

4.3 What are the sources of uncertainty?

The study's findings are the result of a lengthy modeling process. It is important to interpret them with care, considering the uncertainties associated with each step. These uncertainties are listed below.

4.3.1 Atmospheric forcing errors

The SAFRAN reanalysis utilized as the historical forcing combines observations and model data. Consequently, it may be subject to observational and meteorological model bias, as well as bias due to the coarse grid resolution. Similar limitations can be listed for the projected forcing. It is important to note the individual biases of each GCM and RCM. Theoretically, these biases tend to cancel each other out when using an ensemble of simulations. The ensemble we use has only 6 members. Adding more members, such as GCMs and RCMs, as well as multiple realizations of the same GCMs and RCMs to explore internal variability, would increase the robustness and reliability of the results. The weighting of each model contribution to the ensemble mean could also be considered (Ribes et al., 2021; Ribes et al., 2022). Model-to-model comparisons (as described in section 3.2) have identified trends of varying statistical significance. The climate modeling framework adopted here involves feeding the same CO_2 evolution into a set of models and assessing their response over time. However, it is not stated that all models warm at the same rate. Other approaches exist that take this element into account. For instance, Samaniego et al. (2018) analyze drought as a function of warming instead of time. This choice enhances model consistency and reduces uncertainty.

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

4.3.2 Model shortcomings

There are several sources of uncertainty associated with ISBA. The coarse model resolution of 8 km × 8 km is the first limitation, as clay shrinkage occurs on a local scale. ISBA provides an averaged meteorological forcing to a grid cell, resulting in model outputs that are correct on average but do not cover the range of possible values found within a real 64 km² square. The patch concept addresses the sub-grid land cover variability by generating as many outputs as there are vegetation types, and a single vegetation patch is used in this study. While vegetation type is a relevant factor contributing to heterogeneity, it is not the only one. At the grid cell scale, there may be several vegetation species of the same type, as well as different soil textures in various vertical combinations. To improve the approach, simulations could be repeated for a wider range of soil and vegetation parameters to better reflect reality. Changes in soil hydromechanical parameters with depth could also be implemented. Several databases such as SoilGrids (ISRIC, 2020) already provide such information of vertical texture variations. A disadvantage of ISBA is that its hydromechanical soil parameters remain stationary and do not change with time. However, the literature has shown that pore structure is affected by both drying-wetting and freeze-thaw cycles (Zhao et al., 2021). This oversimplification misrepresents the actual water content. Additionally, the lack of CO_2 feedback in the projected experiment, mentioned in the previous subsection, contributes to uncertainty. Although ISBA has been validated with in situ data (Decharme et al., 2011), it relies on simplifying assumptions that introduce bias. These assumptions are specific to each model and result from choices made during the design phase. To compensate for individual biases, it would be appropriate to base drought assessments on simulations from multiple land surface models, as done by Samaniego et al. (2018).

4.3.3 The lack of representativeness of the drought index

Finally, uncertainty arises from the method used to calculate the drought index. YDMI is an annual index, chosen to give it an operational focus, particularly towards the insurance industry, which mostly operates on a calendar year scale. This approach is consistent with soil moisture variations, as soil moisture periodically decreases each year during the summer season. However, this choice sacrifices our ability to identify and characterize multi-year drought events, which are expected to become more frequent with global warming (Hari et al., 2020). We define drought as occurring when the daily soil moisture value falls within the lowest 1 to 5% relative to a given period. In contrast to the historical experiment, the drought threshold for the projected experiment is defined over the first 20 years (2006-2025). Implicitly, this means that a drought event will always have the same impact, whether it occurs in 2006 or 2065. Regarding clay shrinkage, this hypothesis is realistic because most exposed assets already exist and are at risk. However, it is a particularly pessimistic view that completely ignores the impact of prevention. Preventive measures have recently been implemented to establish rules for the construction of new buildings (LégiFrance, 2018). In the meantime, research initiatives are being undertaken to prevent subsidence damage using various strategies. For instance, some aim to limit shrinkage by injecting resin (Al-Atroush et al., 2021), while others propose maintaining soil hydration during drought events (Ighil Ameur, 2021).

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

4.4 What are the limitations and possible applications of this work?

One limitation of this analysis is that mountainous areas were removed, which resulted in spatial gaps where clay shrinkage-induced subsidence may occur in limited locations. To address this, soil moisture calculations could be based on total water content instead of just liquid water content, which would allow these areas to be considered. However, the coarse resolution of ISBA is still a problem when dealing with rugged terrain and local biophysical processes.

In spite of these limitations, calculating YDMI can help characterize the conditions that cause clay shrinkage and add to our understanding of the phenomenon. This information is critical for policymakers and the insurance industry to better anticipate and manage the consequences of extreme drought events. The projected assessment is particularly relevant in the context of global warming. However, YDMI alone cannot be used for this purpose. For a complete assessment, triggering factors must be combined with the soil predisposing factor. The predisposing factor is the soil's ability to shrink, which is called susceptibility. Previous studies have provided an analysis of the susceptibility to the shrinkage-swell phenomenon (e.g. Olive et al., 1989), based on the combination of several predisposing factors, such as: proportion of clay material, heterogeneity of the layers, proportion of minerals susceptible to swelling, swelling-shrinkage intensity of the clay, layer thickness, and preferential water paths. Future improvements will involve combining the current drought index with BRGM's regional susceptibility maps. However, there are several challenges to overcome, particularly the spatial heterogeneity of clay deposits at large scales (Georisques, 2019).

5 Conclusion

This paper analyzes past and future drought events in France from 2000 to 2065 that are likely to cause clay shrinkage, using an adapted index. The YDMI index is based on soil moisture simulations from the ISBA land surface model, which is successively forced by an atmospheric reanalysis and an ensemble of six climate models. The index characteristics were adjusted in a previous study (Barthelemy et al., 2023) by comparison with insurance data. The results are as follows:

- Significant soil moisture drought events occurred in France in 2003, 2018, 2019, 2020, and 2022. In 2022, high index values were observed throughout the country.
- Drought characteristics are expected to worsen in the future, particularly under RCP8.5. The scenarios diverge significantly after 2046, and both the north and south of the country will be affected.
- Unexpected differences were found between historical and projected indices (projections are more pessimistic on average and more optimistic regarding extreme events). These differences may be due to either the climate forcing or the vegetation response of the land surface scheme.

Finally, YDMI quantifies drought on an annual scale for deciduous broadleaf tree vegetation, making it applicable in agricultural or forest management contexts beyond this specific example. This methodology can be applied to other countries, as clay shrinkage-induced subsidence is not unique to France. The SURFEX modeling platform that supports ISBA is applicable anywhere, provided that high-resolution atmospheric forcing is available.

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

Code availability

The analysis was carried out with Python codes that can be made available upon request. SURFEX can be downloaded freely at http://www.umr-cnrm.fr/surfex/data/OPEN-SURFEX/open_surfex_v8_1_20210914.tar.gz (CNRM, 2016). It is provided under a CECILL-C license (French equivalent to the LGPL license).

Data availability

The data presented in the Figures are available online at:

Supplement

The supplement related to this article is available online at:

Author contributions

SBa and JCC designed the experiments. SBa performed the investigation, did the formal analysis and wrote the paper. BB did the historical simulations. MTB did the projected simulations. All co-authors participated in the interpretation of the results and the revision of the paper.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgments

The authors would like to thank the operational services of Météo-France (DCSC).

Financial support

The PhD thesis work of Sophie Barthelemy was co-funded by BRGM (Bureau de Recherches Géologiques et Minières), CCR (Caisse Centrale de Réassurance) and Météo-France.

References

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

Al-Atroush, M. E., Shabbir, O., Almeshari, B., Waly, M., and Sebaey, T. A.: A Novel Application of the Hydrophobic Polyurethane Foam: Expansive Soil Stabilization, Polymers, 13, 1335, https://doi.org/10.3390/polym13081335, 2021.

Barbosa, P., Masante, D., Arias Muños, C., Cammalleri, C., de Jager, A., Magni, D., Mazzeschi, M., McCormick, N., Naumann, G., Spinoni, J., and Vogt, J.: Drought in Europe September 2020, Publications Office of the European Union, Available at: https://edo.jrc.ec.europa.eu/documents/news/EDODroughtNews202009_Europe.pdf (last access: 27 February 2024), 2020.

Barthelemy, S., Bonan, B., Calvet, J.-C., Grandjean, G., Kapsambelis, D., and Bernardie, S.: A new approach for drought index adjustment to clay-shrinkage-induced subsidence over France: advantages of the interactive leaf area index, Nat. Hazards Earth Syst. Sci., 24, 999–1016, https://doi.org/10.5194/nhess-24-999-2024, 2024.

Berg, A. and Sheffield, J.: Climate Change and Drought: the Soil Moisture Perspective, Curr. Clim. Change Rep., 4, 180–191, https://doi.org/10.1007/s40641-018-0095-0, 2018.

Boone, A., Masson, V., Meyers, T., and Noilhan, J.: The Influence of the Inclusion of Soil Freezing on Simulations by a Soil–Vegetation–Atmosphere Transfer Scheme, J. Appl. Meteorol., 39, 1544–1569, https://doi.org/10.1175/1520-0450(2000)039<1544:TIOTIO>2.0.CO;2, 2000.

Brut, A., Rüdiger, C., Lafont, S., Roujean, J. L., Calvet, J. C., Jarlan, L., Gibelin, A. L., Albergel, C., Le Moigne, P., Soussana, J. F., Klumpp, K., Guyon, D., Wigneron, J. P., and Ceschia, E.: Modelling LAI at a regional scale with ISBA-A-gs: Comparison with satellite-derived LAI over southwestern France, Biogeosciences, 6, 1389–1404, https://doi.org/10.5194/bg-6-1389-2009, 2009.

Calvet, J.-C.: Investigating soil and atmospheric plant water stress using physiological and micrometeorological data, Agric. For. Meteorol., 103, 229–247, https://doi.org/10.1016/S0168-1923(00)00130-1, 2000.

Calvet, J.-C. and Soussana, J.-F.: Modelling CO_2-enrichment effects using an interactive vegetation SVAT scheme, Agric. For. Meteorol., 108, 129–152, https://doi.org/10.1016/S0168-1923(01)00235-0, 2001.

Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A., and Wigneron, J.-P.: An interactive vegetation SVAT model tested against data from six contrasting sites, Agric. For. Meteorol., 92, 73–95, https://doi.org/10.1016/S0168-1923(98)00091-4, 1998.

Calvet, J.-C., Rivalland, V., Picon-Cochard, C., and Guehl, J.-M.: Modelling forest transpiration and CO_2 fluxes—response to soil moisture stress, Agric. For. Meteorol., 124, 143–156, https://doi.org/10.1016/j.agrformet.2004.01.007, 2004.

CCR: Conséquences du changement climatique sur le coût des catastrophes naturelles en France à horizon 2050, Available at: https://www.ccr.fr/documents/35794/1255983/CCR+Etude+climat+BAG+23102023+page+22mo.pdf/68b95f6e-8238-4dcc-6c56-025fa410257b?t=1698161402128 (last access: 27 February 2024), 2023a.

CCR: Les Catastrophes Naturelles en France Bilan 1982-2022, Available at: https://catastrophes-naturelles.ccr.fr/documents/148935/368920/BILAN+Cat+Nat+2022_HD_12062023.pdf/5b5397f0-2118-2252-608f-76ece6e195a1?t=1686731142230 (last access: 27 February 2024), 2023b.

Ciais, Ph., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, Chr., Carrara, A., Chevallier, F., De Noblet, N., Friend, A. D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T., and Valentini, R.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, Nature, 437, 529–533, https://doi.org/10.1038/nature03972, 2005.

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water Resour. Res., 14, 601–604, https://doi.org/10.1029/WR014i004p00601, 1978.

Covéa and RiskWeatherTech: Changement climatique et Assurance: Quelles conséquences sur la sinistralité à horizon 2050?, Available at : https://www.covea.com/sites/default/files/2022-02/202202_Livre_Blanc_Cov%C3%A9a_Risques_Climatiques.pdf (last access: 27 February 2024), 2022.

Dayon, G., Boé, J., Martin, É., and Gailhard, J.: Impacts of climate change on the hydrological cycle over France and associated uncertainties, Comptes Rendus - Geosci., 350, 141–153, https://doi.org/10.1016/j.crte.2018.03.001, 2018.

Decharme, B., Boone, A., Delire, C., and Noilhan, J.: Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions, J. Geophys. Res. Atmospheres, 116, 1–29, https://doi.org/10.1029/2011JD016002, 2011.

Dirmeyer, P. A., Balsamo, G., Blyth, E. M., Morrison, R., and Cooper, H. M.: Land-Atmosphere Interactions Exacerbated the Drought and Heatwave Over Northern Europe During Summer 2018, AGU Adv., 2, e2020AV000283, https://doi.org/10.1029/2020AV000283, 2021.

Explore2 - des futurs de l'eau: https://professionnels.ofb.fr/fr/node/1244, last access: 27 February 2024.

Faroux, S., Kaptué Tchuenté, A. T., Roujean, J.-L., Masson, V., Martin, E., and Le Moigne, P.: ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, Geosci. Model Dev., 6, 563–582, https://doi.org/10.5194/gmd-6-563-2013, 2013.

Freeman, T. J., Burford, D., and Crilly, M. S.: Seasonal foundation movements in London Clay, in: Proceedings of the 4th International Conference on Ground Movements and Structures, Cardiff, UK, 8–11 July 1991, 485–501, 1992.

Georisques: Dossier expert sur le retrait-gonflement des argiles – Exposition du territoire au phénomène, https://www.georisques.gouv.fr/articles-risques/retrait-gonflement-des-argiles/exposition-du-territoire-au-phenomene (last access: 2 April 2024), 2019.

Gibelin, A. L., Calvet, J. C., Roujean, J. L., Jarlan, L., and Los, S. O.: Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products, J. Geophys. Res. Atmospheres, 111, https://doi.org/10.1029/2005JD006691, 2006.

Gleick, P. H., Pacific Institute for Studies in Development, Environment, and Security, Stockholm Environment Institute, and Pacific Institute for Studies in Development, Environment, and Security (Eds.): Water in crisis: a guide to the world's fresh water resources, Oxford Univ. Press, New York, NY, 473 pp., ISBN 978-0-19-507628-8, 1993.

Gourdier, S. and Plat, E. : Impact du changement climatique sur la sinistralité due au retrait-gonflement des argiles. Journées Nationales de Géotechnique et Géologie de l'Ingénieur (JNGG), Champs-sur-Marne, France, https://brgm.hal.science/hal-01768395/ (last access: 27 February 2024), 2018.

Gudmundsson, L. and Seneviratne, S. I.: European drought trends, Proc. Int. Assoc. Hydrol. Sci., 369, 75–79, https://doi.org/10.5194/piahs-369-75-2015, 2015.

Habets, F., Boone, A., Champeaux, J. L., Etchevers, P., Franchistéguy, L., Leblois, E., Ledoux, E., Le Moigne, P., Martin, E., Morel, S., Noilhan, J., Seguí, P. Q., Rousset-Regimbeau, F., and Viennot, P.: The SAFRAN-ISBA-MODCOU

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

hydrometeorological model applied over France, J. Geophys. Res. Atmospheres, 113, 1–18, https://doi.org/10.1029/2007JD008548, 2008.

Hansen, J., Kharecha, P., Sato, M., Masson-Delmotte, V., Ackerman, F., Beerling, D.J., Hearty, P.J., Hoegh-Guldberg, O., Hsu, S.-L., and Parmesan, C.: Assessing "dangerous climate change": required reduction of carbon emissions to protect young people, future generations and nature, PLoS ONE, 8, e81648, https://doi.org/10.1371/journal.pone.0081648, 2013.

Hari, V., Rakovec, O., Markonis, Y., Hanel, M., and Kumar, R.: Increased future occurrences of the exceptional 2018–2019 Central European drought under global warming, Sci. Rep., 10, 12207, https://doi.org/10.1038/s41598-020-68872-9, 2020.

Hawkins, A. B.: Some engineering geological effects of drought: Examples from the UK, Bull. Eng. Geol. Environ., 72, 37–59, https://doi.org/10.1007/s10064-013-0458-7, 2013.

Ighil Ameur, L.: Experimental Analysis of Shrinkage-swelling Phenomenon of Clays Application to an Individual House Affected by Drought Under Climate Change Effects, ASCE EMI/PMC, Available at: https://www.researchgate.net/publication/356284599_Experimental_Analysis_of_Shrinkage-swelling_Phenomenon_of_Clays_-Application_to_an_Individual_House_Affected_by_Drought_Under_Climate_Change_Effects (last access: 27 February 2024), 2021.

ISRIC: SoilGrids — global gridded soil information, https://www.isric.org/explore/soilgrids (last access: 2 April 2024), 2020.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Reg. Environ. Change, 14, 563–578, https://doi.org/10.1007/s10113-013-0499-2, 2014.

Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., Benestad, R., Boberg, F., Buonomo, E., Cardoso, R. M., Casanueva, A., Christensen, O. B., Christensen, J. H., Coppola, E., De Cruz, L., Davin, E. L., Dobler, A., Domínguez, M., Fealy, R., Fernandez, J., Gaertner, M. A., García-Díez, M., Giorgi, F., Gobiet, A., Goergen, K., Gómez-Navarro, J. J., Alemán, J. J. G., Gutiérrez, C., Gutiérrez, J. M., Güttler, I., Haensler, A., Halenka, T., Jerez, S., Jiménez-Guerrero, P., Jones, R. G., Keuler, K., Kjellström, E., Knist, S., Kotlarski, S., Maraun, D., van Meijgaard, E., Mercogliano, P., Montávez, J. P., Navarra, A., Nikulin, G., de Noblet-Ducoudré, N., Panitz, H.-J., Pfeifer, S., Piazza, M., Pichelli, E., Pietikäinen, J.-P., Prein, A. F., Preuschmann, S., Rechid, D., Rockel, B., Romera, R., Sánchez, E., Sieck, K., Soares, P. M. M., Somot, S., Srnec, L., Sørland, S. L., Termonia, P., Truhetz, H., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community, Reg. Environ. Change, 20, 51, https://doi.org/10.1007/s10113-020-01606-9, 2020.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7, 1297–1333, https://doi.org/10.5194/gmd-7-1297-2014, 2014.

Laanaia, N., Carrer, D., Calvet, J.-C., and Pagé, C.: How will climate change affect the vegetation cycle over France? A generic modeling approach, Clim. Risk Manag., 13, 31–42, https://doi.org/10.1016/j.crm.2016.06.001, 2016.

Lashof, D.A., and Ahuja, D.R.: Relative contributions of greenhouse gas emissions to global warming, Nature, 344, 529–

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

531, https://doi.org/10.1038/344529a0, 1990.

Le Moigne, P., Besson, F., Martin, E., Boé, J., Boone, A., Decharme, B., Etchevers, P., Faroux, S., Habets, F., Lafaysse, M., Leroux, D., and Rousset-Regimbeau, F.: The latest improvements with SURFEX v8.0 of the Safran-Isba-Modcou hydrometeorological model for France, Geosci. Model Dev., 13, 3925–3946, https://doi.org/10.5194/gmd-13-3925-2020, 2020.

Légifrance: Loi n° 2018-1021 du 23 novembre 2018 portant évolution du logement, de l'aménagement et du numérique, 2018-1021. Available at: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000037639478 (last access: 2 April 2024), 2018.

Manning, C., Widmann, M., Bevacqua, E., Van Loon, A. F., Maraun, D., and Vrac, M.: Increased probability of compound long-duration dry and hot events in Europe during summer (1950–2013), Environ. Res. Lett., 14, 094006, https://doi.org/10.1088/1748-9326/ab23bf, 2019.

Masante, D., Vogt, J., Cammalleri, C., Spinoni, J., and Barbosa, P.: Drought in Central-Northern Europe - September 2018, Publications Office of the European Union, Available at: https://edo.jrc.ec.europa.eu/documents/news/EDODroughtNews201809_Central_North_Europe.pdf (last access: 27 February 2024), 2018.

Masante, D., Barbosa, P., and Magni, D.: Drought in Europe August 2019, Publications Office of the European Union, Available at: https://edo.jrc.ec.europa.eu/documents/news/EDODroughtNews201908_Europe.pdf (last access: 27 February 2024), 2019.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J. C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A. L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J. F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geosci. Model Dev., 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, 2013.

Meisina, C., Zucca, F., Fossati, D., Ceriani, M., and Allievi, J.: Ground deformation monitoring by using the Permanent Scatterers Technique: The example of the Oltrepo Pavese (Lombardia, Italy), Eng. Geol., 88, 240–259, https://doi.org/10.1016/j.enggeo.2006.09.010, 2006.

Micale, F., Vogt, J., and Cammalleri, C.: European Drought Observatory (EDO) Drought News August 2015, Publications Office of the European Union, Available at: https://edo.jrc.ec.europa.eu/documents/news/EDODroughtNews201508.pdf (last access: 27 February 2024), 2015.

Moravec, V., Markonis, Y., Rakovec, O., Svoboda, M., Trnka, M., Kumar, R., and Hanel, M.: Europe under multi-year droughts: how severe was the 2014–2018 drought period?, Environ. Res. Lett., 16, 034062, https://doi.org/10.1088/1748-9326/abe828, 2021.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J.: The next generation of scenarios for climate change research and assessment, Nature, 463, 747–756, https://doi.org/10.1038/nature08823, 2010.

Mostafiz, R. B., Friedland, C. J., Rohli, R. V., Bushra, N., and Held, C. L.: Property Risk Assessment for Expansive Soils in

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

Louisiana, Front. Built Environ., 7, 1–10, https://doi.org/10.3389/fbuil.2021.754761, 2021.

MTES: Cartographie de l'exposition des maisons individuelles au retrait-gonflement des argiles, Available at: https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2021-06/note_methode_croisement_retrait_gonflement_argiles_juin2021v3.pdf (last access: 27 February 2024), 2021.

Nachtergaele, F., Velthuizen, H., Verelst, L., and Wiberg, D.: Harmonized World Soil Database Version 1.2, Available at: https://esdac.jrc.ec.europa.eu/ESDB_Archive/Soil_Data/Docs_GlobalData/Harmonized_World_Soi_Database_v1.2.pdf (last access: 27 February 2024), 2012.

Noilhan, J. and Lacarrère, P.: GCM Grid-Scale Evaporation from Mesoscale Modeling, J. Clim., 8, 206–223, https://doi.org/10.1175/1520-0442(1995)008<0206:GGSEFM>2.0.CO;2, 1995.

Noilhan, J. and Mahfouf, J.-F.: The ISBA land surface parameterisation scheme, Global Planet. Change, 13, 145–159, https://doi.org/10.1016/0921-8181(95)00043-7, 1996.

Noilhan, J. and Planton, S.: A simple parameterization of land surface processes for meteorological models, Mon. Weather Rev., 117, 536–549, 1989.

Olive, W. W., Chleborad, A. F., Frahme, C. W., Schlocker, J., Schneider, R. R., and Schuster, R. L.: Swelling clays map of the conterminous United States, https://doi.org/10.3133/i1940, 1989.

Page, R. C. J.: Reducing the cost of subsidence damage despite global warming, Struct. Surv., 16, 67–75, https://doi.org/10.1108/02630809810219641, 1998.

Ribes, A., Boé, J., Qasmi, S., Dubuisson, B., Douville, H., and Terray, L.: An updated assessment of past and future warming over France based on a regional observational constraint, Earth Syst. Dynam., 13, 1397–1415, https://doi.org/10.5194/esd-13-1397-2022, 2022.

Ribes, A., Qasmi S., and Gillett N. P.: Making climate projections conditional on historical observations. Sci. Adv., 7, eabc0671, https://doi.org/10.1126/sciadv.abc0671, 2021.

Robin, Y., Corre, L., Marson, P., Bernus, S., Vrac, M., and Thao, S.: Projections climatiques régionalisées : correction de biais et changements futurs. Available at: https://entrepot.recherche.data.gouv.fr/file.xhtml?persistentId=doi:10.57745/99X4CD (last access: 27 February 2024), 2023.

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France, J. Appl. Meteorol. Climatol., 47, 92–107, https://doi.org/10.1175/2007JAMC1636.1, 2008.

Samaniego, L., Thober, S., Kumar, R., Wanders, N., Rakovec, O., Pan, M., Zink, M., Sheffield, J., Wood, E. F., and Marx, A.: Anthropogenic warming exacerbates European soil moisture droughts, Nat. Clim. Change, 8, 421–426, https://doi.org/10.1038/s41558-018-0138-5, 2018.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Sci. Rev., 99, 125–161, https://doi.org/10.1016/j.earscirev.2010.02.004, 2010.

Solomon, S., Plattner, G.-K., Knutti, R., and Friedlingstein, P.: Irreversible climate change due to carbon dioxide emissions. Proc. Natl. Acad. Sci. USA, 106, 1704–1709, https://doi.org/10.1073/pnas.0812721106, 2009.

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

Soubeyroux, M., Grigis, L., Rousset, F., Corre, L., and Boe, J.: Evolution des sécheresses météorologiques et agricoles en France en contexte de changement climatique, 36th Conference of the International Association of Climatology, Available at: https://drive.google.com/file/d/1Fw18eT6FZz756wwla0vh5Rhxy6XwpNYt/view (last access: 27 February 2024), 2023.

Soubeyroux, J.-M., Vidal, J.-P., Najac, J., Kitova, N., Blanchard, M., Dandin, P., Martin, E., Pagé, C., Habets, F. : Projet ClimSec : Impact du changement climatique en France sur la sécheresse et l'eau du sol. Rapport final du projet, https://www.umr-cnrm.fr/IMG/pdf/2011_fmaif_rapport_final_v2.2.pdf (last access: 27 February 2024), 2011.

Spinoni, J., Vogt, J. V., Naumann, G., Barbosa, P., and Dosio, A.: Will drought events become more frequent and severe in Europe?, Int. J. Climatol., 38, 1718–1736, https://doi.org/10.1002/joc.5291, 2018.

Szczypta, C., Calvet, J. C., Maignan, F., Dorigo, W., Baret, F., and Ciais, P.: Suitability of modelled and remotely sensed essential climate variables for monitoring Euro-Mediterranean droughts, Geosci. Model Dev., 7, 931–946, https://doi.org/10.5194/gmd-7-931-2014, 2014.

Toreti, A., Bavera, D., Acosta Navarro, J., Cammalleri, C., de Jager, A., Di Ciollo, C. i, Hrast Essenfelder, A., Maetens, W., Masante, D., Magni, D., Mazzeschi, M., and Spinoni, J.: Drought in Europe - August 2022, JRC Global Drought Observatory, https://doi.org/10.2760/264241, 2022.

Tzampoglou, P., Loukidis, D., and Koulermou, N.: Seasonal Ground Movement Due to Swelling/Shrinkage of Nicosia Marl, Remote Sens., 14, 1440, https://doi.org/10.3390/rs14061440, 2022.

Vautard, R., Cattiaux, J., Happé, T., Singh, J., Bonnet, R., Cassou, C., Coumou, D., D'Andrea, F., Faranda, D., Fischer, E., Ribes, A., Sippel, S., and Yiou, P.: Heat extremes in Western Europe increasing faster than simulated due to atmospheric circulation trends, Nat. Commun., 14, 6803, https://doi.org/10.1038/s41467-023-42143-3, 2023.

Verfaillie, D., Déqué, M., Morin, S., and Lafaysse, M.: The method ADAMONT v1.0 for statistical adjustment of climate projections applicable to energy balance land surface models, Geosci. Model Dev., 10, 4257–4283, https://doi.org/10.5194/gmd-10-4257-2017, 2017.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, Int. J. Climatol., 30, 1627–1644, https://doi.org/10.1002/joc.2003, 2010.

Vidal, J.-P., Martin, E., Kitova, N., Najac, J., and Soubeyroux, J.-M.: Evolution of spatio-temporal drought characteristics: validation, projections and effect of adaptation scenarios, Hydrol. Earth Syst. Sci., 16, 2935–2955, https://doi.org/10.5194/hess-16-2935-2012, 2012.

Vincent, M., Cojean, R., Fleureau, J.-M., Cui, Y. J., Jacquard, C., Kazmierczak, J.-B., Masrouri, F., Tessier, D., Alimi-Ichola, I., Magnan, J.-P., Blanchard, M., Fabre, D., Pantet, A., Audiguier, M., Plat, E., Souli, H., Taibi, S., Tang, A.-M., Morlock, C., Maison, T., Mrad, M., Bréda, N., Djeran-Maigre, I., Duc, M., Soubeyroux, J.-M., Denis, A., Proust, D., Geremew, Z., Le Roy, S., Dumont, M., Hemmati, S., Nowamooz, H., Coquet, Y., Pothier, C., Makki, L., Chrétien, M., Fontaine, C. : Rapport de synthèse final du projet ARGIC (Analyse du retrait-gonflement et de ses Incidences sur les Constructions), Projet ANR-05-PRGCU-005, BRGM/RP-57011-FR report, http://infoterre.brgm.fr/rapports/RP-57011-FR.pdf (last access: 27 February 2024), 2009.

Wilks, D. S.: Statistical methods in the atmospheric sciences, 3rd ed., Elsevier/Academic Press, Amsterdam ; Boston, 676 pp., ISBN 13: 978-0-12-751966-1, 2011.

Zhao, G., Zou, W., Han, Z., Wang, D., and Wang, X.: Evolution of soil-water and shrinkage characteristics of an expansive clay during freeze-thaw and drying-wetting cycles, Cold Reg. Sci. Technol., 186, 103275,

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

https://doi.org/10.1016/j.coldregions.2021.103275, 2021.

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

740 **Table 1: Six GCM-RCM atmospheric model combinations selected as forcing for the projected YDMI derived from the simulations of the ISBA land surface model.**

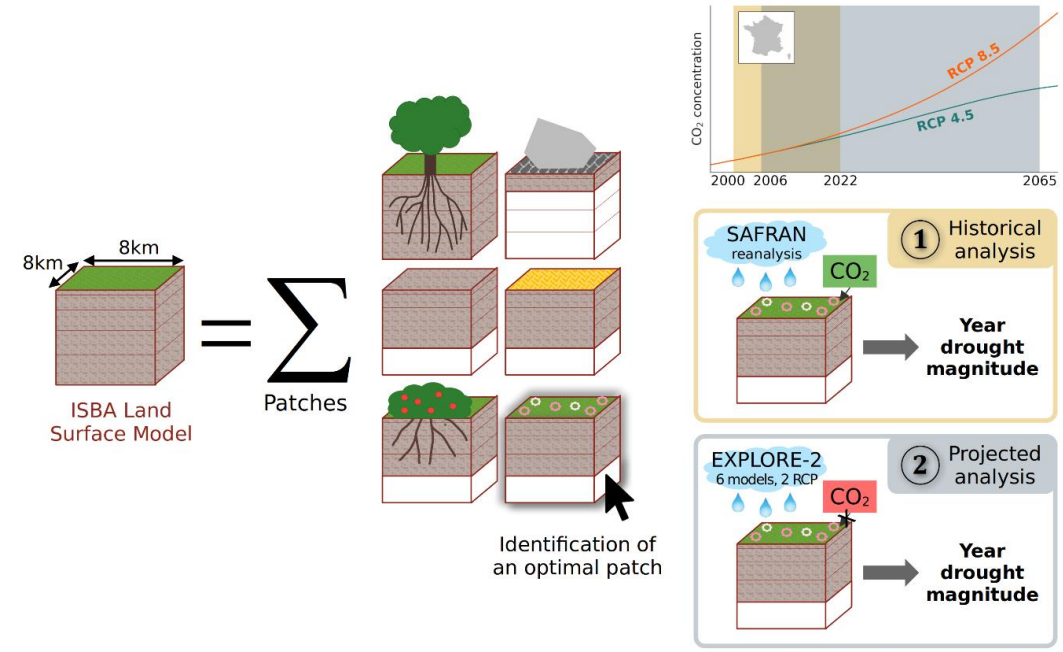
GCM-RCM atmospheric model combinations	GCM	RCM
CNRM-ALAD	CNRM-CM5	ALADIN63
CNRM-RACM	CNRM-CM5	RACMO22E
EC-RACM	EC-EARTH	RACMO22E
EC-RCA4	EC-EARTH	RCA4
MPI-CCLM	MPI-ESM	CCLM4-8-17
MPI-REMO	MPI-ESM	REMO2009

745

Table 2: Third quartile (75th percentile) of historical and projected YDMI statistical distribution for RCP 4.5 and RCP8.5, separating time horizons and atmospheric models. YDMI values larger or equal to 0.4 are in bold.

Atmospheric forcing		2000-2022	2006-2025	2026-2045	2046-2065
SAFRAN (historical)		0.18			
RCP 4.5	**Pooled data from the 6 models**		0.20	0.29	**0.40**
	CNRM_ALAD		0.24	**0.62**	0.27
	CNRM_RACM		0.24	**0.77**	0.33
	EC_RAC4		0.15	0.04	**0.75**
	EC_RACM		0.15	0.03	**0.76**
	MPI_CCLM		0.23	0.32	0.29
	MPI_REMO		0.20	0.33	0.19
RCP 8.5	**Pooled data from the 6 models**		0.21	0.28	**0.62**
	CNRM_ALAD		0.22	0.16	0.22
	CNRM_RACM		0.25	0.05	0.25
	EC_RAC4		0.15	0.16	**0.70**
	EC_RACM		0.19	0.15	**0.40**
	MPI_CCLM		0.24	**0.91**	**1.61**
	MPI_REMO		0.22	**0.59**	0.99

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.



750

Figure 1: Flowchart of the methodology deployed in this study with focus on vegetation patch selection and historical vs. projected analyses.

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

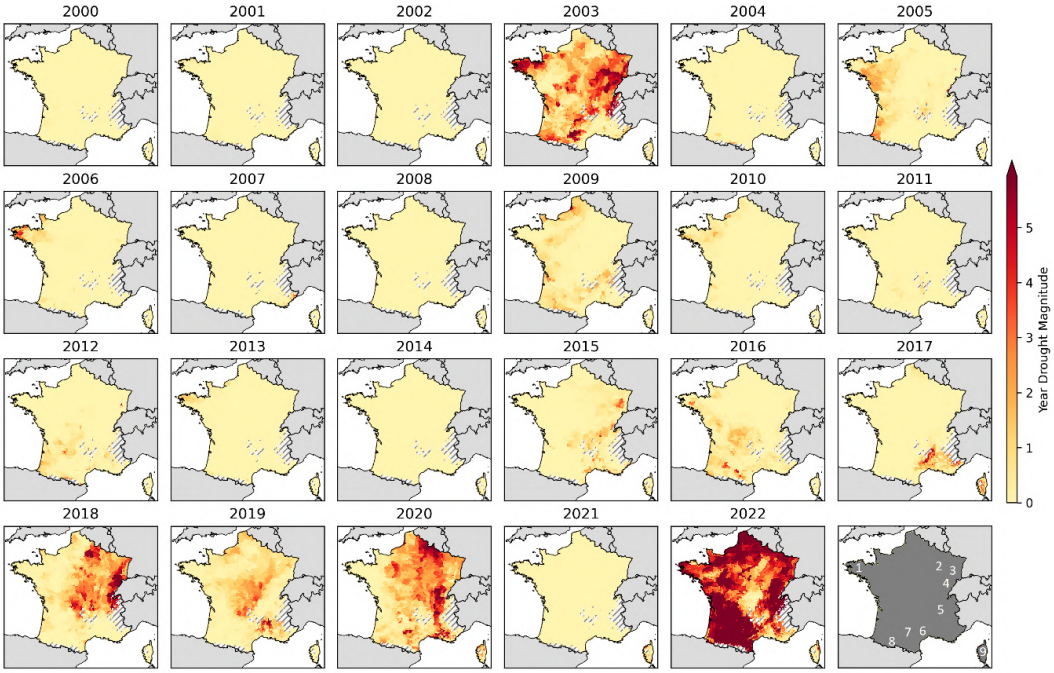


Figure 2: Year drought magnitudes for years 2000 to 2022, computed over France in each grid cell from the ISBA historical simulation. Areas with gray hatching correspond to filtered mountain areas (average altitude > 1100 meters). In the bottom-right subfigure the following areas are indicated by numbers, from 1 to 9: Brittany ("1"), Grand Est ("2", also see Fig. 6), Vosges ("3"), Jura ("4"), Rhône corridor ("5"), Cévennes ("6"), Occitanie ("7", also see Fig. 6), Pyrénées ("8"), and Corsica ("9").

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

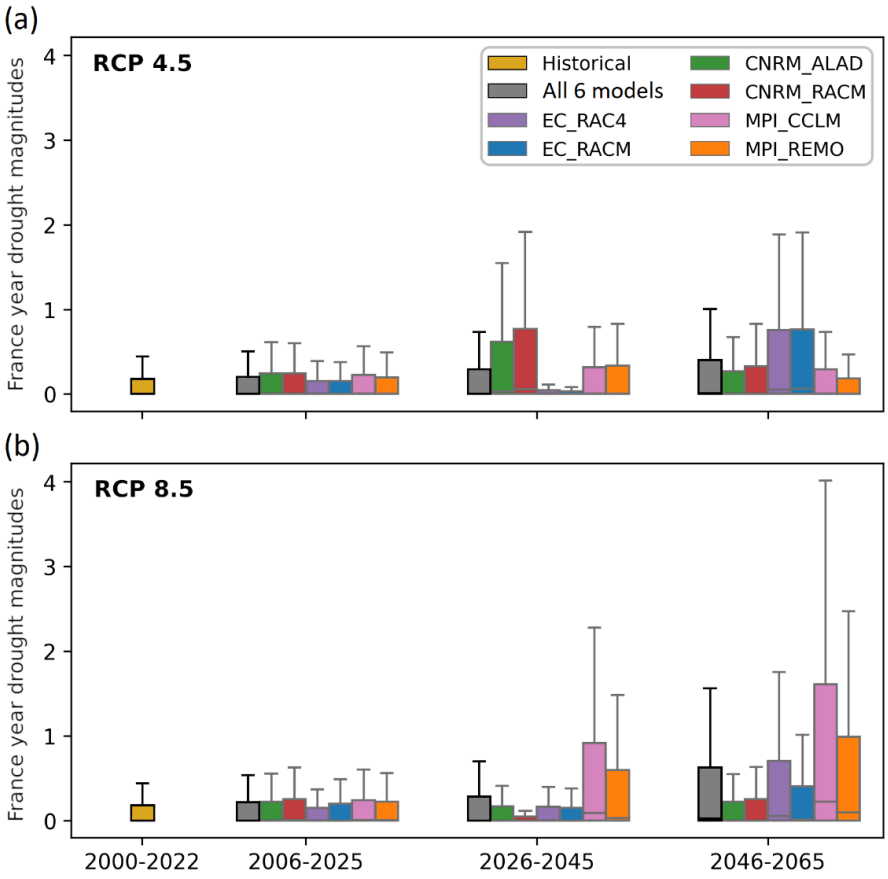


Figure 3: Projected YDMI distributions for (a) RCP 4.5 and (b) RCP8.5, separating time horizons and models. The grey box corresponds to the distribution of all models combined. The distribution of historical magnitudes is plotted as well (yellow box) for comparison. Box edges and center correspond to 25%, 75% and 50% percentiles. Whiskers extend to 1.5 times the interquartile range, equal to the difference between 75% and 25% percentiles. Outliers are not displayed.

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

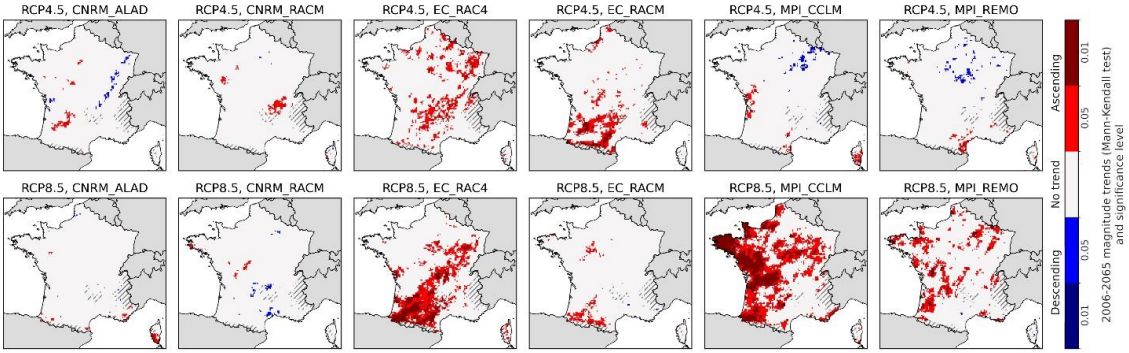


Figure 4: Results of the Mann-Kendall trend test applied to YDMI values from the 12 projected simulations, over the period 2006-2065. The ascending trends are marked with red color, while the descending trends are marked with blue color. The darker, the higher (smaller) the statistical significance (p-value). The two rows correspond to the two RCPs, and the six columns to the six models. Areas with gray hatching correspond to filtered mountain areas (average altitude > 1100 meters).

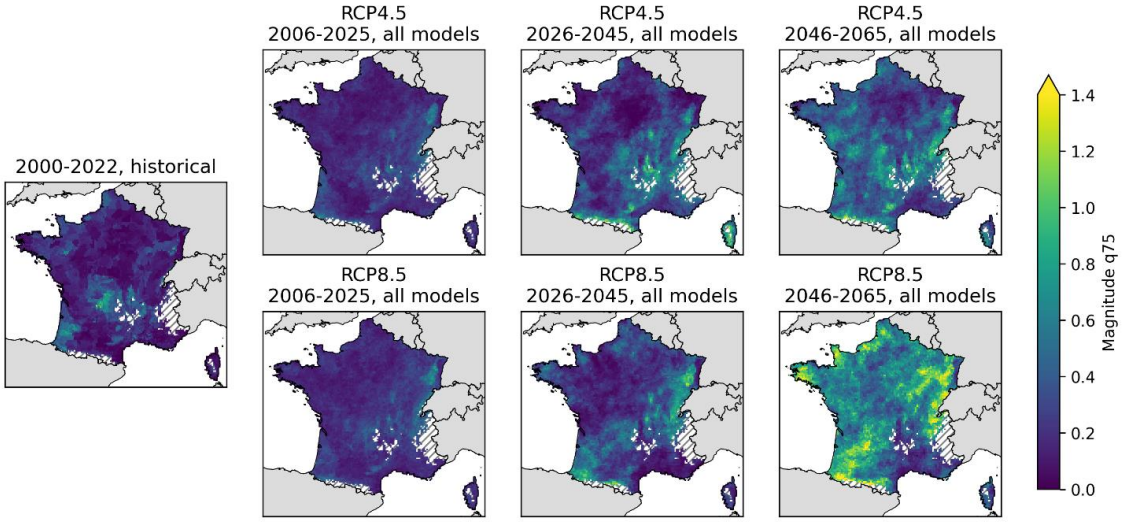
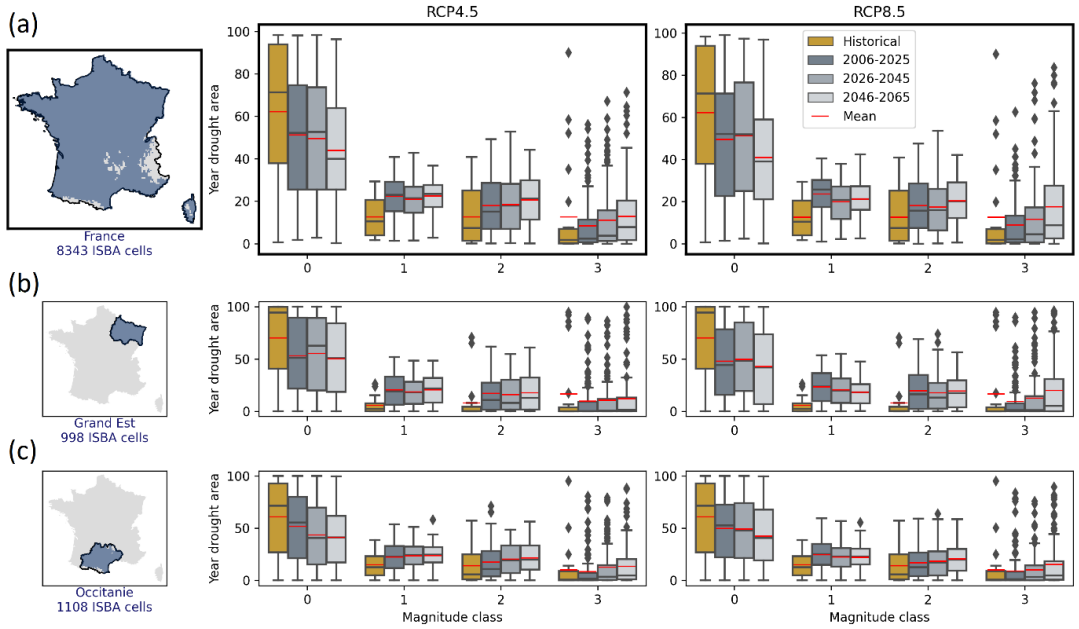


Figure 5: Third quartile (75th percentile) of YDMI, separating time horizon and RCP. Areas with gray hatching correspond to filtered mountain areas (average altitude > 1100 meters).

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.



785 **Figure 6**: YDMI class fractional area expressed as % of model grid cells, separating time horizon and RCP scenario, for (a) France, (b) Grand Est region, (c) Occitanie region. Whiskers extend either to the maximum or to 1.5 * the interquartile range, equal to the difference between 75% and 25% percentiles, if inferior to the maximum. Means and outliers are shown as red features and black diamonds, respectively

https://doi.org/10.5194/egusphere-2024-1079
Preprint. Discussion started: 8 May 2024
© Author(s) 2024. CC BY 4.0 License.

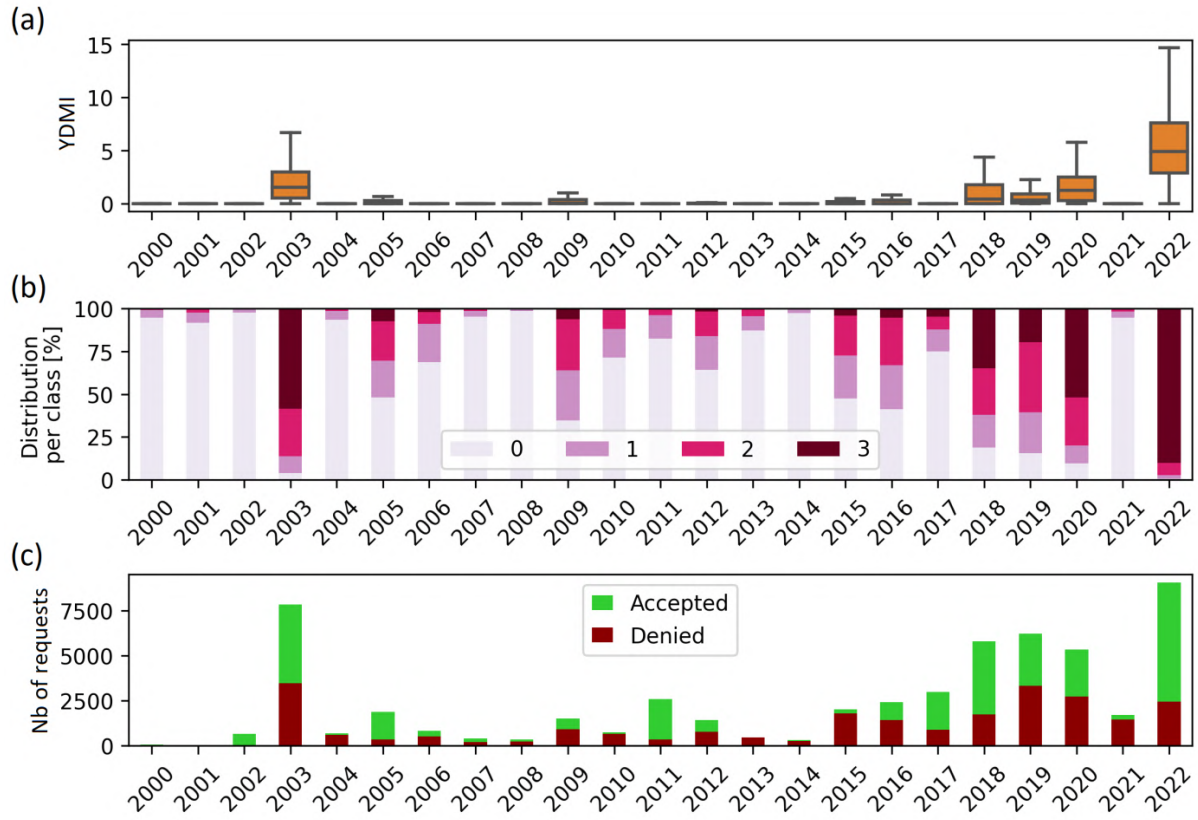


Figure 7: Year by year (a) statistical distribution of YDMI values, (b) fraction of YDMI classes (from 0 to 3), and (c) number of NatCat recognition requests, for France from 2000 to 2022. In (a), box edges and center correspond to 25%, 75% and 50% percentiles. Whiskers extend to 1.5 times the interquartile range, equal to the difference between 75% and 25% percentiles. Outliers are not displayed.

5.3 Drought temporality

A weakness of the YDMI is its coarse annual time scale, which prevents more detailed analysis of drought patterns. To explore this aspect, we assess the dominant drought month of each year between 2000 and 2022. This is done by counting the number of days below a drought threshold in each month, and keeping the month with the highest number. We consider the same thresholds than for computing optimal YDMI, which are the 1^{st} to the 5^{th} daily SWI distribution percentiles, and average over them. The results are shown in Figure 5.1. It is important to note that this figure only describes the timing of the drought, not its intensity. For a complete assessment, it should be read in conjunction with the YDMI maps.

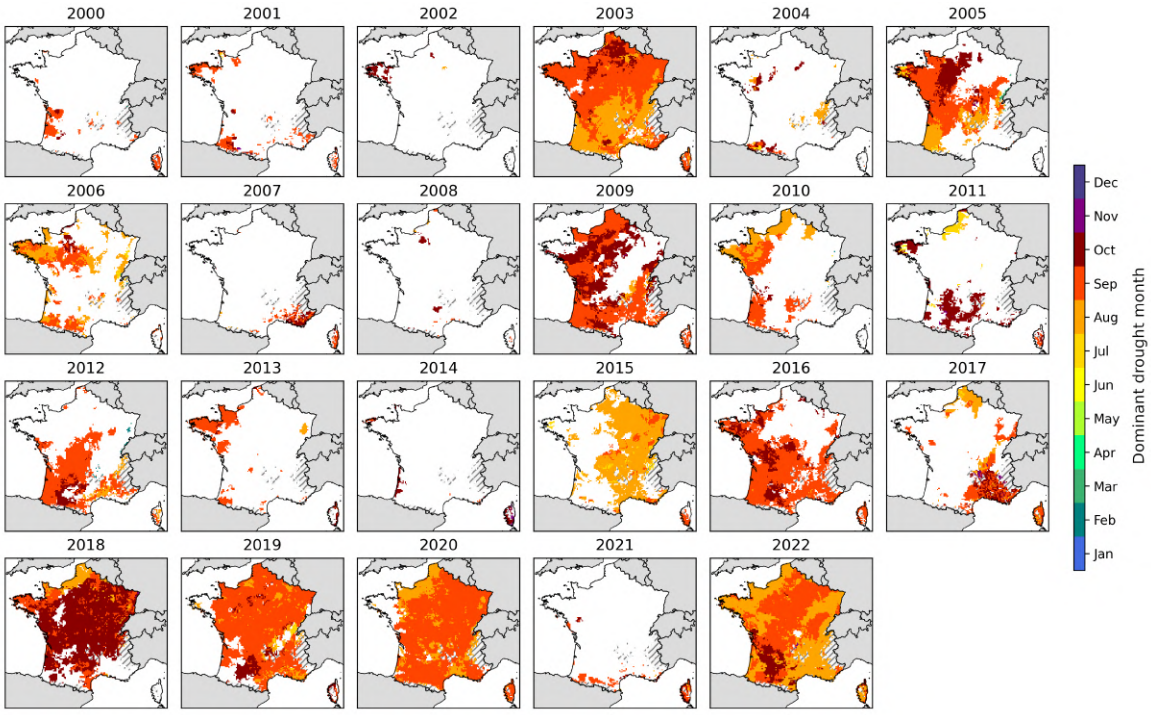


Figure 5.1: Dominant drought month (with the highest number of days below the 1^{st} to the 5^{th} daily SWI distribution percentiles) for each year, from 2000 to 2022

We find that almost all the droughts identified by the optimal YDMI occur mainly in late summer to early fall, from August to October.

We focus first on the year 2015, whose drought was not exceptional according to the YDMI, but was described as significant by the European Drought Observatory. The report [Micale et al., 2015] mentions a significant water deficit in the summer in eastern France. Positive YDMI were indeed observed in this area. Based the dominant drought month, the event peaked in August. It is likely that dry conditions did indeed occur in 2015, but did not last long enough to settle in depth.

Focusing on the 5 most significant years based on the YDMI (2003, 2018, 2019, 2020 and 2022), the 2018 event stands out in particular for its late timing, reaching its peak in October across France. This is in agreement with Masante et al. [2018]. In 2022, the drought lasted

until fall only in the southwest, where it was the most intense.

The development of the YDMI (chapter 3) demonstrated that moisture variations of deep layers are relevant for monitoring clay shrinkage-induced subsidence. There is a time lag between these variations and surface weather conditions, as observed at the Chaingy site (section 2.4.1). The droughts that reach deep layers integrate summer dry conditions with a time lag, which is why they are at their peak in late summer to early fall, just before the rehydration phase. This is what these maps show. This is consistent with the timing of clay shrinkage-induced damage, described in section 1.2.2.

5.4 Resampling YDMI

A second limitation of YDMIs is their coarse spatial resolution of 8 km, which can result in abrupt and therefore unrealistic transitions between two adjacent grid cells. This issue, directly linked to ISBA, is problematic in an operational context, as two neighboring municipalities can receive contrasting drought assessments.

To address this issue, we resample YDMI using bilinear interpolation. This technique consists in generating values at a finer spatial scale by performing a double linear interpolation on the x and y axes. We tested output resolutions of 4 km, 2 km, and 1 km. Figure 5.2 shows the results of the YDMI resampling for the Haute-Garonne department and the year 2016 (ISBA layer 8, and deciduous broadleaf tree patch).

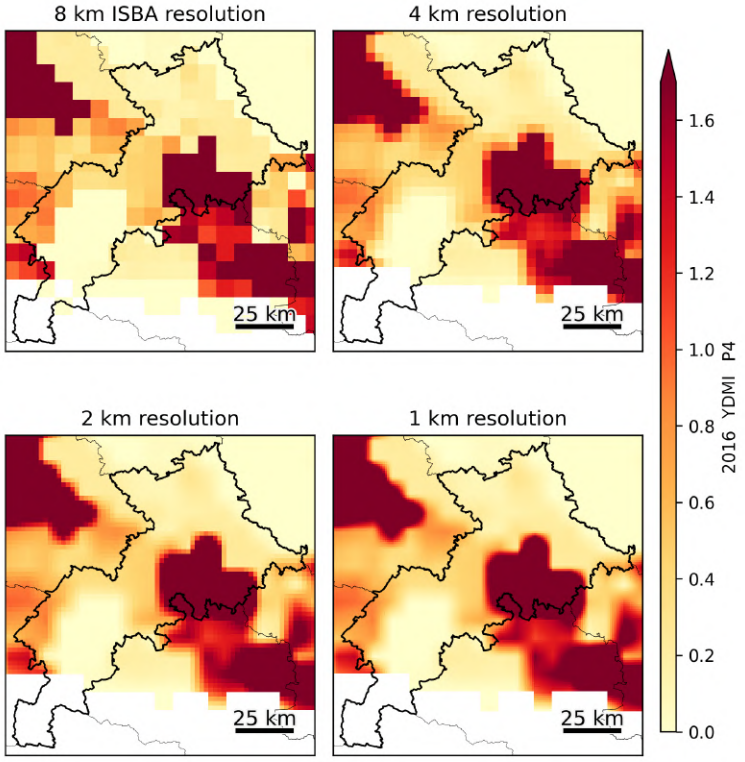


Figure 5.2: Testing different resolutions (4km, 2 km and 1 km) for resampling optimal YDMI with bilinear interpolation, year 2016 and Haute-Garonne.

The drought pattern at the initial 8 km resolution presents high contrasts between adjacent grid cells from the northwest to the southeast of the area. Resampling by interpolation blurs this pattern and smooths transitions, but results below 4 km become unrealistic because they replicate the 8 km grid pattern.

Therefore, we refine the YDMI to 4 km. The resampled YDMI is then sorted into classes based on the 8 km class boundaries. As a reminder, YDMI classes have been defined as follows: null YDMI are in class 0, and boundaries between classes 1, 2 and 3 are obtained by separating positive YDMI in three equal groups. Figure 5.3 shows 2016 YDMI over the country, before and after resampling, continuously and by class.

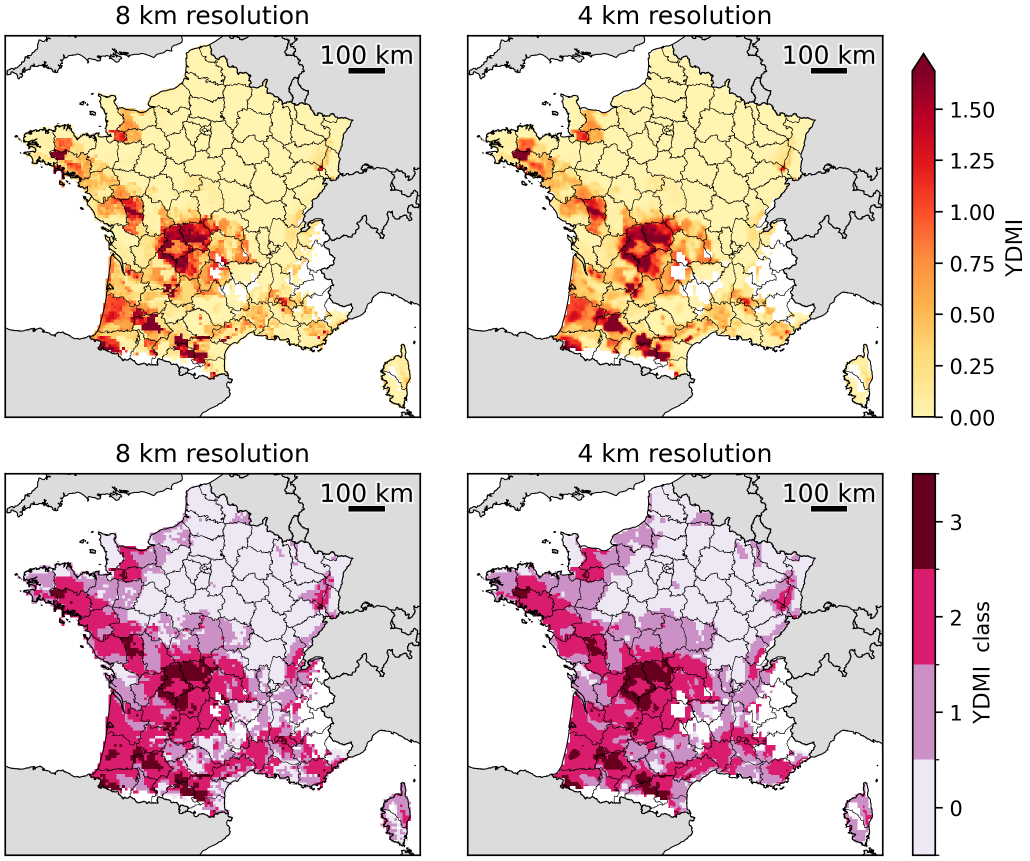


Figure 5.3: 2016 YDMI, before (left) and after (right) resampling to 4 km, continuous (top) and by classes (bottom). The YDMI are based the ISBA layer 8 and the deciduous broadleaf tree patch.

Consistent with what we observed for Haute-Garonne, resampling smooths the spatial pattern of drought, resulting in more gradual transitions between model grid cells. The changes are more pronounced for the YDMI in classes. In the southwest particularly, jumps of more than one class for adjacent cells are less frequent after resampling. However, the disadvantage of this method is that the influence of neighbors on the value of a cell dilutes the signal: for example, class 3 values in the southern Vosges are reduced to class 2 after resampling.

YDMI after resampling to 4 km for the years 2000 to 2022 are available in Figure B.2, continuous and by class. The linear interpolation technique used is based only on the four

pixels surrounding the value to be resampled. This is both a disadvantage, as the entire pattern is not considered when generating values, and an advantage, as only a small amount of data is lost around coastlines and filtered out in mountainous areas. It should be noted, however, that while this method is convenient for smoothing drought estimates, it is only statistical and has no physical basis: it does not take into account local geographic features that may affect soil water content, such as elevation, soil texture, and so on.

As the next chapter focuses on combining YDMI with hazard maps at a finer scale in order to better understand claim occurrence, we will examine whether increasing the resolution through bilinear interpolation has a positive impact.

5.5 Conclusion

In this chapter, we obtained optimal YDMI for a historical period and a future period under two distinct RCP scenarios. The historical analysis highlights the years 2003, 2018, 2019, 2020 and especially 2022 for their intense drought events. This identification is corroborated by external information such as European Drought Observatory reports. An additional study shows that droughts identified by the optimal YDMI last until fall. On the other hand, the projected analysis shows that the intensity and frequency of droughts likely to induce subsidence are bound to increase in the future. The trends are more marked for RCP 8.5 than for RCP 4.5. Nevertheless, these results are affected by a significant source of uncertainty: a mismatch is observed between past and projected YDMI, as projected YDMIs are unable to reproduce extreme events on the scale of 2022.

Historical YDMI will be used in the following chapter 6 as a proxy for clay shrinkage triggering conditions. They will be combined with hazard maps which quantify the susceptibility factor, with the goal of improving damage characterization. The predictive potential of YDMI interpolated at 4-km will be compared to that of YDMI at ISBA native resolution of 8-km.

Chapter 6

Merging drought index and susceptibility

Contents

6.1 Introduction

The previous chapters of this thesis have resulted in a new drought metric, the Yearly Drought Magnitude Index (YDMI), which provides an annual estimate of the intensity of the triggering factor for clay shrinkage. It represents a step forward on the issue, but is not sufficient by itself to fully characterize the hazard: it is the effect of drought on expansive soils that induces damaging volume variations.

The susceptibility of French soils to shrinkage and swelling was assessed by BRGM resulting in several maps described in detail in section 2.3. Combining the YDMI with such maps should

improve the characterization of the clay shrinkage hazard. However, it remains unclear how to proceed and what is the importance of each factor.

This chapter presents a machine-learning-based approach using a random forest model to predict subsidence claim occurrence from both the YDMI quantifying the triggering factor and the expansive soil maps representing the susceptibility factor. In particular, annual numbers of claims will be forecast from the model and commented, as this information is crucial to the insurance business. The model will be calibrated and validated using insurance claim and policy data issued from the ATHENA databases (section 2.2).

6.2 Methodology

In order to predict the occurrence of claims from soil and drought characteristics, these data first need to be merged in a single dataset. We refer to this dataset as the *policies* dataset. The successive steps to prepare this dataset are detailed in section 6.2.1. Its characteristics will be investigated in section 6.2.2. Then, the random forest model used for prediction will be detailed in section 6.2.3.

6.2.1 Preparation of the *policies* dataset

The different steps for preparing the *policies* dataset are summarized through the flowchart of Figure 6.1. They are commented below.

Policy extraction

The starting point is an extraction from the ATHENA insurance policy database. We query only policies associated with owners of houses, with a spatial precision at the address. Such a filtering is applied, because (i) houses are the most vulnerable type of building to subsidence, (ii) although both owners and tenants of a building are insured, claims are only associated with the owner's policy, and (iii) a precise location is required for a meaningful match with the expansive soil maps (which have a 100 meter accuracy). Each line corresponds to a policy for a specific year.

Merging with claims

The table of policies is then merged with another table containing information of clay shrinkage induced subsidence damage at the address, issued from the ATHENA claims database. A binary is associated with each policy, taking the value 1 if a claim was reported during the year and 0 otherwise. Duplicates are then deleted, and a filter on completeness is applied (with a 10% threshold on policy and claims- see section 2.2.1). As a reminder, completeness quantifies the extent to which the individual policy and claim data submitted by insurers is representative, in comparison to the figures they provide. Filtering on completeness is important for obtaining unbiased claim frequencies.

NatCat filter

Building damage caused by clay shrinkage is only compensated under the NatCat system, which requires that the town be recognized for the given year (section 1.5.1). If this condition

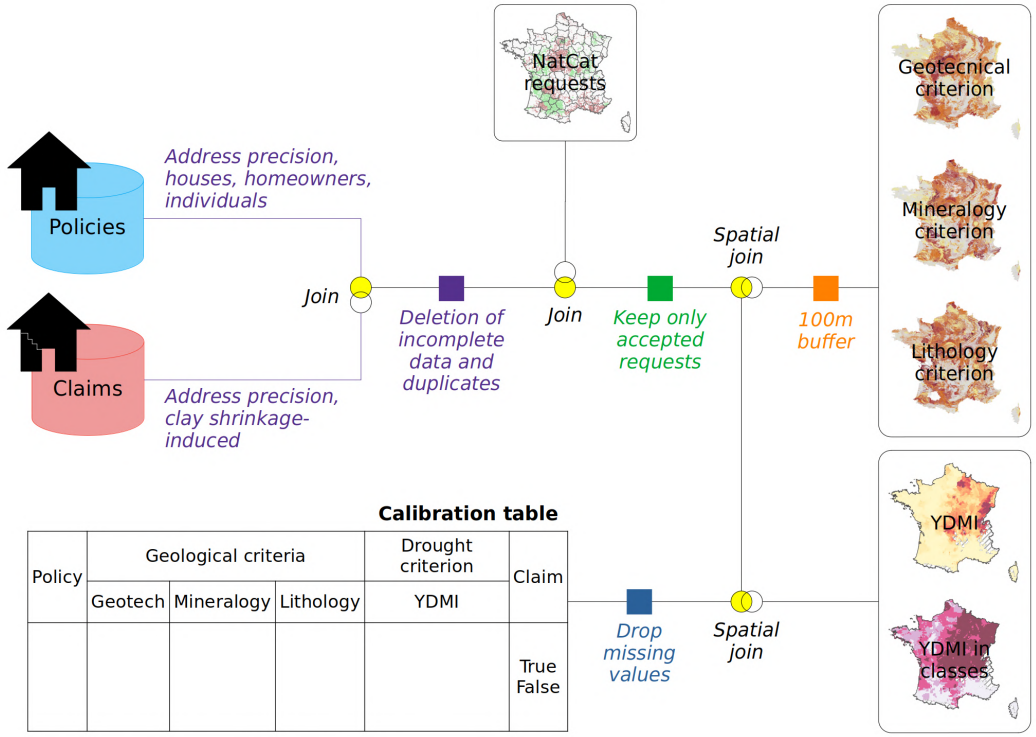


Figure 6.1: Preparation of the *policies* dataset

is not met, there is no compensation. To overcome this bias, our table is merged with all NatCat recognition requests, and only policies associated with accepted requests are retained. The filtering deletes data for 2001, a year associated to very few NatCat recognitions as shown in Figure 7 of Barthelemy et al. [2024b]. After this step, the table numbers 1,921,768 rows.

Spatial association with geological criteria

A first spatial join is done with the three geological criteria related to lithology, geotechnical properties and mineralogy, mapped by BRGM. Only these three criteria are selected, as exposure depends on historical claims and on susceptibility, which is calculated from the three criteria. Map precision is accounted for by applying a 100 m buffer zone to the polygons of geological units, and in case of overlapping the maximum value of the criterion is retained.

Spatial association with magnitudes

A second spatial join is done with drought magnitudes, continuous and in classes, (at the native 8 km resolution and re-sampled to 4 km- see section 5.4). Policies prior to 2000 are deleted as magnitudes cover the period from 2000 to 2022. This shifts the number of rows to 1,708,793.

Finally, missing values are removed, changing the number of rows from 1,708,793 to 1,550,848.

These missing values correspond to policies located on geological units that have not been characterized for either mineralogy, geotechnical properties, or lithology.

The *policies* dataset created by these successive steps therefore has 1,550,848 rows and 8 columns of data (3 geological criteria, 4 drought criteria, and a boolean for the occurrence of claims). The aim of this chapter being to use hazard criteria to better understand damage occurrence, the geological criteria and the YDMI will be used as predictors, and claims as the target variable. The modeling will be done on a policy-by-policy basis, as subsidence is a very localized risk [Charpentier et al., 2022].

6.2.2 Description of the *policies* dataset

6.2.2.1 Distribution among years

The distribution of the policies and claims of the *policies* dataset among years after the successive filtering steps is given in Figure 6.2. We notice differences for policies compared to the whole ATHENA database (in Figure 2.6), which is explained by the filtering on accepted NatCat requests - see the number of requests issued per year in Figure 7 of Barthelemy et al. [2024b]. There is no data for the years 2001 and 2017, and the policies are mainly concentrated over five years: 2011, 2016, 2018, 2019 and 2020, with more than 100,000 policies in each. On the other hand, the distribution of claims is similar to that of ATHENA.

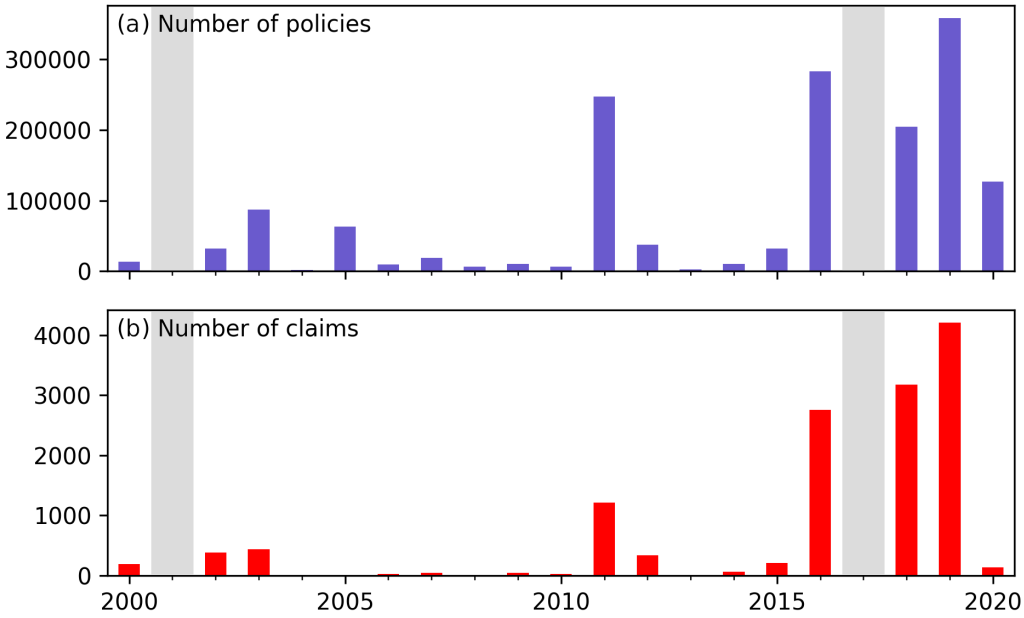


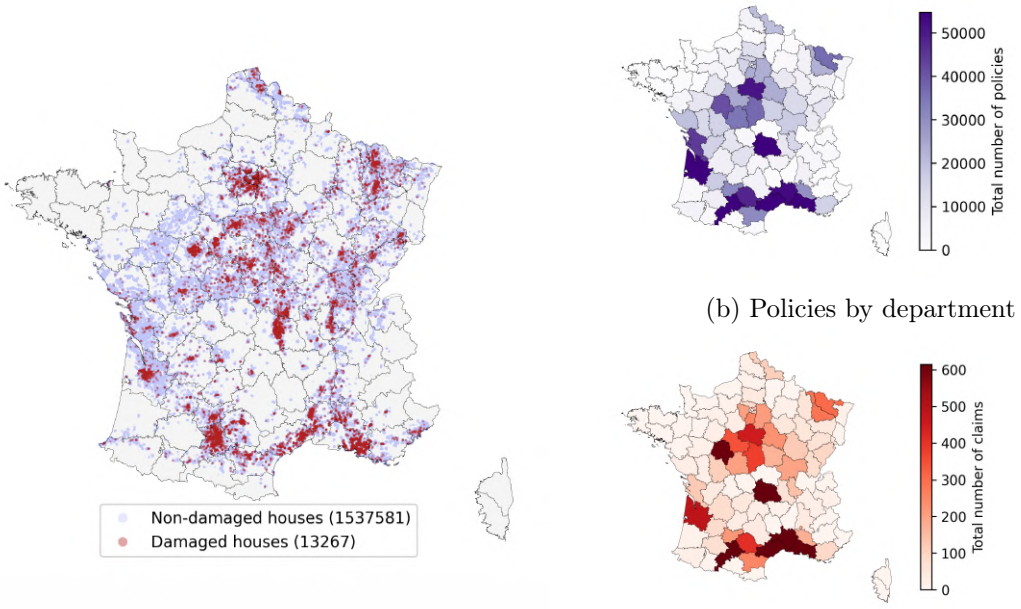
Figure 6.2: Distribution of the (a) policies and (b) claims of the *policies* dataset among years. The grey bars indicate years with no data.

In total, we count 13,267 damaged policies (0.86%) and 1,537,581 (99.14%) non-damaged policies. The target variable is therefore very imbalanced towards the absence of claims.

6.2.2.2 Spatial distribution

2000-2020

The spatial distribution of policies and claims of the *policies* dataset is shown in Figure 6.3.



(a) Distribution of policies and claims

(b) Policies by department

(c) Claims by department

Figure 6.3: Spatial distribution of (a) policies and claims, and (b-c) aggregation

Figure 6.3a shows that both policies and claims are concentrated around metropolitan areas: Toulouse, Marseille, Bordeaux, Clermont-Ferrand, Paris and Tours notably stand out.

The distribution of policies by department, in Figure 6.3b, differs from that in the ATHENA database (in Figure 2.5). In the *policies* dataset, the policies are mainly located in the south from Haute-Garonne to Bouches-du-Rhône, Gironde, Puy-de-Dôme and in the center-west of the country. In contrast, the distribution of ATHENA policies is similar to that of houses, with a large number of policies in the north and west of France. The distribution of claims is similar whether we consider all the data or only the *policies* dataset.

We explain this discrepancy concerning only policies by the filtering on accepted NatCat requests. Figure 6.4 shows the average number of accepted requests per town by department between 2000

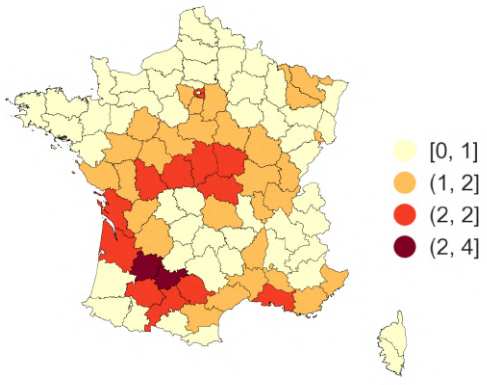


Figure 6.4: Average number of accepted NatCat requests per town by department 2000-2020

and 2020. It is higher in the southern departments and in the center-west of the country.

The distribution of policies in the *policies* dataset results from the total number of policies and the accepted decrees. For example, towns in the Nord department may have an important number of houses but they are poorly represented in the calibration dataset because they have not been much recognized within the NatCat system. Conversely, the Tarn-et-Garonne department has a comparatively small number of houses, but stands out in the calibration table because of the large number of accepted requests. The claims are not affected by this filtering step because their existence is already conditioned by an accepted request.

The filtering step based on accepted NatCat requests was implemented to avoid a bias resulting from the fact that claims can only be compensated following a recognition. However, it introduces an issue: the representation of spatial areas is not proportional to the number of houses they contain. For example, the northern departments are under-represented despite their significant number of houses.

Focus on years 2011, 2016, 2018, 2019 and 2020

As the distribution of policies between years is heterogeneous, we analyze the 5 years with the most data: 2011, 2016, 2018, 2019 and 2020. The spatial distribution of policies and claims of these five years is represented on Figure 6.5.

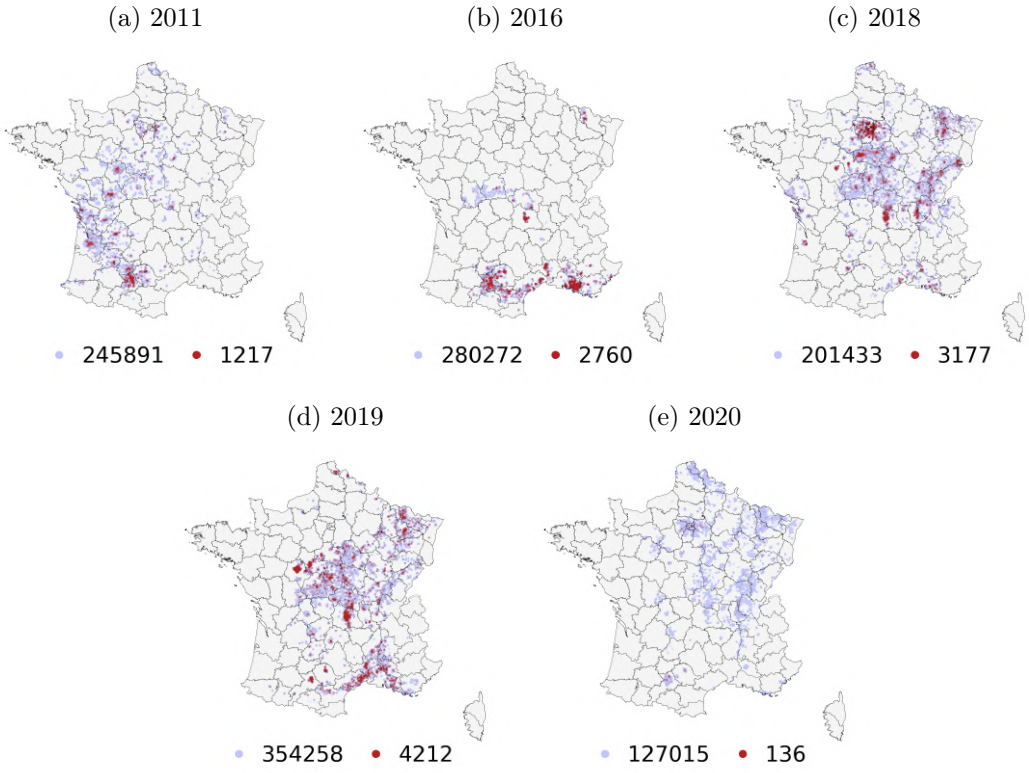


Figure 6.5: Distribution of the *policies* dataset for the most represented years. Purple markers indicate policies, and red markers indicate claims.

Policies and claims of each of these years present a different spatial distribution, which results

from the NatCat recognition step (see Figure A.2). The insurance policies of years 2011, 2016, and 2018 to 2020 are mainly located in the west, south and north-east of France, respectively. Sorting the years by increasing claim frequency, the year 2020 comes first with a frequency of 0.11%, followed by 2011, 2016, 2019 and finally 2018 with a claim frequency of 1.18%.

These maps demonstrate that our data show spatial heterogeneity in addition to temporal heterogeneity. Focusing on a single year is not relevant, as for each of them the data are concentrated in a specific region of France.

6.2.2.3 Distribution of policies in geological and drought classes

Each one of the geological criteria can take a value from 0 to 4, expressing increasing expansive potential. The correspondence between soil characteristics and criterion value is detailed in section 2.3. Class zoning has been also defined for YDMI (section 3.2). Class 0 designates null YDMI, and limits between classes 1, 2, and 3 have been determined by splitting positive YDMI into three equal groups.

Global distribution

The 1,550,848 policies of the *policies* dataset are distributed among classes for the different geological and drought criteria as in Table 6.1.

Table 6.1: Distribution of policies among classes of geological criteria and drought index

	Fraction of policies among classes					
Geological criteria	0	1	2	3	4	Total
Lithology	0.070	0.142	0.358	0.291	0.139	1
Mineralogy	0.070	0.209	0.332	0.286	0.102	1
Geotechnical	0.070	0.067	0.511	0.302	0.050	1
Drought index	0	1	2	3		Total
8-km YDMI	0.279	0.160	0.286	0.275		1
4-km YDMI	0.223	0.185	0.311	0.281		1

The purpose of this table is to assess whether each class of each criterion concentrates a sufficient number of policies to draw conclusions from later on. This is especially important since our dataset is very unbalanced in favor of the absence of claims (average ratio of less than 1 claim per 100 policies). Classes of all criteria are well represented in the dataset as each of them concentrates 5% or more of all policies.

The fraction of policies outside clayey areas (geological class 0) are identical because the three criteria are based on the same contours. The distribution of policies within geotechnical classes differs from that of the mineralogy and lithology criteria, suggesting that the information provided by this criterion is less redundant. For the 8-km YDMI and the 4-km YDMI, the distribution of policies among the classes is homogeneous with the exception of class 1, which is less represented.

Annual distribution

We also analyze the annual distribution of policies across classes of geological criteria and drought index. They are plotted in Figure 6.6.

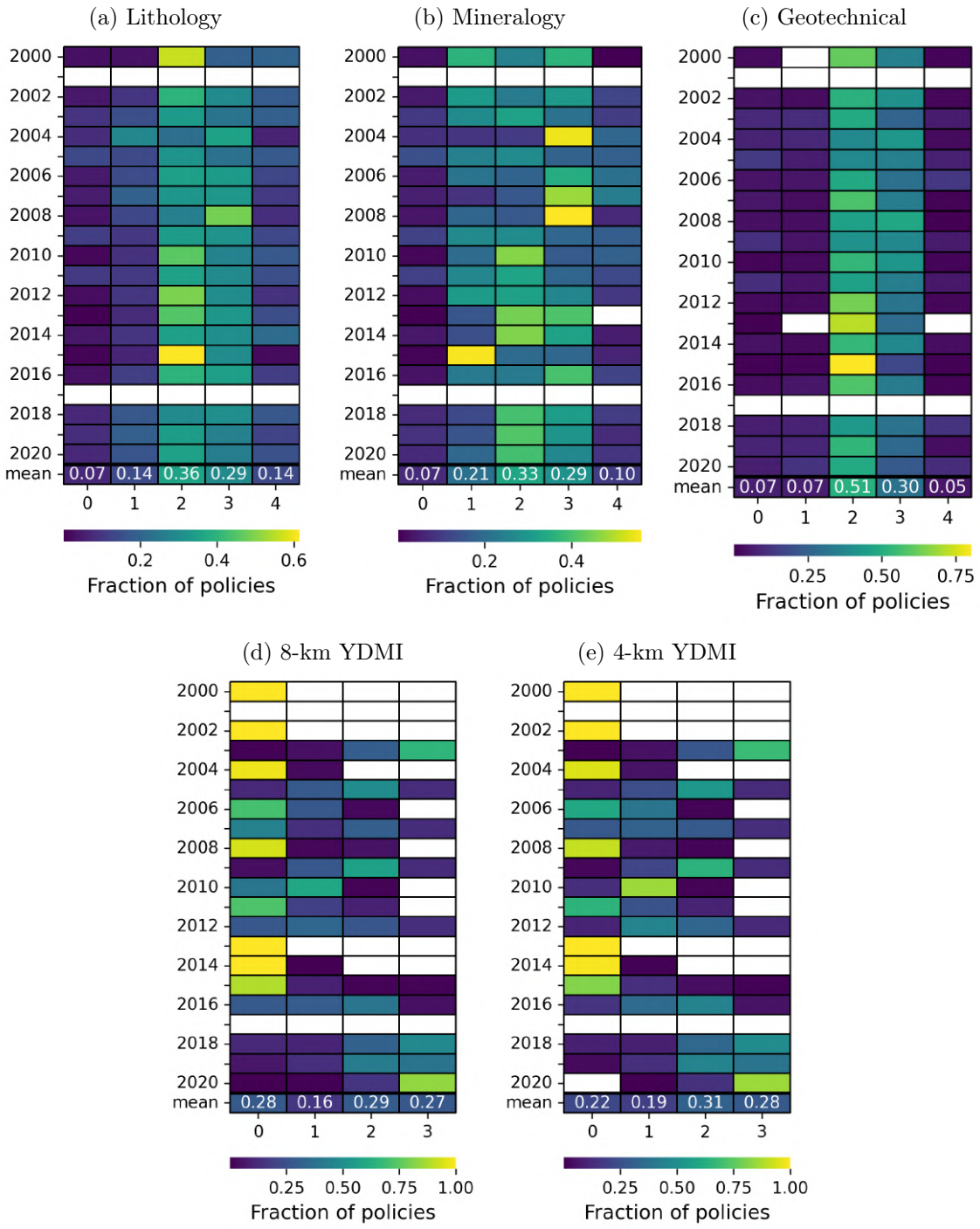


Figure 6.6: Annual distribution of policies by geological criteria and YDMI class. The colorbar shows the fraction of policies in each class, with the sum of each line equal to 1. The last line shows fractions for the entire dataset.

This analysis is crucial for understanding results better, since we will be making annual predictions.

The annual distributions of policies within classes of geological criteria are similar to those for the entire dataset. Only a few years differ. For example, in 2015, the vast majority of policies were in lithology class 2, mineralogy class 1, and geotechnical class 3. The slight differences between years are explained by the varying locations of the policies induced by the filtering step on the NatCat decrees.

On the other hand, the annual distributions of policies within YDMI classes are more heterogeneous, which was expected since it is a dynamic predictor and not a static one like the previous. The results are very similar for the 8 km and 4 km YDMI, as they are highly correlated. The years 2000, 2002, 2004, 2008, 2013, 2014 and 2015 stand out for their high fractions of policies associated with null YDMI, indicating the absence of a major drought event during these years. In contrast, notable drought events occurred in 2003, 2005, 2009, 2018, 2019, and 2020 based on the fraction of policies in superior YDMI classes.

6.2.2.4 Claim frequency in geological and drought classes

The claim frequency per class for the different geological and drought criteria, estimated by dividing the number of claims by the number of policies, is given in Table 6.2.

Table 6.2: Claim frequency per classes of geological criteria and drought index

	Claim frequency per class [%]				
Geological criteria	0	1	2	3	4
Lithology	0.37	0.55	0.72	1.07	1.32
Mineralogy	0.37	0.49	0.80	1.17	1.22
Geotechnical	0.37	0.50	0.80	1.08	1.21
Drought index	0	1	2	3	
8-km YDMI	0.64	0.86	0.95	0.97	
4-km YDMI	0.55	0.93	0.92	0.98	

Damage frequencies are positively correlated with classes. This effect is more pronounced for geological criteria than for YDMI. They exhibit maximum variation ranges of 0.37 to 1.32 and 0.55 to 0.98, respectively. This was expected and shows that claims are more frequent in higher geological and YDMI classes, hence the relevance of the approach.

However, we did not expect positive claim frequencies to be associated with class 0 of geological criteria or with null YDMI. Indeed, no damage is expected outside of clayey areas, or in the absence of a drought. The number of claims is 402 for class 0 of geological criteria, and 2792 and 1886 for 8-km YDMI and 4-km YDMI, respectively. We can explain claims occurring outside of clayey areas by the scale of the hazard maps: their precision being of 100 m, very local clayey zones may be unidentified. This point will not be explored further in this study, although it deserves to be. On the other hand, claims associated to null YDMI will be investigated in section 6.4.1.1.

6.2.3 Classification model for claim prediction

The goal of the chapter is to use the information conveyed by the soil and drought characteristics to forecast damaged policies. This is actually a classification problem that relies on both discrete and continuous predictors, called *features* in this context. Since the target variable (claim occurrence) is known, it can be used to train the model using a supervised learning approach.

Classifier

We consider a random forest model. This type of model, widely used for classification, has the advantages of (i) being robust due to its ensemble nature, (ii) requiring no assumption about the distribution of the data, and (iii) handling correlated features. The three geological criteria and the 8-km and 4-km YDMI variables are chosen as features, and the binary damage variable is chosen as the target. The concept of the random forest model is explained section 6.2.3.1.

Probabilistic predictions

Given the characteristics of our dataset, i.e. many policies, most of them undamaged, it is more appropriate to obtain probabilities of belonging to a class rather than binary predictions. The number of claims is obtained by summing the probabilities.

In fact, any model that makes binary predictions, given such data, would predict only no claims and still have a majority of successful predictions. There are strategies to correct for this behavior by focusing predictions on the minority class (e.g., oversampling, undersampling, or even choosing weights), but they would result in a drastic overprediction of claims, which is equally problematic.

To illustrate this point, consider a sample of 1000 policies with identical drought and soil conditions, only one of which is damaged. A binary prediction could result in either 0 or 1000 claims, when summing probabilities of 0.001 would allow the correct number to be obtained.

Experiments

Probabilistic predictions will be made for each policy and summed to give the number of claims per year. The quality of the predictions will be evaluated using performance metrics, which will be described in section 6.2.3.2. The metric related to features will be presented in section 6.2.3.3. Several experiments will be done, iterating on features. They will be described in section 6.2.3.4.

6.2.3.1 Random forest model

Random forests, introduced by Breiman [2001], is a supervised ensemble machine learning technique based on decision trees that can be applied to both classification and regression problems.

Decision tree

A decision tree functions similarly to expert judgment, making a decision based on a superposition of conditions. It works by iteratively splitting an initial dataset into groups to separate the different instances of the target variable. The groups of data in the tree are called nodes. The initial data is called the root node, and the final groups are called leaf nodes. At each split, the parent node is separated into two child nodes based on a condition applied to a feature. If a node contains only one instance of the target variable, it is called pure. Otherwise, the node is impure. Once the tree is built, it assigns to any combination of features a probability of being in a class equal to the fraction of samples of the same class in the corresponding leaf. This probability can be converted into a binary prediction.

Impurity criterion

A decision tree is not generated randomly. The condition on which the split is based is obtained by optimizing a specific criterion. The criterion can either be a measure of the probability of misclassifying a sample (Gini) or of the disorder (entropy) within the resulting nodes. The predictive capacity of a feature is based on its *importance*, which is calculated as the normalized total reduction of either the Gini or entropy criterion brought by the feature.

Random forest

Random forests consist in a *bagging* (bootstrap aggregation) technique applied to decision trees. Samples are generated by bootstrapping (i.e., random sampling with replacement) from a dataset, and a decision tree is constructed for each sample. For a classification problem, the model's output is the mean predicted class probability, or the majority vote of all trees in the case of a binary prediction. Random forests belong to ensemble methods, based on the concept of the *wisdom of crowds*, which posits that the average prediction of numerous weak learners is more reliable than that of a single expert. Each tree is a relatively weak learner, as it is trained on a subset of the data. The aggregation of all trees is supposed to be robust enough to prevent over-fitting.

Example

To illustrate the approach, we've built a decision tree of limited depth in Figure 6.7 from a sample of the data, that we will explain below.

The tree is constructed from a random sample of 10,000 policies from the the *policies* table, considering only the mineralogy class and YDMI as features. The two classes labeled as 0 and 1 correspond to no claims and claims. Out of the 10,000 policies, 9913 are undamaged and 87 are damaged. We select here the entropy impurity criterion. The algorithm chooses itself optimal conditions at each split as to minimize final leaf impurity based on this criterion.

The first split separates policies based on their YDMI score: null YDMI on the left, and positive YDMI on the right. This condition "purifies" the left group as its entropy is equal to 0.04 in contrast to the initial entropy of 0.07. The second split divides the two groups into four groups based on a mineralogy criterion. For null YDMI (left node), a condition of mineralogy = 0 is applied, resulting in the separation of 224 undamaged policies. This node is pure, as indicated by the entropy of 0. In contrast, the other nodes still have positive

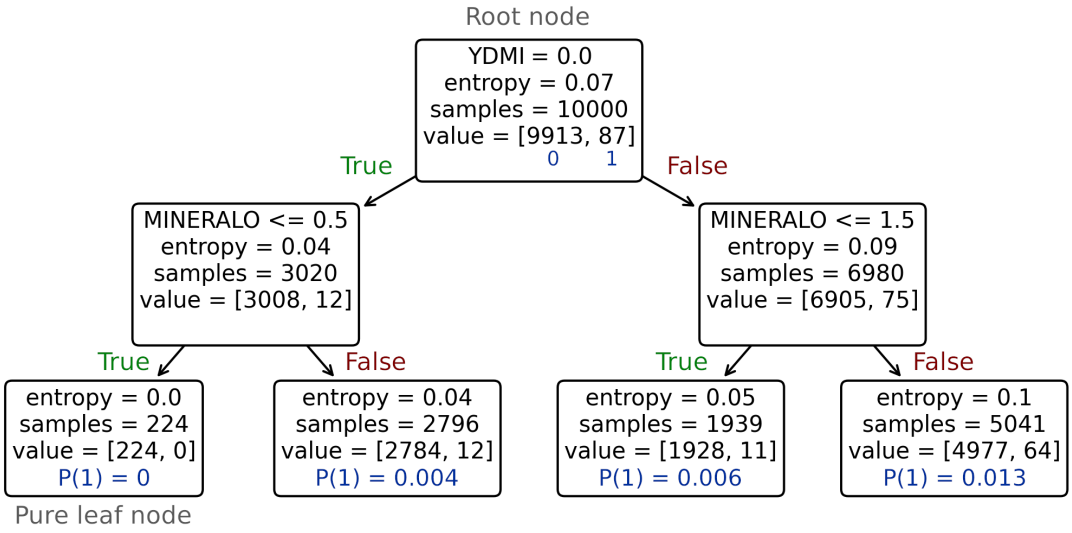


Figure 6.7: Example of decision tree created from a random draw of 10,000 policies from the *policies* dataset

entropies. In this simplified example, we have set the maximum depth of the tree to two splits.

The probability of a policy to be damaged is equal to the proportion of damaged policies in the corresponding leaf node. In this specific example, a policy associated with a YDMI of 1 and a mineralogy score of 3 has a probability of 0.013 to be damaged. Although we have limited the maximum depth of the tree here to 2, the more splits there are, the more possibilities there are for different predictions. In this example, the feature importance of the mineralogy criterion and the YDMI are 0.42 and 0.58, respectively, quantifying the superior ability of YDMI-based conditions to separate policies.

6.2.3.2 Performance metrics

By summing the probabilities predicted by the model, it is possible to obtain numbers of claims per year. The quality of the predictions is measured using two different metrics. First, the accuracy of individual predicted probabilities is assessed using the ROC AUC metric. Then, the annual predictions are assessed by calculating the root mean square error (RMSE) between actual and forecast claims. These metrics are described in detail in this section.

Evaluating binary classifiers

There are only 4 possible prediction outcomes for a binary classification model: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). They are summarized in a confusion matrix, the concept of which is shown in Table 6.3. Usually, scores are derived from the confusion matrix. This is not done here because we are deriving probabilities and not binary predictions from the models. However, it is still possible to compute scores from probabilities using ROC analysis.

ROC analysis

The Receiver Operating Characteristic (ROC) curve [Fawcett, 2006] evaluates the performance of binary classifiers by testing different class assignment thresholds from probabilities. A curve is created by plotting two scores obtained by varying the class assignment threshold. These two scores are the true positive rate (TPR), which corresponds to the proportion of correctly identified true positives, and the false positive rate (FPR), which is equal to the number of false positives divided by the actual number of negatives. Examples of ROC curves are given on figure 6.8. The closer the classifier is to the point (0,1), the better its performance. On the contrary, if it coincides with the diagonal, this means that it is equivalent to a random classifier. The ROC curve can be summarized by the area under curve (AUC) metric. The AUC of a perfect classifier is 1.0, while that of a random classifier is 0.5.

Table 6.3: Confusion matrix

		Observed	
		Yes	No
Predicted	Yes	TP	FP
	No	FN	TN

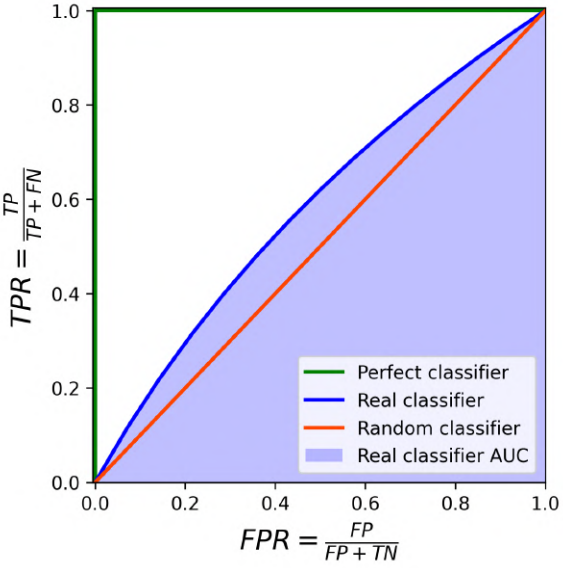


Figure 6.8: ROC curve examples

Root mean square error

The RMSE is an error metric used in machine learning to assess how well predictions match actual data. It will be used to evaluate the forecast number of claims. The RMSE consists of the square root of the average of squared errors between model and observations, as detailed in Equation 6.1. The lower the RMSE, the better the prediction, with a minimum of 0 obtained for a perfect match between observation and forecast. No distinction is made between positive and negative errors. The main advantage of RMSE over other metrics is that it has the same unit as the measured variable, making it easy to interpret.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{6.1}$$

$$y_i : \textit{ actual values}, \ \hat{y}_i : \textit{ predicted values}$$

$$n : \textit{ number of observations}$$

6.2.3.3 Feature importance

While the metrics presented above describe the predictive capacity of the models, the feature importance relates to the weight of each feature. It is determined during the training phase. The feature importance is specific to the random forest model, and is equal to the total normalized impurity reduction brought by features, i.e. their ability to separate instances of the target variable within trees.

6.2.3.4 Experiments and validation strategy

Experiments

We run different experiments, iterating over the features to assess their respective contributions to the forecast. A first experiment is performed with all (5) features, then with the geological criteria only (3 features). For each of these experiments, we select the entropy criterion for building the random forest. The Gini has the advantage of being faster to compute, but is not well suited to data with highly unbalanced classes.

Repeated Stratified K-Fold validation

We use a *Repeated Stratified K-Fold* validation strategy. The k-fold approach consists in dividing the dataset into k equal parts: the model is iteratively calibrated on k-1 parts and validated on the remaining part. We use 5 folds, which means that 80% of the data is used for training and 20% for validation. Since the target variable is highly unbalanced, we choose a stratification strategy: the k-folds are built by preserving the initial proportions of the target variable. This validation method enables to model to be trained and then validated on homogeneous data, which is an optimal context. It however does not correspond to the operational situation in which the model has to predict damage for an unseen year whose features probably differ from that of the entire dataset.

6.3 Results

A random forest model was iteratively trained and validated on the *policies* dataset using the stratified 5-fold approach, with geological features only (3-feature case) and with both geological criteria and YDMI as features (5-feature case). We present in this section the resulting performance and feature scores. Annual numbers of claims and claim frequencies will also be commented.

ROC AUC score

We obtain ROC AUC scores equal to 0.657 and 0.626 in the 5-feature and 3-feature cases, respectively. Scores vary little from fold to fold. These scores indicate that including the YDMI is a relevant feature for characterizing policies. The model performs better than a random classifier, but that there is still room for improvement when it comes to identifying whether or not a policy is damaged.

Feature importances

The random forest feature importances are indicated in Table 6.4.

Table 6.4: Average random forest feature importances over the 5 folds. Best features of each group are indicated in bold.

Number of features	Geological criteria			Drought indices	
	Lithology	Mineralogy	Geotechnical	8-km YDMI	4-km YDMI
5	0.037	**0.044**	0.032	0.381	**0.507**
3	0.361	**0.445**	0.194		

Feature importance scores reveal a significantly higher predictive capacity for climate factors than soil susceptibility factors. This recalls the conclusions of Corti et al. [2009], which identified a strong influence of meteorological forcing on the inter-annual variability of subsidence damage. As a reminder, the sum of feature importances is normalized to 1, which explains the different orders of magnitude in the 3-feature and in the 5-feature experiments. The ranking of geological features is unchanged without the YDMI. Of the two types of features, the method yields the highest scores for the mineralogy criterion, and for the 4-km YDMI.

Annual numbers of claims

Summing the probabilities of policies to be damaged for each year gives the annual number of claims in Figure 6.9. The number of claims are lower than that of the entire *policies* dataset (in Figure 6.2), because it corresponds here to the iterative 20% of the data used for validation in the 5-fold approach.

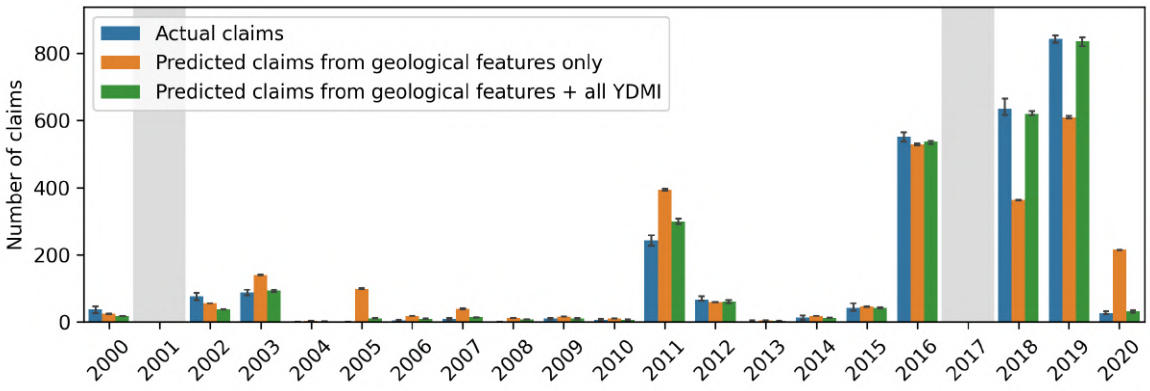


Figure 6.9: Number of claims per year, actual and predicted by the random forest, for 20% of the data. The error bars show dispersion for the 5 folds. The grey bars indicate years with no policies.

First, the error bars show that actual claims are distributed across years similarly to the initial claims in each iteration. They also show that there is little variability between the folds for the predictions. This indicates that the methodology is robust.

Consistent with the high importance of the YDMI, we find that its use as a feature has a significant impact on the annual number of claims. For example, using the YDMI for prediction leads to a 33% increase in the predicted number of claims in 2019 (600 to 800

claims), which is in line with the actual number. Visually, a better fit is obtained with observations in the 5-feature case than in the 3-feature case. RMSE values, of 17 in the former and 103 in the latter, are corroborate this interpretation.

However, a counterintuitive element is the large inter-annual variability in the number of claims predicted solely on the basis of geological features, which have been shown to vary little from year to year (see Figure 6.6). In fact, this variability is due to the heterogeneous distribution of policies between years (in Figure 6.2).

Annual claim frequency

To erase this effect, we analyze claim frequencies which are equal to the number of claims divided by the number of policies, plotted in Figure 6.10. This is equivalent to a normalization.

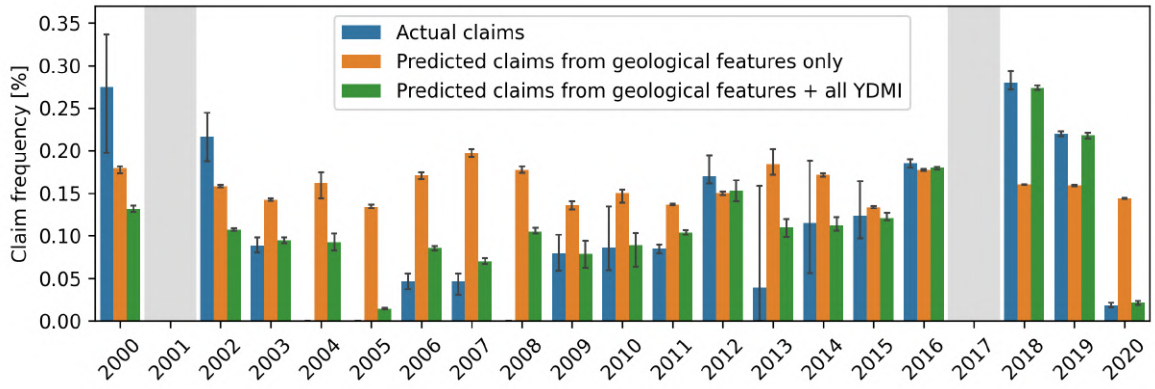


Figure 6.10: Annual claim frequencies, actual and predicted by the random forest, for 20% of the data. The error bars show dispersion for the 5 folds. The grey bars indicate years with no policies.

We notice that the claim frequency is similar every year in the 3-feature case, fluctuating around 0.15%. The small variability is attributed to the annual distribution of policies in the classes: for example, the maximum frequency of 0.20% in 2007 is explained by the fact that in this year a higher proportion of policies are located on soils with expansive properties (as seen in Figure 6.6). Therefore, the inter-annual variability in the number of claims in the 3-feature experiment is indeed due to the heterogeneous distribution of policies by year.

On the other hand, the inter-annual variation in claim frequency in the 5-feature experiment is of the same order of magnitude as the actual claim frequencies. Comparing the effect of YDMI on claim frequencies with the annual distribution of policies in YDMI classes reveals different situations:

- For several years (2004, 2006, 2007, 2008, 2011, 2018, and 2019), the effect of YDMI on claim frequency is physically consistent, and improves the agreement with actual claims. For example, in 2004, almost all policies were associated with zero YDMI. Using it as a feature reduced the claim frequency, narrowing the gap with the actual claims. Conversely, in 2018, the significant proportion of policies in higher YDMI classes led to an increase in the claim frequency, bringing it closer to the actual data.

- In some cases (2000, 2002), the effect of YDMI on the number of claims is consistent but degrades the agreement with observations. For instance, in 2000, there was no drought based on YDMI, which led to a decrease of the claim frequency, when the actual claim frequency was surprisingly higher.
- In a final case (2003, 2005, 2009 and 2020), the effect of the YDMI improves the agreement with damage frequencies, but is physically inconsistent. For example, in 2020, the YDMI reduced the frequency of claims even though it was a record year for drought. These four years have in common that notable droughts based on the YDMI are associated with lower than average claim frequencies. This suggests that this inconsistency is a sign of overfitting: the model is learning and replicating a counterintuitive pattern that exists in the data.

6.4 Discussion

In summary, our random forest model trained on YDMIs and geological criteria correctly predicts the annual number of claims, although it performs only slightly better than a random classifier on a policy-by-policy basis. We show that YDMIs contribute significantly to model predictions and improve the match with actual claims. However, inconsistencies emerge as the model replicates counter-intuitive patterns inherent to the data. This suggests that this model, while promising, is not yet ready for operational use.

In this section, we will discuss the causes of the low values of the ROC AUC score, which shows the difficulty of predicting whether an insurance policy is damaged or not based solely on soil and drought characteristics.

6.4.1 How can we explain the low ROC AUC score ?

Apart from the weaknesses of the YDMI, such as the fact that it does not take into account the recurrence of drought events, two points help to explain the low ROC AUC score. These are the catch-up effect of the NatCat system and the lack of information on local characteristics of insurance policies.

6.4.1.1 Catch-up effect of the NatCat regime

An unexpected element was noted earlier in the chapter: a significant number of claims is associated to null YDMIs. The 1886 claims associated to null 4-km YDMI will be analyzed to find out what lies behind. These claims are spatially distributed as shown in Figure 6.11.

We observe that 6 years account for almost 90% of the claims associated to null magnitudes: 2000, 2002, 2011, 2015, 2016 and 2018. In particular, 26% of all false negatives occur in 2011. We will investigate this year in particular: Figure 6.12 shows the history of YDMI and NatCat recognition requests from 2009 to 2011.

In 2011, 91% (451 out of 497) of false negatives were concentrated in the western area of the country, distributed among 249 towns. While western France exhibited null YDMI in 2011, this was not the case in 2009. Also, in 2009 and 2010, 136 and 93 Natcat recognition requests

were submitted by the 249 towns, respectively, with 28% and 14% of these requests being accepted. All 249 towns were recognized in 2011, resulting in the reported claims.

The high YDMI in 2009 associated to the low Natcat request acceptance rate indicates that some of the false negatives reported in 2011 may have been caused by the climatic conditions of 2009. For this to happen, the meteorological criterion leading to recognition of the municipalities would have been met in 2011 but not in 2009. This criterion does not account for the same drought characteristics than the YDMI. The delay would then be explained by the meteorological criterion of the NatCat regime.

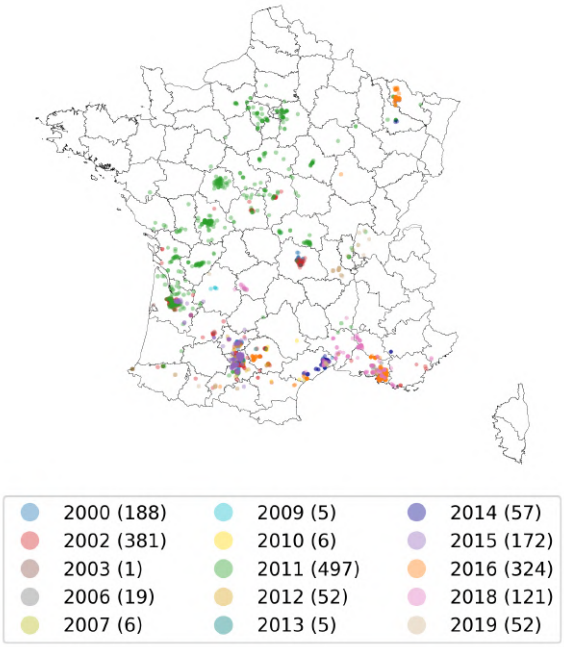


Figure 6.11: Spatial distribution of claims associated to null 4km-YDMI

However, this only partly explains the high number of false negatives in 2011: only 166 towns have submitted a request in 2009 or 2010, which remains inferior to the 249 towns of 2011. In addition to this effect, it is likely that the YDMI does not accurately reflect the full geographic extent of areas experiencing significant drought conditions. In line with this point, 2011 is described as a year with a particularly dry spring, in the southwest to west of the country [CCR, 2022], while the YDMI only identifies something in the southwest. It's possible that the YDMI does not identify the drought in 2011 because it did not extend to the depth of 1 m, being a spring event. It is relevant to wonder if water content variations in shallower layers should be considered, provided they are smoothed out to eliminate short-term variations.

Based on this demonstration, we identify a temporal bias probably due to the NatCat recognition criterion, which questions the temporal accuracy of the damages indexed by the CCR databases. It contributes to explain why both models struggle to make correct predictions at the policy scale. This is also in line with the slow kinetics of the phenomenon, whereby cracks in buildings are often not noticed immediately by homeowners, resulting in a delay in their reporting.

6.4.1.2 Insufficient predictors

The obtained values of the ROC AUC metric show the difficulty of the model in predicting whether an insurance policy is damaged or not based solely on its soil and drought characteristics. This is not surprising given our data set.

Analyses at the sub-municipal level have shown that during drought events, only a very small proportion of buildings within the same geographical area were damaged. For example, in

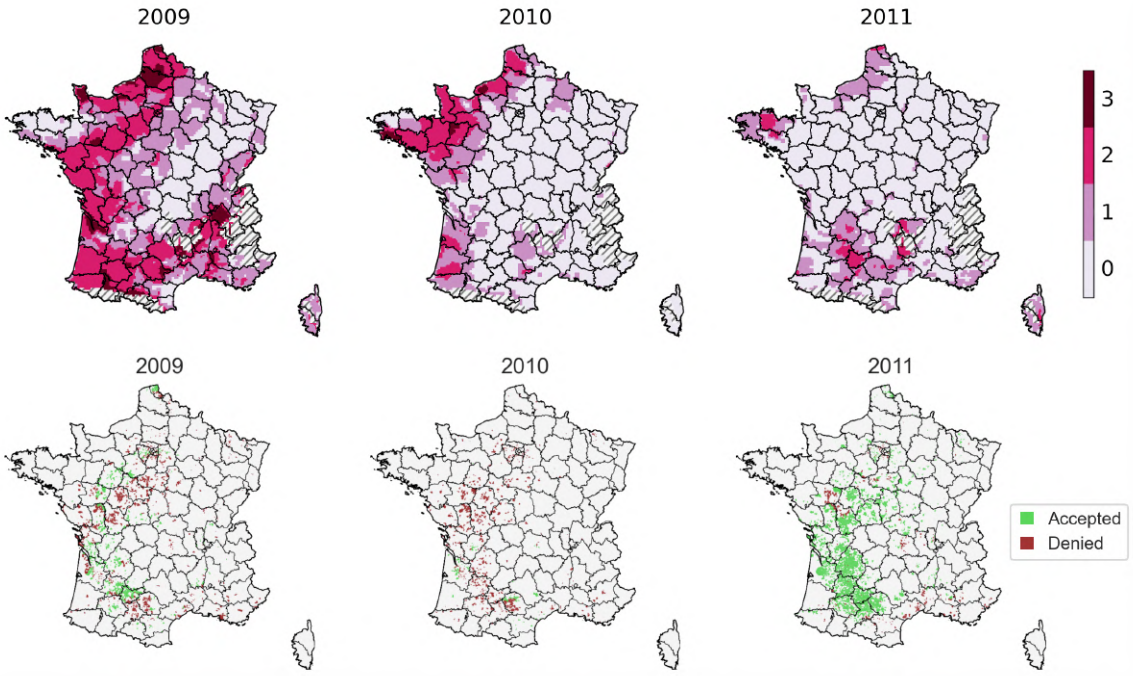


Figure 6.12: YDMI in classes (top) and NatCat requests (bottom) for 2009-2011

2016 in the town of Saint-Orens-de-Gameville, on the outskirts of Toulouse, we counted only 6 claims for 112 houses in the same housing block, when all of them were exposed to the same favorable conditions (drought combined to a high susceptibility based on mineralogy). In this situation and without additional information, a classification model can either identify all the houses as undamaged or damaged. In both cases the prediction is inadequate. This shows, without questioning their relevance, that our predictors are not sufficient to fully understand the occurrence of damage.

In fact, local characteristics concerning the building or its close environment, such as its age, foundation depth, or the presence of trees, are of significant importance (see section 1.2). Such information is not yet widely available for France.

6.4.2 What next steps could improve results ?

Without knowledge of local characteristics, damage prediction cannot be improved by any model. The approach we have developed is promising, although significant improvements are to be expected with the addition of more local data.

In the meantime, it would be useful to carry out local analyses to understand which local factors are relevant. This is the logical continuation of this work, as the research done so far has resulted in an optimal drought index (the 4-km resampled YDMI) and has identified an optimal susceptibility criterion (mineralogy).

The neighborhood scale seems to be a good compromise, because it allows to study a large number of houses, under soil and drought conditions that vary little, and individual building characteristics could be easily retrieved, for example using high-resolution satellite imagery.

These characteristics could then be used to understand why some buildings are affected and others are not. The approach could be applied across multiple departments, to identify regional trends.

6.5 Conclusion

In summary, this work using a random forest classification model shows us that it is possible to estimate the annual number of claims from soil and drought characteristics. However, challenges remain because (i) inconsistencies arise from model learning due to counter-intuitive aspects of the data, and (ii) the model has not proven to be able to accurately determine whether an insurance policy will be damaged or not. We explain this second point by the temporal catch-up effect of the NatCat system and the lack of information on local characteristics.

The random forest model emphasizes the superior predictive capacity if YDMI, and especially the 4-km YDMI in comparison to geological criteria. Of the latter, the mineralogy criterion performs best. It is important to recall the weakness of the mineralogy criterion, which is based on a small number of XRD values.

To provide perspectives, we suggest repeating the approach, adding predictors that describe local characteristics. Since this is currently not possible due to lack of data, it would be interesting to find out which factors are the most important based on local neighborhood-level analyses.

Conclusion and Perspectives

The natural phenomenon of clay shrinkage and swelling is a major problem in France, due to the damage it causes. In addition, its impact is set to worsen as a result of climate change. However, the phenomenon remains poorly understood, partly due to its complex, multi-factorial nature. The chapter 1 gives an overview of the state of the art.

The aim of this thesis is to improve the understanding and modeling of clay shrinkage and swelling and of its impacts, including statistically quantifying the overall number of claims. The goal is not to determine the risk of loss for each house. The perspective is to improve the tools currently used by public authorities to manage this risk. The work carried out during the thesis has led to a number of conclusions, which are detailed below. They are based on soil moisture simulations using Météo-France's ISBA model, on insurance data supplied by the CCR, on measurements from a site instrumented by BRGM in the commune of Chaingy (45), and on clay shrink-swell hazard maps relating to soil properties also established by BRGM. Each of these tools is detailed in chapter 2.

Summary of results

As a first step, we developed a method for creating a drought index based on the ISBA model that best matches a history of insurance claims. This index is called Yearly Drought Magnitude Index (YDMI). The method is based on Kendall's tau statistical test of association. Applied to a sample of 20 municipalities located in various geographical areas, better results are obtained when the YDMI is based on a deep layer of the model, considering deciduous tree vegetation and a strict definition of drought. Each of these elements has a scientific justification. We note a considerable improvement in scores with an interactive representation of vegetation, which is not yet operationally deployed. All these results are detailed in chapter 3.

A comparison was then made between ISBA simulations and measurements of soil moisture content and vertical displacement from the Chaingy instrumented site. It shows that while ISBA provides reliable trends, the model neglects the variability of soil moisture within an 8 km grid, which can in fact be considerable. The analysis is presented in chapter 4.

Next, the YDMI index is evaluated for France, for a historical period and a future period, enabling us to analyze past droughts, and to refine the characterization of trends to come under the effect of climate change. The years 2003, 2018, 2019, 2020 and 2022 stand out for their remarkable summer droughts. The projected analysis shows that the frequency and intensity of events will increase in the future, more so under the RCP 8.5 scenario than under the RCP 4.5 scenario. These results must be qualified, however, by the fact that past and projected YDMIs do not match. In particular, the year 2022 is unequaled by all the projected YDMIs. These results are developed in chapter 5.

Finally, soil characteristics derived from hazard maps and YDMI were spatially merged to train a random forest machine learning model to predict the occurrence of claims. Three geological criteria are compared: a lithological criterion, a mineralogical criterion and a geotechnical criterion. The model provides a good estimate of the annual number of claims.

Nevertheless, (i) the prediction is not always consistent with the YDMI indices and (ii) the predictors are insufficient to reliably estimate the damage occurrence for a given insurance policy. However, the 4-km YDMI and the mineralogical geological criterion stand out for their better predictive ability. This experiment is the subject of chapter 6.

Perspectives

This thesis work contributes to improving the characterization of the clay shrink-swell phenomenon. The methods developed could prove useful for understanding its other effects, as the phenomenon affects not only houses, but also roads and railways, for example. However, these results remain exploratory, and are not yet mature enough to be applied operationally. In addition, significant approximations have been made, and there are still major scientific hurdles to be overcome in this area.

Local factors

As indicated throughout this work, our approach neglects local factors that are nonetheless crucial to understanding the clay shrink-swell phenomenon. These include the presence of trees near a building, or the nature of the building itself. However, to our knowledge, there is no freely accessible database containing all this information for France. As a first step, it would be useful to prioritize these various local factors by carrying out a study of a small area, which could be characterized using high-resolution satellite data.

Cumulative effect of droughts

Another approximation made in this work concerns the cumulative effect of droughts. The YDMI index only takes into account the soil moisture of a single year. However, it has been proven that repeated droughts have an impact on soils, which become irreversibly deformed, and on buildings, which become more fragile with each cycle. This needs to be taken into account in order to better characterize the phenomenon, for example through an additional drought index.

Improving the temporal and spatial resolution of YDMI

Although YDMI is calculated on a yearly basis, it is entirely possible to estimate it on a finer temporal scale, since it is obtained by summing daily data. Calculating it on an operational day-by-day basis would enable to monitor drought conditions in real time, and thus be more reactive to the clay shrink-swell phenomenon. It is also possible to improve the spatial resolution of YDMI in a physical way, by using finer-scale climate forcings in ISBA, such as those from the AROME model [Seity et al., 2011], which have a resolution of 1.3 km. This has already been done in Rojas-Munoz et al. [2023], for example.

Imprecise soil mapping

The hazard maps used in the thesis to estimate the predisposition factor have certain weaknesses. In particular, their resolution is too low to identify very local clay layers, their superficial nature does not take into account the presence of clay at depth, and in some cases, the number of measurements on which they are based is very small. These maps could

be improved to overcome these drawbacks by merging with other data, such as the texture maps produced by INRAE [Roman Dobarco et al., 2022] or geophysical acquisitions. For this point, as for the previous one, it is not necessarily relevant to focus on the whole country, but rather on the areas most at risk, such as suburban areas.

Uncertainty about the effects of climate change on soils

Under current climatic conditions, our results showed that a YDMI based on a soil layer between 80 cm and 1 m deep could satisfactorily account for the number of claims. However, this result could be called into question by climate change, as the depth of the drying front is linked to the intensity of drought. To better anticipate future changes, these aspects could be monitored, using instrumented sites, for example.

EGMS data potential

As mentioned in chapter 1, high-precision satellite displacement data are now available. In particular, the European Ground Motion Service (EGMS) [Crosetto et al., 2021] offers data across Europe since 2018. These data were not exploited in this thesis, as their temporal depth is short, and under these conditions the signal is difficult to interpret. Nevertheless, they offer considerable potential for analyzing the clay shrink-swell phenomenon.

Conclusion et Perspectives

Le phénomène naturel de retrait-gonflement des argiles représente un problème majeur en France en raison du montant des dégâts qu'il engendre. De plus, ses impacts devraient s'amplifier sous l'effet du changement climatique. Cependant, c'est un phénomène qui reste méconnu, ce que l'on explique en partie par sa nature multifactorielle complexe. Le chapitre 1 donne un aperçu de l'état de l'art.

Cette thèse a pour objectif d'améliorer la compréhension et la modélisation du retrait-gonflement des argiles, ainsi que de ses impacts, et à quantifier statistiquement le nombre global de sinistres. Il ne s'agit pas ici de déterminer un risque de sinistre pour chaque maison. Elle s'inscrit dans la perspective opérationnelle d'apporter des pistes d'amélioration aux outils actuellement utilisés par les pouvoirs publics pour gérer ce risque. Les travaux accomplis au cours des trois ans ont permis de conclure sur certains points que nous détaillerons ci-dessous. Ils reposent sur des simulations d'humidité du sol réalisées à l'aide du modèle ISBA de Météo-France, sur des données assurantielles fournies par la CCR, sur des mesures provenant d'un site instrumenté par le BRGM dans la commune de Chaingy (45), et sur des cartes d'aléa RGA relatives aux propriétés des sols également établies par le BRGM. Chacun de ces outils est détaillé dans le chapitre 2.

Résumé des résultats

Dans un premier temps, nous avons développé une méthode pour créer un indice de sécheresse basé sur le modèle ISBA qui concorde au mieux avec un historique de sinistralité. Cet indice s'appelle le Yearly Drought Magnitude Index (YDMI). La méthode repose sur le test statistique d'association du tau de Kendall. Appliquée à un échantillon de 20 communes situées dans des zones géographiques variées, de meilleurs résultats sont obtenus lorsque le YDMI est basé sur une couche profonde du modèle, en considérant une végétation d'arbres feuillus et une définition stricte de la sécheresse. Chacun de ces éléments a une justification scientifique. Nous remarquons une amélioration considérable des scores avec une représentation interactive de la végétation, laquelle n'est pas encore déployée opérationnellement. L'ensemble de ces résultats est détaillé dans le chapitre 3.

Une comparaison a ensuite été établie entre les simulations réalisées avec ISBA et des mesures de teneur en eau des sols et de déplacement vertical provenant du site instrumenté de Chaingy. Elle montre que si ISBA restitue des tendances fiables, le modèle néglige la variabilité de l'humidité du sol au sein d'une maille de 8 km, qui peut en réalité s'avérer considérable. L'analyse est présentée dans le chapitre 4.

Ensuite, l'indice YDMI est évalué pour la France, pour une période historique et une période future, ce qui permet de dresser un bilan des sécheresses passées, et d'affiner la caractérisation des tendances à venir sous l'effet du changement climatique. Les années 2003, 2018, 2019, 2020 et 2022 se démarquent par leurs sécheresses estivales remarquables. L'analyse projetée montre quant à elle que la fréquence et l'intensité des événements augmenteraient dans le futur, davantage sous le scénario RCP 8.5 que sous le scénario RCP 4.5. Ces résultats sont néanmoins à nuancer par le fait que les YDMI passés et projetés ne concordent pas. Notamment, l'année

2022 est inégalée par l'ensemble des YDMI projetés. Ces résultats sont développés dans le chapitre 5.

Enfin, les caractéristiques de sol issues des cartes d'aléa et YDMI ont été croisés spatialement afin d'entraîner un modèle d'apprentissage automatique de type random forest, pour prédire la survenue de sinistres. Trois critères géologiques sont comparés : un critère lithologique, un critère minéralogique et un critère géotechnique. Le modèle permettent d'estimer de manière correcte le nombre annuel de sinistres. Néanmoins, (i) la prédiction n'est pas toujours cohérente avec les indices YDMI et (ii) les prédicteurs sont insuffisants pour estimer de manière fiable la sinistralité pour une police d'assurance donnée. Le YDMI à 4 km et le critère minéralogique se démarquent cependant par leur meilleure capacité prédictive. Cette expérience fait l'objet du chapitre 6.

Perspectives

Ces travaux de thèse contribuent à améliorer la caractérisation du phénomène de retrait-gonflement des argiles. Les méthodes mises au point pourraient s'avérer utiles pour comprendre d'autres effets du RGA, le phénomène n'affectant pas uniquement les maisons mais aussi les routes ou les voies ferrées par exemple. Cependant, ces résultats restent exploratoires, et ne sont pas encore matures pour être appliqués opérationnellement. De plus, des approximations importantes ont été faites et il reste encore des verrous scientifiques majeurs à lever sur la thématique.

Facteurs locaux

Comme indiqué tout au long des travaux, notre approche néglige les facteurs locaux qui sont pourtant primordiaux pour comprendre le phénomène RGA. Parmi ceux-ci, on peut citer la présence d'arbres à proximité d'un bâtiment ou la nature même de ce bâtiment. Cependant, il n'existe pas, à notre connaissance, de base de données librement accessible contenant ces informations de manière exhaustive sur la France. Il serait pertinent, dans un premier temps, de hiérarchiser ces différents facteurs locaux en faisant une étude sur une zone réduite, qui pourrait être caractérisée à l'aide de données satellitaires haute résolution.

Effet cumulatif des sécheresses

Une autre approximation faite dans ces travaux concerne l'effet cumulatif des sécheresses. En effet, l'indice YDMI intègre uniquement l'humidité du sol d'une seule année. Or, il a été prouvé par ailleurs que les sécheresses à répétition ont des conséquences sur les sols, qui se déforment de manière irréversible, et sur les bâtiments, qui sont fragilisés à chaque cycle. Il serait nécessaire d'en tenir compte afin de mieux caractériser le phénomène, par exemple par le biais d'un indice de sécheresse complémentaire.

Amélioration de la résolution temporelle et spatiale du YDMI

Si dans ces travaux le YDMI est calculé par années, il est tout à fait possible de l'estimer à une échelle temporelle plus fine, puisque celui-ci est obtenu en sommant des données journalières. Le calculer au jour le jour opérationnellement permettrait notamment de suivre les conditions de sécheresse en temps réel, et ainsi d'être plus réactif face au phénomène RGA.

Il est également possible d'améliorer la résolution spatiale du YDMI de manière physique, en utilisant dans ISBA des forçages climatiques à une échelle plus fine, comme par exemple ceux issus du modèle AROME [Seity et al., 2011] qui ont une résolution de 1,3 km. Ceci a déjà été fait dans Rojas-Munoz et al. [2023], par exemple.

Caractéristiques de sol imprécises

Les cartes d'aléa utilisées dans la thèse pour estimer le facteur de prédisposition présentent certaines faiblesses. On peut citer notamment leur résolution trop faible pour identifier des couches argileuses très locales, leur caractère superficiel, qui ne tient pas compte de la présence d'argile en profondeur et dans certains cas le nombre réduit de mesures sur lesquelles elles reposent. Ces cartes pourraient être améliorées afin de pallier ces inconvénients par croisement avec d'autres données, par exemple les cartes de texture réalisées par INRAE [Roman Dobarco et al., 2022] ou des acquisitions géophysiques. Pour ce point, comme pour le point précédent, il n'est pas forcément pertinent de se focaliser sur l'ensemble du pays, mais plutôt sur les zones les plus à risques telles que les zones périurbaines.

Incertitude sur les effets du changement climatique sur les sols

Dans les conditions climatiques actuelles, nos résultats ont montré qu'un YDMI basé sur une couche de sol située entre 80 cm et 1 m de profondeur permettait de rendre compte du nombre de sinistres de manière satisfaisante. Cependant, ce résultat pourrait être remis en question par le changement climatique, la profondeur du front de dessèchement étant liée à l'intensité de la sécheresse. Pour une meilleure anticipation des changements à venir, ces aspects pourraient être surveillés, par le biais de sites instrumentés, par exemple.

Potentiel des données EGMS

Comme indiqué dans le chapitre 1, il existe désormais des données satellitaires de déplacement des sols à haute précision spatiale. L'European Ground Motion Service (EGMS) [Crosetto et al., 2021] propose notamment des données depuis 2018 sur toute l'Europe. Ces données n'ont pas été exploitées dans cette thèse car leur profondeur temporelle est courte, et dans ces conditions le signal est difficile à interpréter. Elles présentent néanmoins un potentiel considérable pour l'analyse du phénomène RGA.

Appendices

Appendix A

Additional elements

A.1 SWCC characteristics derived from soil texture

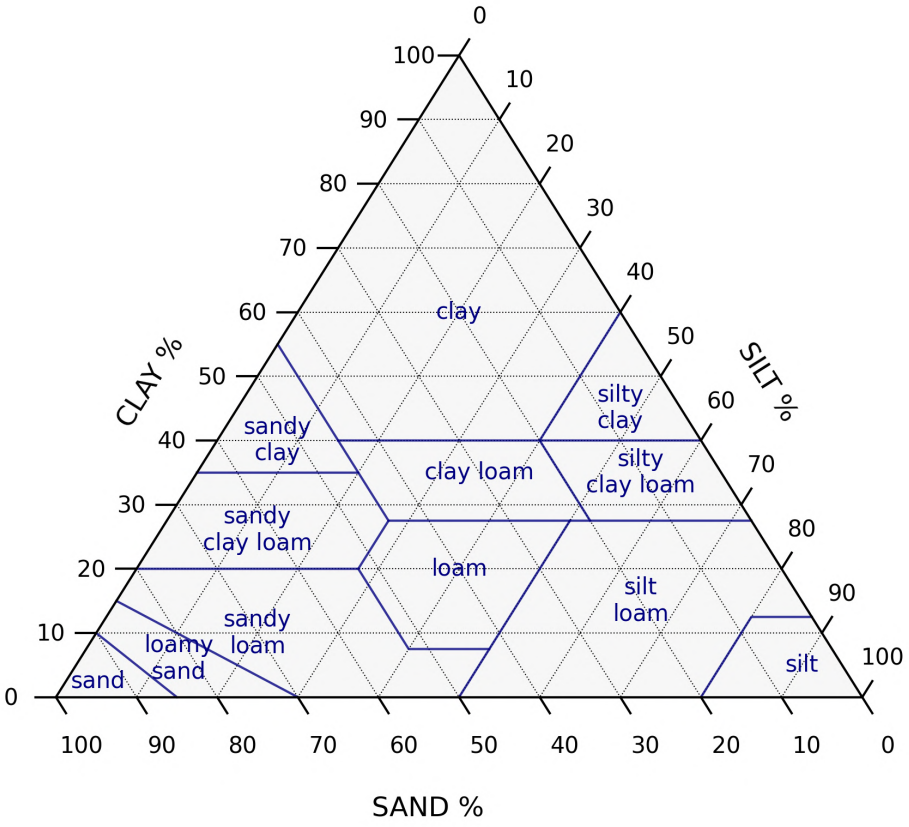


Figure A.1: Soil texture triangle, showing the 12 main textural classes as defined by the USDA

Table A.1: Parameter values for the Campbell [1974] and the Van Genuchten [1980] SWCC and conductivity functions for different soil texture. Table compiled by Bonan [2019], with data from Clapp and Hornberger [1978] and Carsel and Parrish [1988].

	Campbell (1974)				Van Genuchten (1980)				
soil type	θ_{sat}	ψ_{sat} (cm)	b	K_{sat} $(cm.h^{-1})$	θ_{sat}	θ_{res}	α (cm^{-1})	n	K_{sat} $(cm.h^{-1})$
sand	0.395	-12.1	4.05	63.36	0.43	0.045	0.145	2.68	29.7
loamy sand	0.41	-9	4.38	56.28	0.41	0.057	0.124	2.28	14.59
sandy loam	0.435	-21.8	4.9	12.48	0.41	0.065	0.075	1.89	4.42
silt loam	0.485	-78.6	5.3	2.59	0.45	0.067	0.02	1.41	0.45
loam	0.451	-47.8	5.39	2.5	0.43	0.078	0.036	1.56	1.04
sandy clay loam	0.42	-29.9	7.12	2.27	0.39	0.1	0.059	1.48	1.31
silty clay loam	0.477	-35.6	7.75	0.61	0.43	0.089	0.01	1.23	0.07
clay loam	0.476	-63	8.52	0.88	0.41	0.095	0.019	1.31	0.26
sandy clay	0.426	-15.3	10.4	0.78	0.38	0.1	0.027	1.23	0.12
silty clay	0.492	-49	10.4	0.37	0.36	0.07	0.005	1.09	0.02
clay	0.482	-40.5	11.4	0.46	0.38	0.068	0.008	1.09	0.2

A.2 Detail of NatCat requests between 1989 and 2022

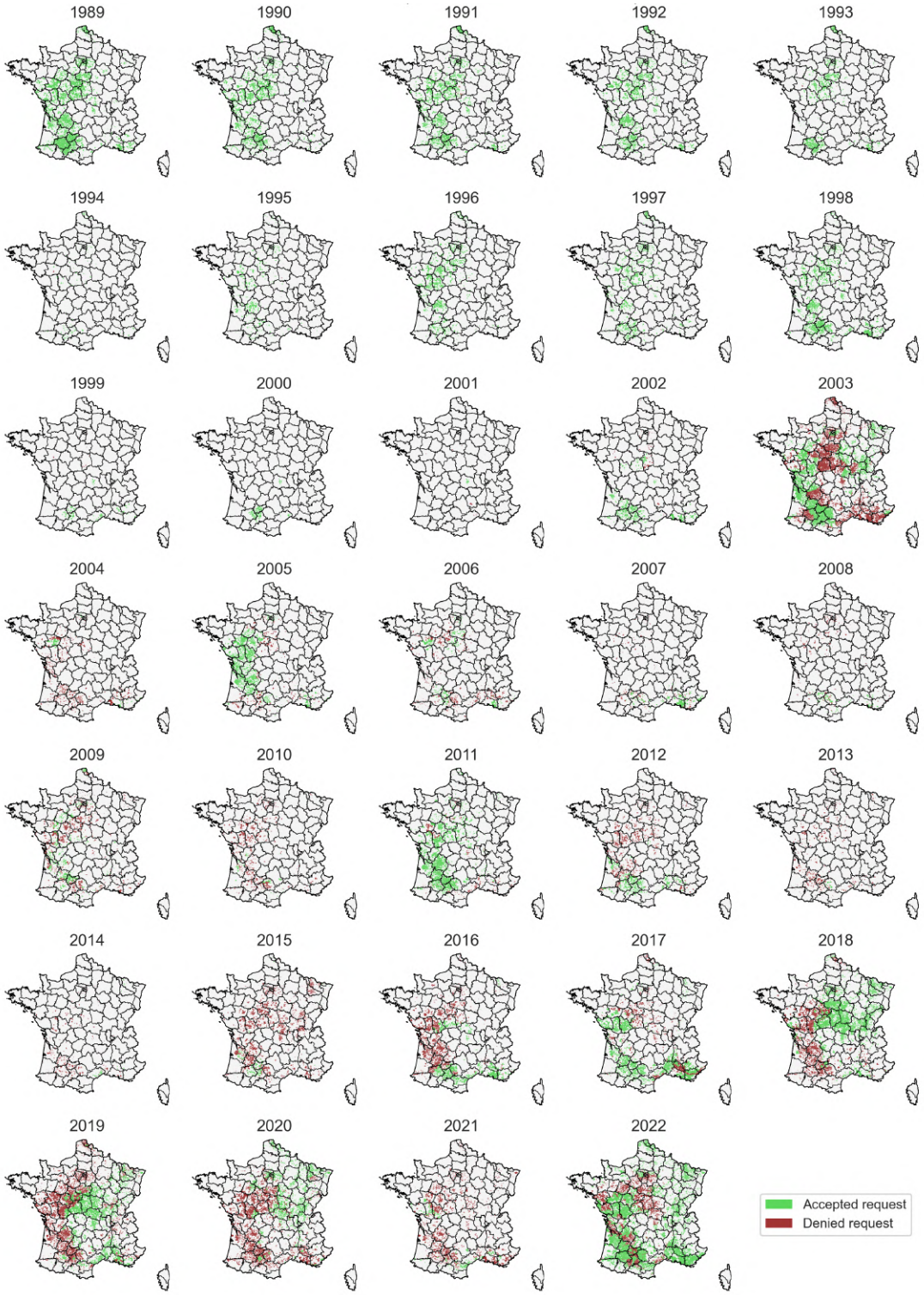


Figure A.2: Accepted and refused NatCat requests per year between 1989 and 2022 (last update 27/05/2024)

A.3 Number of houses in France

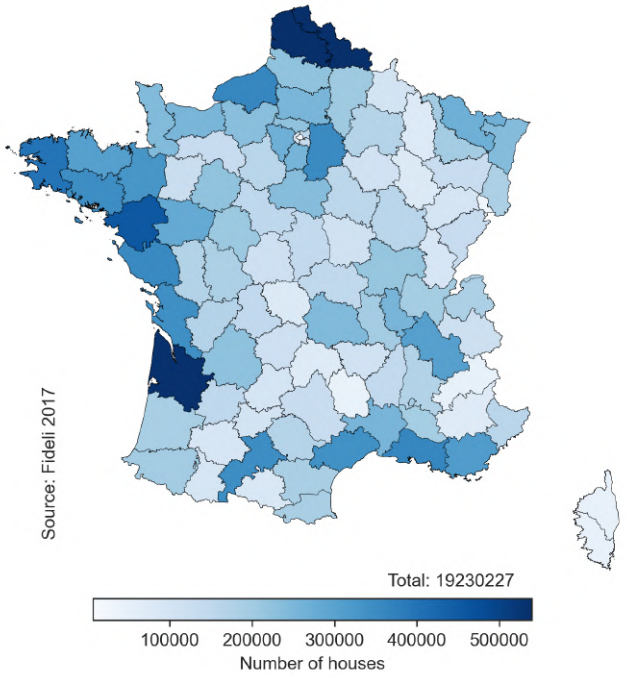


Figure A.3: Number of houses by department in France, INSEE

A.4 BRGM expansive soil mapping reports

Department	Year	Report number and link
Ain (01)	2009	RP-57699-FR
Aisne (02)	2009	RP-57155-FR
Allier (03)	2011	RP-52769-FR
Alpes-de-Haute-Provence (04)	2006	RP-54213-FR
Hautes-Alpes (05)	2010	RP-58667-FR
Alpes-Maritimes (06)	2010	RP-57979-FR
Ardèche (07)	2009	RP-57688-FR
Ardennes (08)	2010	RP-59147-FR
Ariège (09)	2009	RP-59147-FR
Aube (10)	2008	RP-56710-FR
Aude (11)	2007	RP-55419-FR
Aveyron (12)	2009	RP-57409-FR
Bouches-du-Rhône (13)	2004	RP-53237-FR
	2007	RP-55403-FR
Calvados (14)	2008	RP-56219-FR
Cantal (15)	2010	RP-58889-FR
Charente (16)	2007	RP-55432-FR

Table A.2 continued from previous page

Department	Year	Report number and link
Charente-Maritime (17)	2003	RP-52408-FR
Cher (18)	2004	RP-52682-FR
Corrèze (19)	2010	RP-57879-FR
Corse-du-sud (2A)	2010	RP-59012-FR
Haute-Corse (2B)	2010	RP-59013-FR
Côte-d'Or (21)	2007	RP-55427-FR
Côtes-d'Armor (22)	2010	RP-58917-FR
Creuse (23)	2010	RP-59210-FR
Dordogne (24)	2002	RP-51266-FR
Doubs (25)	2009	RP-57338-FR
Drôme (26)	2008	RP-56648-FR
Eure (27)	2008	RP-56485-FR
Eure-et-Loir (28)	2003	RP-52519-FR
Finistère (29)	2010	RP-58916-FR
Gard (30)	2007	RP-55420-FR
Haute-Garonne (31)	2002	RP-51894-FR
Gers (32)	2001	RP-51251-FR
Gironde (33)	2004	RP-53382-FR
Hérault (34)	2005	RP-54236-FR
Ille-et-Vilaine (35)	2010	RP-58120-FR
Indre (36)	2002	RP-51879-FR
Indre-et-Loire (37)	2005	RP-53351-FR
Isère (38)	2009	RP-57472-FR
Jura (39)	2009	RP-57419-FR
Landes (40)	2008	RP-56583-FR
Loir-et-Cher (41)	2004	RP-52950-FR
Loire (42)	2009	RP-57391-FR
Haute-Loire (43)	2010	RP-58946-FR
Loire-Atlantique (44)	2009	RP-56818-FR
Loiret (45)	2004	RP-53316-FR
Lot (46)	2006	RP-55122-FR
Lot-et-Garonne (47)	2002	RP-51752-FR
	2013	RP-61485-FR
Lozère (48)	2010	RP-57711-FR
Maine-et-Loire (49)	2005	RP-53753-FR
Manche (50)	2010	RP-58168-FR
Marne (51)	2008	RP-56828-FR
Haute-Marne (52)	2007	RP-55957-FR
Mayenne (53)	2010	RP-58896-FR
Meurthe-et-Moselle (54)	2006	RP-54860-FR
Meuse (55)	2008	RP-56295-FR
Morbihan (56)	2010	RP-58041-FR
Moselle (57)	2008	RP-56535-FR

Table A.2 continued from previous page

Department	Year	Report number and link
Nièvre (58)	2007	RP-54897-FR
Nord (59)	2003	RP-52107-FR
Oise (60)	2009	RP-57154-FR
Orne (61)	2008	RP-56594-FR
Pas-de-Calais (62)	2006	RP-53817-FR
Puy-de-Dôme (63)	2010	RP-59125-FR
Pyrénées-Atlantiques (64)	2008	RP-56566-FR
Hautes-Pyrénées (65)	2007	RP-55228-FR
Pyrénées-Orientales (66)	2010	RP-58054-FR
Bas-Rhin (67)	2009	RP-57373-FR
Haut-Rhin (68)	2009	RP-57374-FR
Rhône (69)	2008	RP-56842-FR
Haute-Saône (70)	2009	RP-57196-FR
Saône-et-Loire (71)	2008	RP-56635-FR
Sarthe (72)	2007	RP-55777-FR
Savoie (73)	2009	RP-57560-FR
Haute-Savoie (74)	2009	RP-57559-FR
Seine-Maritime (76)	2009	RP-57440-FR
Seine-et-Marne (77)	2006	RP-53212-FR
Yvelines (78)	2005	RP-53074-FR
Deux-Sèvres (79)	1998	RP-39957-FR
Somme (80)	2010	RP-58580-FR
Tarn (81)	2005	RP-53531-FR
Tarn-et-Garonne (82)	2002	RP-51893-FR
Var (83)	2007	RP-55471-FR
Vaucluse (84)	2004	RP-53187-FR
	2007	RP-55468-FR
Vendée (85)	2010	RP-57994-FR
Vienne (86)	2002	RP-51264-FR
Haute-Vienne (87)	2010	RP-58290-FR
Vosges (88)	2009	RP-57115-FR
Yonne (89)	2006	RP-54100-FR
	2007	RP-55546-FR
Territoire de Belfort (90)	2010	RP-57743-FR
Essonne (91)	2000	RP-50376-FR
	2007	RP-55526-FR
Hauts-de-seine (92)	2007	RP-53867-FR
Seine-Saint-Denis (93)	2001	RP-51198-FR
	2007	RP-55527-FR
Val-de-Marne (94)	2003	RP-52224-FR
	2007	RP-55525-FR
Val-d'Oise (95)	2004	RP-52598-FR

Appendix B

Additional results

B.1 YDMI and claims for layers 8 and 10

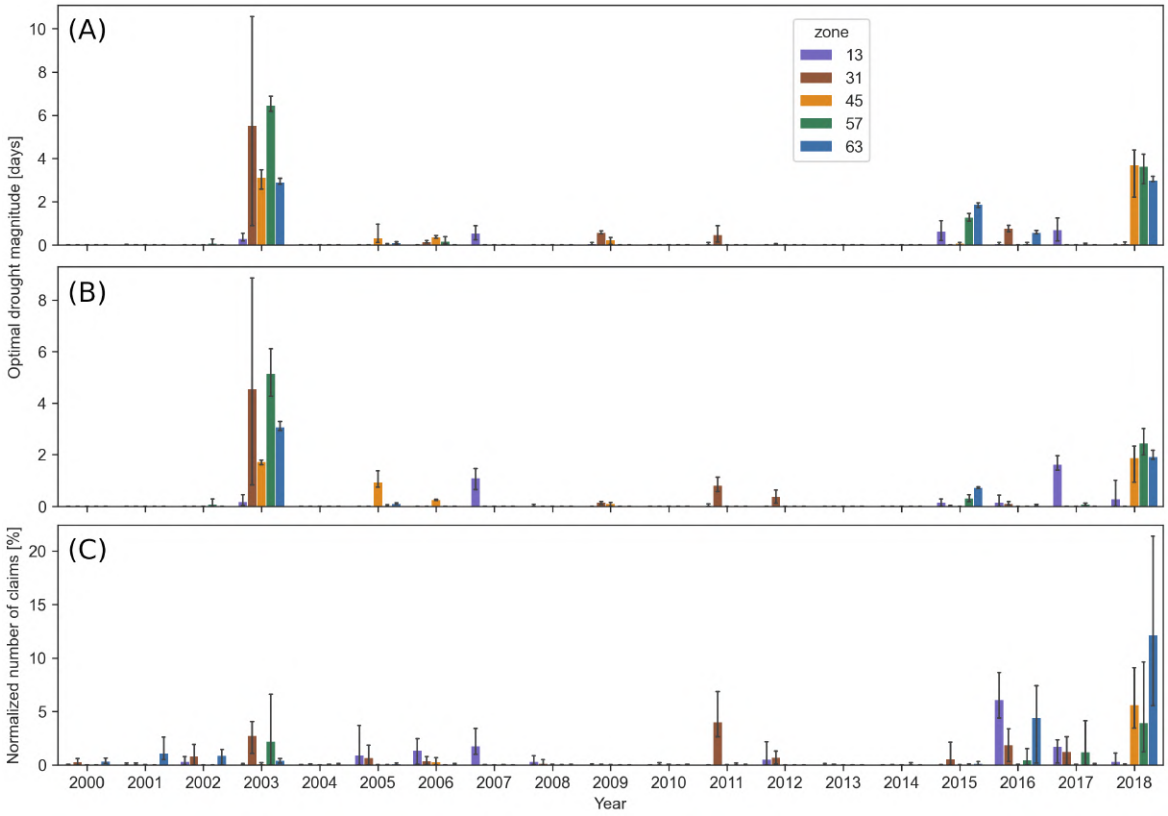
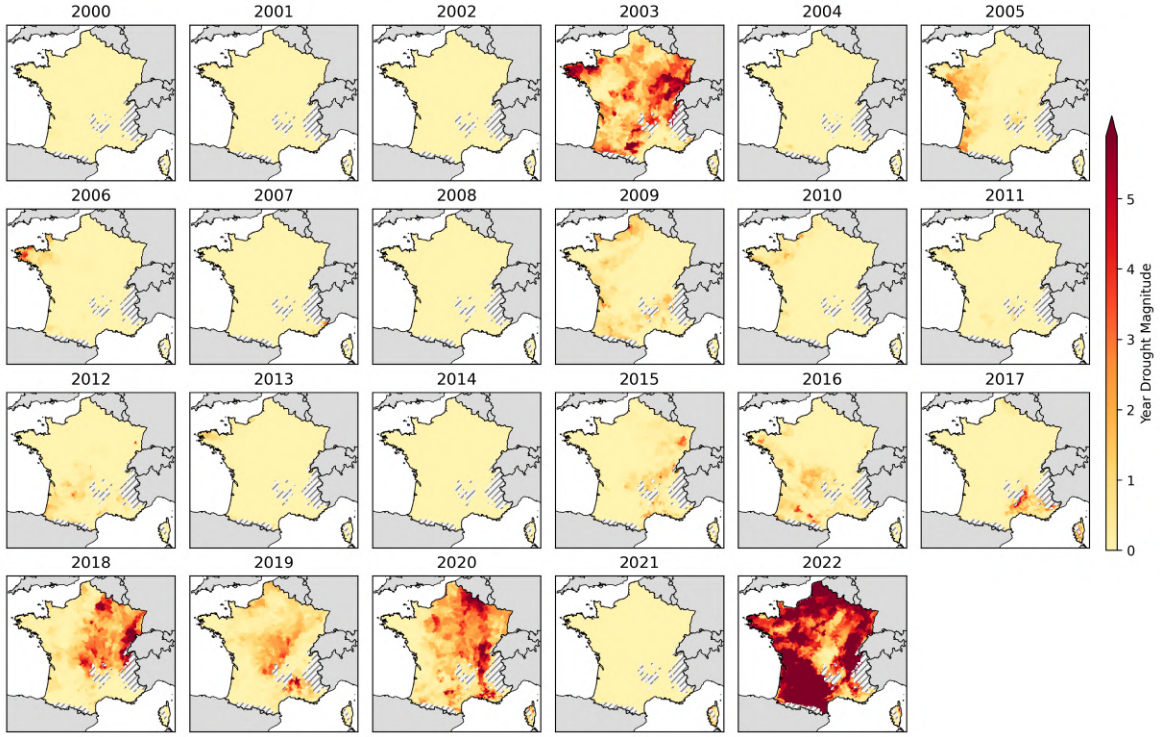
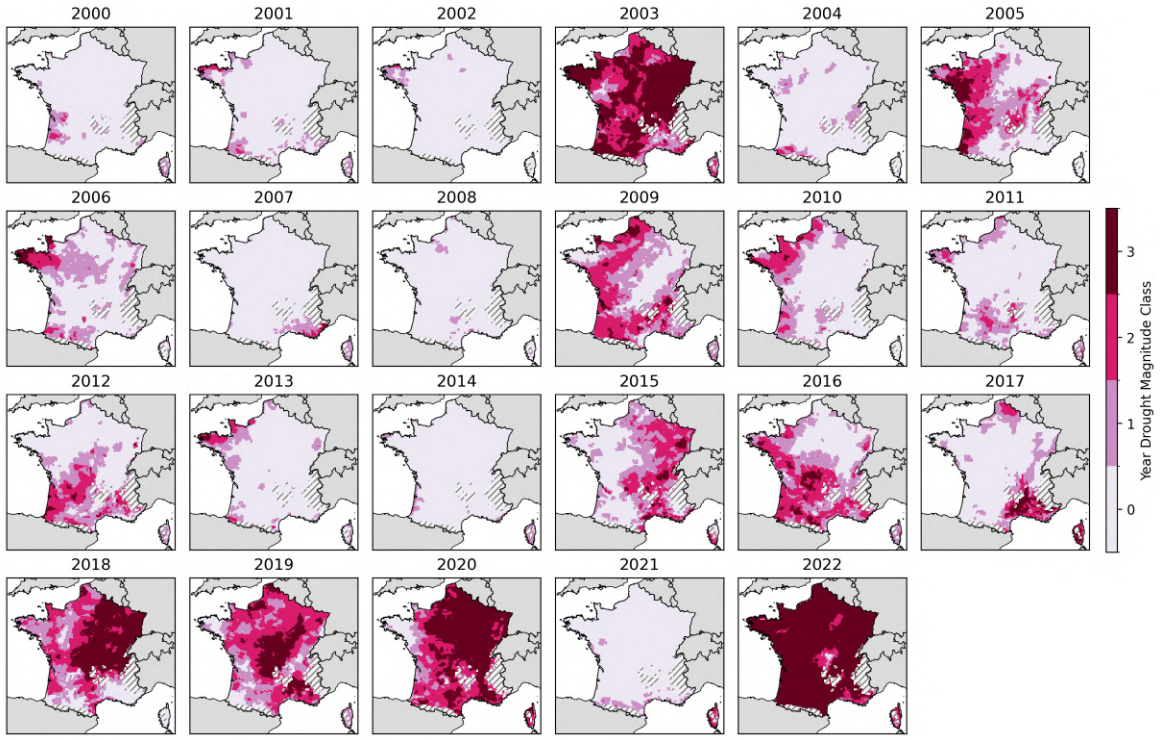


Figure B.1: YDMI obtained for patch 4 and (A) layer 8 or (B) layer 10, compared to the (C) normalized number of claims. Sample of 20 towns located in departments 13, 31, 45, 57, 63.

B.2 YDMI resampling to 4 km, 2000-2022



(a)



(b)

Figure B.2: YDMI after resampling to 4 km, (1) continuous and (2) by class

Bibliography

H. H. Adem and S. K. Vanapalli. Review of methods for predicting in situ volume change movement of expansive soil over time. *Journal of Rock Mechanics and Geotechnical Engineering*, 7(1):73–86, Feb. 2015. ISSN 16747755. doi:10.1016/j.jrmge.2014.11.002. (Cited in page 26.)

H. Assadollahi and H. Nowamooz. Long-term analysis of the shrinkage and swelling of clayey soils in a climate change context by numerical modelling and field monitoring. *Computers and Geotechnics*, 127:103763, Nov. 2020. ISSN 0266352X. doi:10.1016/j.compgeo.2020.103763. (Cited in page 28.)

P. Barbosa, D. Masante, C. Arias Muños, C. Cammalleri, A. de Jager, D. Magni, M. Mazzeschi, N. McCormick, G. Naumann, J. Spinoni, and J. Vogt. Drought in Europe September 2020. Technical report, Publications Office of the European Union, 2020. URL `https://edo.jrc.ec.europa.eu/documents/news/EDODroughtNews202009_Europe.pdf`. (Cited in page 86.)

S. Barthelemy, B. Bonan, J.-C. Calvet, G. Grandjean, D. Moncoulon, D. Kapsambelis, and S. Bernardie. A new approach for drought index adjustment to clay-shrinkage-induced subsidence over France: Advantages of the interactive leaf area index. *Nat. Hazards Earth Syst. Sci.*, 2024a. doi:10.5194/nhess-24-999-2024. (Cited in pages 54 and 74.)

S. Barthelemy, B. Bonan, M. Tomas-Burguera, G. Grandjean, S. Bernardie, J.-P. Naulin, P. Le Moigne, A. Boone, and J.-C. Calvet. Analyzing past and future droughts that induce clay shrinkage in France using an index based on water budget simulation for trees. May 2024b. doi:10.5194/egusphere-2024-1079. (Cited in pages 85, 86, 121 and 122.)

M. J. Best, M. Pryor, D. B. Clark, G. G. Rooney, R. L. H. Essery, C. B. Ménard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding. The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development*, 4(3):677–699, Sept. 2011. ISSN 1991-9603. doi:10.5194/gmd-4-677-2011. (Cited in page 25.)

C. M. Bhumralkar. Numerical Experiments on the Computation of Ground Surface Temperature in an Atmospheric General Circulation Model. *Journal of Applied Meteorology and Climatology*, 14(7):1246–1258, 1975. doi:10.1175/1520-0450(1975)014%3C1246:NEOTCO%3E2.0.CO;2. (Cited in page 33.)

A. K. Blackadar. Modeling the Nocturnal Boundary Layer. In *Proceedings of the Third Symposium on Atmospheric Turbulence, Diffusion, and Air Quality*, pages 46–49, Rayleigh, 1976. American Meteorological Society. (Cited in page 33.)

G. Bonan. *Climate Change and Terrestrial Ecosystem Modeling*. Cambridge University Press, 1 edition, Feb. 2019. ISBN 978-1-107-33921-7 978-1-107-04378-7 978-1-107-61907-4. doi:10.1017/9781107339217. (Cited in page 150.)

A. Boone, J.-C. Calvet, and J. Noilhan. Inclusion of a Third Soil Layer in a Land Surface Scheme Using the Force–Restore Method. *Journal of Applied Meteorology*, 38(11):1611–1630, Nov. 1999. ISSN 0894-8763, 1520-0450. doi:10.1175/1520-0450(1999)038<1611:IOATSL>2.0.CO;2. (Cited in page 33.)

A. Boone, V. Masson, T. Meyers, and J. Noilhan. The Influence of the Inclusion of Soil Freezing on Simulations by a Soil–Vegetation–Atmosphere Transfer Scheme. *Journal of Applied Meteorology*, 39(9):1544–1569, Sept. 2000. ISSN 0894-8763, 1520-0450. doi:10.1175/1520-0450(2000)039<1544:TIOTIO>2.0.CO;2. (Cited in page 33.)

L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 08856125. doi:10.1023/A:1010933404324. (Cited in page 128.)

J.-L. Briaud, X. Zhang, and S. Moon. Shrink Test–Water Content Method for Shrink and Swell Predictions. *Journal of Geotechnical and Geoenvironmental Engineering*, 129(7):590–600, July 2003. ISSN 1090-0241, 1943-5606. doi:10.1061/(ASCE)1090-0241(2003)129:7(590). (Cited in page 22.)

J. J. B. Bronswijk. Prediction of actual cracking and subsidence in clay soils. *Soil Science*, 148(2):87–93, Aug. 1989. ISSN 0038-075X. doi:10.1097/00010694-198908000-00002. (Cited in page 12.)

A. Brut, C. Rüdiger, S. Lafont, J. L. Roujean, J. C. Calvet, L. Jarlan, A. L. Gibelin, C. Albergel, P. Le Moigne, J. F. Soussana, K. Klumpp, D. Guyon, J. P. Wigneron, and E. Ceschia. Modelling LAI at a regional scale with ISBA-A-gs: Comparison with satellite-derived LAI over southwestern France. *Biogeosciences*, 6(8):1389–1404, 2009. ISSN 17264189. doi:10.5194/bg-6-1389-2009. (Cited in page 35.)

A. Burnol, M. Foumelis, S. Gourdier, J. Deparis, and D. Raucoules. Monitoring of expansive clays over drought-rewetting cycles using satellite remote sensing. *Atmosphere*, 12(10), 2021. ISSN 20734433. doi:10.3390/atmos12101262. (Cited in page 24.)

J.-C. Calvet. Investigating soil and atmospheric plant water stress using physiological and micrometeorological data. *Agricultural and Forest Meteorology*, 103(3):229–247, 2000. ISSN 01681923. doi:10.1016/S0168-1923(00)00130-1. (Cited in page 34.)

J.-C. Calvet and J.-F. Soussana. Modelling CO2-enrichment effects using an interactive vegetation SVAT scheme. *Agricultural and Forest Meteorology*, 108(2):129–152, 2001. ISSN 01681923. doi:10.1016/S0168-1923(01)00235-0. (Cited in page 35.)

J.-C. Calvet, J. Noilhan, J.-L. Roujean, P. Bessemoulin, M. Cabelguenne, A. Olioso, and J.-P. Wigneron. An interactive vegetation SVAT model tested against data from six contrasting sites. *Agricultural and Forest Meteorology*, 92(2):73–95, 1998. ISSN 01681923. doi:10.1016/S0168-1923(98)00091-4. (Cited in page 34.)

J.-C. Calvet, V. Rivalland, C. Picon-Cochard, and J.-M. Guehl. Modelling forest transpiration and CO2 fluxes—response to soil moisture stress. *Agricultural and Forest Meteorology*, 124 (3-4):143–156, Aug. 2004. ISSN 01681923. doi:10.1016/j.agrformet.2004.01.007. (Cited in page 34.)

J.-C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, and B. Piguet. In situ soil moisture observations for the CAL/VAL of SMOS: The SMOSMANIA network. In *2007 IEEE International Geoscience and Remote Sensing Symposium*, pages 1196–1199, Barcelona, Spain, 2007. IEEE. ISBN 978-1-4244-1211-2. doi:10.1109/IGARSS.2007.4423019. (Cited in page 23.)

J.-C. Calvet, S. Lafont, E. Cloppet, F. Souverain, V. Badeau, and C. Le Bas. Use of agricultural statistics to verify the interannual variability in land surface models: A case study over France with ISBA-A-gs. *Geoscientific Model Development*, 5(1):37–54, Jan. 2012. ISSN 1991-9603. doi:10.5194/gmd-5-37-2012. (Cited in page 34.)

J.-C. Calvet, N. Fritz, C. Berne, B. Piguet, W. Maurel, and C. Meurey. Deriving pedotransfer functions for soil quartz fraction in southern France from reverse modeling. *SOIL*, 2(4): 615–629, Dec. 2016. ISSN 2199-398X. doi:10.5194/soil-2-615-2016. (Cited in page 23.)

G. S. Campbell. A simple method for determining unsaturated conductivity from moisture retention data. *Soil Science*, 117(6):311–314, June 1974. ISSN 0038-075X. doi:10.1097/00010694-197406000-00001. (Cited in pages 13, 14, 15, 38 and 150.)

R. F. Carsel and R. S. Parrish. Developing joint probability distributions of soil water retention characteristics. *Water Resources Research*, 24(5):755–769, May 1988. ISSN 0043-1397, 1944-7973. doi:10.1029/WR024i005p00755. (Cited in page 150.)

CCR. Sécheresse de 2011 en France, 2022. URL https://catastrophes-naturelles.ccr.fr/-/001874_secheresse-de-2011-en-france. (Cited in page 136.)

CCR. Les catastrophes naturelles en France, bilan 1982-2022. Technical report, 2023a. URL https://catastrophes-naturelles.ccr.fr/documents/148935/368920/BILAN+Cat+Nat+2022_HD_12062023.pdf/5b5397f0-2118-2252-608f-76ece6e195a1?t=1686731142230. (Cited in page 2.)

CCR. Conséquences du changement climatique sur le coût des catastrophes naturelles en France à horizon 2050. Technical report, 2023b. URL https://www.ccr.fr/documents/35794/1255983/CCR+Etude+climat+BAG+23102023+page+22mo.pdf/68b95f6e-8238-4dcc-6c56-025fa410257b?t=1698161402128. (Cited in pages 2, 6 and 28.)

N. Chabrat, O. Cuisinier, and F. Masrouri. In Situ Alteration of the Hydro-Mechanical Behaviour of a Compacted Stabilised Expansive Soil. *Geotechnics*, 3(3):921–936, Sept. 2023. ISSN 2673-7094. doi:10.3390/geotechnics3030049. (Cited in page 13.)

S. A. Changnon. Detecting Drought Conditions in Illinois. Technical report, Illinois State Water Survey, 1987. URL https://www.isws.illinois.edu/pubdoc/C/ISWSC-169.pdf. (Cited in pages 1 and 5.)

A. Charpentier, M. James, and H. Ali. Predicting drought and subsidence risks in France. *Natural Hazards and Earth System Sciences*, 22(7):2401–2418, July 2022. ISSN 1684-9981. doi:10.5194/nhess-22-2401-2022. (Cited in pages 26 and 122.)

Ph. Ciais, M. Reichstein, N. Viovy, A. Granier, J. Ogée, V. Allard, M. Aubinet, N. Buchmann, Chr. Bernhofer, A. Carrara, F. Chevallier, N. De Noblet, A. D. Friend, P. Friedlingstein, T. Grünwald, B. Heinesch, P. Keronen, A. Knohl, G. Krinner, D. Loustau, G. Manca, G. Matteucci, F. Miglietta, J. M. Ourcival, D. Papale, K. Pilegaard, S. Rambal, G. Seufert, J. F. Soussana, M. J. Sanz, E. D. Schulze, T. Vesala, and R. Valentini. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437 (7058):529–533, Sept. 2005. ISSN 0028-0836, 1476-4687. doi:10.1038/nature03972. (Cited in pages 2 and 5.)

R. B. Clapp and G. M. Hornberger. Empirical equations for some soil hydraulic properties. *Water Resources Research*, 14(4):601–604, Aug. 1978. ISSN 00431397. doi:10.1029/WR014i004p00601. (Cited in pages 34 and 150.)

R. Cojean. Introduction. *Revue Française de Géotechnique*, 120–121(120-121):7–9, Oct. 2007. ISSN 0181-0529. doi:10.1051/geotech/2007120007. (Cited in page 13.)

E. Çokça. Relationship between methylene blue value, initial soil suction and swell percent of expansive soils. *Turkish Journal of Engineering and Environmental Sciences*, 26(6): 521–529, 2002. ISSN 13000160. URL https://www.scopus.com/inward/record.uri?partnerID=HzOxMe3b&scp=0036405071&origin=inward. (Cited in page 21.)

L. F. Cooling and W. H. Ward. Some examples of foundation movements due to causes other than structural loads. In *Proceedings 2nd International Conference on Soil Mechanics and Foundation Engineering*, pages 162–167, Rotterdam, 1948. doi:10.1007/978-3-319-73568-9_174. (Cited in page 18.)

T. Corti, V. Muccione, P. Köllner-Heck, D. Bresch, and S. I. Seneviratne. Simulating past droughts and associated building damages in France. *Hydrology and Earth System Sciences*, 13(9):1739–1747, 2009. ISSN 16077938. doi:10.5194/hess-13-1739-2009. (Cited in pages 26 and 133.)

T. Corti, M. Wüest, D. Bresch, and S. I. Seneviratne. Drought-induced building damages from simulations at regional scale. *Natural Hazards and Earth System Sciences*, 11(12): 3335–3342, 2011. ISSN 15618633. doi:10.5194/nhess-11-3335-2011. (Cited in pages 20 and 26.)

Covéa and RiskWeatherTech. Changement climatique & Assurance : Quelles conséquences sur la sinistralité à horizon 2050 ? Technical report, 2022. URL https://www.covea.com/sites/default/files/2022-02/202202_Livre_Blanc_Cov%C3%A9a_Risques_Climatiques.pdf. (Cited in page 28.)

M. Crosetto, L. Solari, J. Balasis-Levinsen, L. Bateson, N. Casagli, M. Frei, A. Oyen, D. A. Moldestad, and M. Mróz. Deformation monitoring at european scale: The copernicus ground motion service. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, volume 43, pages 141–146, 2021. doi:10.5194/isprs-archives-XLIII-B3-2021-141-2021. (Cited in pages 24, 141 and 145.)

H. Darcy. *Les fontaines publiques de la ville de Dijon : Détermination des lois d'écoulement de l'eau à travers le sable.* V. Dalmont, 1856. URL https://gallica.bnf.fr/ark:/12148/bpt6k624312/f3.item. (Cited in page 33.)

G. Dayon, J. Boé, É. Martin, and J. Gailhard. Impacts of climate change on the hydrological cycle over France and associated uncertainties. *Comptes Rendus - Geoscience*, 350(4): 141–153, 2018. ISSN 16310713. doi:10.1016/j.crte.2018.03.001. (Cited in page 28.)

G. De Marsily. *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*. Academic Press, 1986. ISBN 0-12-208915-4. URL https://gw-project.org/books/quantitative-hydrogeology-groundwater-hydrology-for-engineers/. (Cited in pages 13 and 16.)

H. De Solère. Retour d'expérience de la sécheresse 2003. *Revue Française de Géotechnique*, (120-121):21–25, Oct. 2007. ISSN 0181-0529. doi:10.1051/geotech/2007120021. (Cited in page 30.)

J. W. Deardorff. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *Journal of Geophysical Research: Oceans*, 83(C4):1889–1903, Apr. 1978. ISSN 0148-0227. doi:10.1029/JC083iC04p01889. (Cited in page 33.)

B. Decharme, A. Boone, C. Delire, and J. Noilhan. Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions. *Journal of Geophysical Research Atmospheres*, 116(20):1–29, 2011. ISSN 01480227. doi:10.1029/2011JD016002. (Cited in pages 26 and 33.)

J. A. Dracup, K. S. Lee, and E. G. Paulson. On the definition of droughts. *Water Resources Research*, 16(2):297–302, 1980. ISSN 19447973. doi:10.1029/WR016i002p00297. (Cited in pages 1 and 5.)

R. Driscoll. The influence of vegetation on the swelling and shrinking of clay soils in Britain. *Géotechnique*, 33(2):93–105, June 1983. ISSN 0016-8505. doi:10.1680/geot.1983.33.2.93. (Cited in page 17.)

E. Ducasse. *Cartographie fine de l'argile minéralogique par démélange d'images hyperspectrales à très haute résolution spatiale*. PhD thesis, Université de Toulouse, 2019. URL https://depozit.isae.fr/theses/2019/2019_Ducasse_Etienne_D.pdf. (Cited in page 21.)

S. Earle. *Environmental Geology*. 2021. URL https://environmental-geol.pressbooks.tru.ca/. (Cited in pages 11 and 12.)

M. B. Ek, K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, 108(D22):2002JD003296, Nov. 2003. ISSN 0148-0227. doi:10.1029/2002JD003296. (Cited in page 25.)

P. Etchevers, F. Besson, and J.-M. Soubeyroux. Contenu en eau des sols : comparaison des données simulées par SIM2 avec les données observées. *La Météorologie*, (118):029, 2022. ISSN 0026-1181. doi:10.37053/lameteorologie-2022-0058. (Cited in page 78.)

A. Farahmand, J. T. Reager, and N. Madani. Drought Cascade in the Terrestrial Water Cycle: Evidence From Remote Sensing. *Geophysical Research Letters*, 48(14):e2021GL093482, July 2021. ISSN 0094-8276, 1944-8007. doi:10.1029/2021GL093482. (Cited in pages 1 and 5.)

S. Faroux, A. T. Kaptué Tchuenté, J.-L. Roujean, V. Masson, E. Martin, and P. Le Moigne. ECOCLIMAP-II/Europe: A twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models. *Geoscientific Model Development*, 6(2):563–582, 2013. ISSN 1991-959X. doi:10.5194/gmd-6-563-2013. (Cited in page 35.)

T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, June 2006. ISSN 01678655. doi:10.1016/j.patrec.2005.10.010. (Cited in page 131.)

M. Fernandes, A. Denis, R. Fabre, J.-F. Lataste, and M. Chrétien. In situ study of the shrinkage-swelling of a clay soil over several cycles of drought-rewetting. *Engineering Geology*, 192:63–75, June 2015. ISSN 00137952. doi:10.1016/j.enggeo.2015.03.017. (Cited in pages 13, 18 and 23.)

J. B. J. Fourier. *Théorie Analytique de La Chaleur*, volume 504. Didot Paris, 1822. URL https://www.academie-sciences.fr/archivage_site/activite/archive/dossiers/Fourier/Fourier_pdf/Mem1828_p581_622.pdf. (Cited in page 33.)

France Assureurs. IMPACT DU CHANGEMENT CLIMATIQUE SUR L'ASSURANCE À L'HORIZON 2050. Technical report, 2021. URL https://www.franceassureurs.fr/wp-content/uploads/2022/09/vf_france-assureurs_impact-du-changement-climatique-2050.pdf. (Cited in page 28.)

T. J. Freeman, D. Burford, and M. S. Crilly. Seasonal foundation movements in London Clay. In *Proceedings of the 4th International Conference on Ground Movements and Structures*, pages 485–501, Cardiff, 1992. (Cited in pages 16 and 18.)

A. L. Gibelin, J. C. Calvet, J. L. Roujean, L. Jarlan, and S. O. Los. Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products. *Journal of Geophysical Research Atmospheres*, 111(18), 2006. ISSN 01480227. doi:10.1029/2005JD006691. (Cited in page 35.)

P. H. Gleick, Pacific Institute for Studies in Development, Environment, and Security, S. E. Institute, and Pacific Institute for Studies in Development, Environment, and Security, editors. *Water in Crisis: A Guide to the World's Fresh Water Resources*. Oxford Univ. Press, New York, NY, 1993. ISBN 978-0-19-507628-8 978-0-19-507627-1. (Cited in pages 1 and 5.)

S. Gourdier. Suivi du site instrumenté « argiles » de Chaingy (Loiret) – Phase 4 : Suivi annuel 2017. Technical Report RP-67802-FR, BRGM, 2018. (Cited in pages 24 and 50.)

S. Gourdier and E. Plat. Impact du changement climatique sur la sinistralité due au retrait-gonflement des argiles. In *Journées Nationales de Géotechnique et Géologie de l'Ingénieur (JNGG)*, 2018. URL https://hal-brgm.archives-ouvertes.fr/hal-01768395. (Cited in pages 2, 6 and 28.)

G. Grandjean, X. Briottet, K. Adeline, A. Bourguignon, and A. Hohmann. Clay Minerals Mapping from Imaging Spectroscopy. In *Geospatial Analyses of Earth Observation (EO) Data*. IntechOpen, May 2019. doi:10.5772/intechopen.86149. (Cited in page 21.)

L. Gudmundsson and S. I. Seneviratne. European drought trends. *Proceedings of the International Association of Hydrological Sciences*, 369:75–79, June 2015. ISSN 2199-899X. doi:10.5194/piahs-369-75-2015. (Cited in page 27.)

J. Hansen, P. Kharecha, M. Sato, V. Masson-Delmotte, F. Ackerman, D. J. Beerling, P. J. Hearty, O. Hoegh-Guldberg, S.-L. Hsu, C. Parmesan, J. Rockstrom, E. J. Rohling, J. Sachs, P. Smith, K. Steffen, L. Van Susteren, K. Von Schuckmann, and J. C. Zachos. Assessing "Dangerous Climate Change": Required Reduction of Carbon Emissions to Protect Young People, Future Generations and Nature. *PLoS ONE*, 8(12):e81648, Dec. 2013. ISSN 1932-6203. doi:10.1371/journal.pone.0081648. (Cited in page 27.)

A. Harrison, J. Plim, M. Harrison, L. Jones, and M. Culshaw. The relationship between shrink–swell occurrence and climate in south-east England. *Proceedings of the Geologists' Association*, 123(4):556–575, Aug. 2012. ISSN 00167878. doi:10.1016/j.pgeola.2012.05.002. (Cited in page 28.)

A. B. Hawkins. Some engineering geological effects of drought: Examples from the UK. *Bulletin of Engineering Geology and the Environment*, 72(1):37–59, 2013. ISSN 14359529. doi:10.1007/s10064-013-0458-7. (Cited in pages 16 and 17.)

S. Hoyer and J. Hamman. Xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*, 5(1):10, Apr. 2017. ISSN 2049-9647. doi:10.5334/jors.148. (Cited in page 52.)

J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. ISSN 1521-9615. doi:10.1109/MCSE.2007.55. (Cited in page 52.)

L. Ighil Ameur. Analysis of climate change impacts on the shrinkage-swelling phenomenon of clayey soils to adapt infrastructures. *E3S Web of Conferences*, 382:01003, Apr. 2023. ISSN 2267-1242. doi:10.1051/e3sconf/202338201003. (Cited in page 13.)

IPCC, editor. *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 1 edition, Mar. 2014. ISBN 978-1-107-05799-9 978-1-107-66182-0 978-1-107-41532-4. doi:10.1017/CBO9781107415324. (Cited in page 27.)

IPCC. *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 1 edition, July 2023. ISBN 978-1-00-915789-6. doi:10.1017/9781009157896. (Cited in pages 1, 5 and 27.)

D. Jacob, J. Petersen, B. Eggert, A. Alias, O. B. Christensen, L. M. Bouwer, A. Braun, A. Colette, M. Déqué, G. Georgievski, E. Georgopoulou, A. Gobiet, L. Menut, G. Nikulin, A. Haensler, N. Hempelmann, C. Jones, K. Keuler, S. Kovats, N. Kröner, S. Kotlarski, A. Kriegsmann, E. Martin, E. van Meijgaard, C. Moseley, S. Pfeifer, S. Preuschmann, C. Radermacher, K. Radtke, D. Rechid, M. Rounsevell, P. Samuelsson, S. Somot, J.-F. Soussana, C. Teichmann, R. Valentini, R. Vautard, B. Weber, and P. Yiou. EURO-CORDEX: New high-resolution climate change projections for European impact research. *Regional Environmental Change*, 14(2):563–578, Apr. 2014. ISSN 1436-3798, 1436-378X. doi:10.1007/s10113-013-0499-2. (Cited in page 27.)

D. Jacob, C. Teichmann, S. Sobolowski, E. Katragkou, I. Anders, M. Belda, R. Benestad, F. Boberg, E. Buonomo, R. M. Cardoso, A. Casanueva, O. B. Christensen, J. H. Christensen, E. Coppola, L. De Cruz, E. L. Davin, A. Dobler, M. Domínguez, R. Fealy, J. Fernandez, M. A. Gaertner, M. García-Díez, F. Giorgi, A. Gobiet, K. Goergen, J. J. Gómez-Navarro, J. J. G. Alemán, C. Gutiérrez, J. M. Gutiérrez, I. Güttler, A. Haensler, T. Halenka, S. Jerez, P. Jiménez-Guerrero, R. G. Jones, K. Keuler, E. Kjellström, S. Knist, S. Kotlarski, D. Maraun, E. van Meijgaard, P. Mercogliano, J. P. Montávez, A. Navarra, G. Nikulin, N. de Noblet-Ducoudré, H.-J. Panitz, S. Pfeifer, M. Piazza, E. Pichelli, J.-P. Pietikäinen, A. F. Prein, S. Preuschmann, D. Rechid, B. Rockel, R. Romera, E. Sánchez, K. Sieck, P. M. M. Soares, S. Somot, L. Srnec, S. L. Sørland, P. Termonia, H. Truhetz, R. Vautard, K. Warrach-Sagi, and V. Wulfmeyer. Regional climate downscaling over Europe: Perspectives from the EURO-CORDEX community. *Regional Environmental Change*, 20 (2):51, June 2020. ISSN 1436-3798, 1436-378X. doi:10.1007/s10113-020-01606-9. (Cited in page 28.)

K. Jordahl, J. V. D. Bossche, M. Fleischmann, J. Wasserman, J. McBride, J. Gerard, J. Tratner, M. Perry, A. G. Badaracco, C. Farmer, G. A. Hjelle, A. D. Snow, M. Cochran, S. Gillies, L. Culbertson, M. Bartos, N. Eubank, Maxalbert, A. Bilogur, S. Rey, C. Ren, D. Arribas-Bel, L. Wasser, L. J. Wolf, M. Journois, J. Wilson, A. Greenhall, C. Holdgraf, Filipe, and F. Leblanc. Geopandas/geopandas: V0.8.1. Zenodo, July 2020. (Cited in page 52.)

JRC. Soil Atlas of Europe - Soil forming processes - The movement of clay particles, 2024. URL https://esdac.jrc.ec.europa.eu/projects/Soil_Atlas/Pages/15.html. (Cited in page 12.)

P. C. Kariuki and T. Woldai. The Role of Remote Sensing in Mapping Swelling Soils. 5(1), 2004. (Cited in pages 2 and 6.)

C. Klein, C. S. Hurlbut, J. D. Dana, and C. Klein. *The 22nd Edition of the Manual of Mineral Science: After James D. Dana*. J. Wiley, New York, 22nd ed edition, 2002. ISBN 978-0-471-25177-4. (Cited in page 12.)

M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel. World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259–263, July 2006. ISSN 0941-2948. doi:10.1127/0941-2948/2006/0130. (Cited in page 13.)

T. Langreney, G. Le Cozannet, and M. Merad. Adapter le système assurantiel français face à l'évolution des risques climatiques. Technical report, 2023. URL https://www.ecologie.gouv.fr/sites/default/files/Rapport_final_Mission-assurance_climat.pdf. (Cited in page 26.)

D. A. Lashof and D. R. Ahuja. Relative contributions of greenhouse gas emissions to global warming. *Nature*, 344(6266):529–531, Apr. 1990. ISSN 0028-0836, 1476-4687. doi:10.1038/344529a0. (Cited in page 27.)

P. Le Moigne, F. Besson, E. Martin, J. Boé, A. Boone, B. Decharme, P. Etchevers, S. Faroux, F. Habets, M. Lafaysse, D. Leroux, and F. Rousset-Regimbeau. The latest improvements with SURFEX v8.0 of the Safran-Isba-Modcou hydrometeorological model for France. *Geoscientific Model Development*, 13(9):3925–3946, 2020. ISSN 19919603. doi:10.5194/gmd-13-3925-2020. (Cited in page 38.)

E. Le Roy. Sur l'intégration des équations de la chaleur. *Annales scientifiques de l'Ecole normale supérieure*, 14:379–465, 1897. ISSN 0012-9593, 1873-2151. doi:10.24033/asens.452. (Cited in page 33.)

M. MacQueen, M. Lawson, and W.-N. Ding. The 2018–2019 UK residential dwellings clay shrinkage subsidence event. *International Journal of Building Pathology and Adaptation*, May 2023. ISSN 2398-4708. doi:10.1108/IJBPA-06-2022-0091. (Cited in pages 2 and 6.)

C. Manning, M. Widmann, E. Bevacqua, A. F. Van Loon, D. Maraun, and M. Vrac. Increased probability of compound long-duration dry and hot events in Europe during summer (1950–2013). *Environmental Research Letters*, 14(9):094006, Sept. 2019. ISSN 1748-9326. doi:10.1088/1748-9326/ab23bf. (Cited in page 27.)

D. Masante, J. Vogt, C. Cammalleri, J. Spinoni, and P. Barbosa. Drought in Central-Northern Europe - September 2018. Technical report, Publications Office of the European Union, 2018. URL https://edo.jrc.ec.europa.eu/documents/news/EDODroughtNews201809_Central_North_Europe.pdf. (Cited in pages 86 and 115.)

D. Masante, P. Barbosa, and D. Magni. Drought in Europe August 2019. Technical report, Publications Office of the European Union, 2019. URL https://edo.jrc.ec.europa.eu/documents/news/EDODroughtNews201908_Europe.pdf. (Cited in page 86.)

V. Masson, P. Le Moigne, E. Martin, S. Faroux, A. Alias, R. Alkama, S. Belamari, A. Barbu, A. Boone, F. Bouyssel, P. Brousseau, E. Brun, J. C. Calvet, D. Carrer, B. Decharme, C. Delire, S. Donier, K. Essaouini, A. L. Gibelin, H. Giordani, F. Habets, M. Jidane, G. Kerdraon, E. Kourzeneva, M. Lafaysse, S. Lafont, C. Lebeaupin Brossier, A. Lemonsu, J. F. Mahfouf, P. Marguinaud, M. Mokhtari, S. Morin, G. Pigeon, R. Salgado, Y. Seity, F. Taillefer, G. Tanguy, P. Tulet, B. Vincendon, V. Vionnet, and A. Voldoire. The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geoscientific Model Development*, 6(4):929–960, 2013. ISSN 1991959X. doi:10.5194/gmd-6-929-2013. (Cited in page 32.)

W. McKinney. Data Structures for Statistical Computing in Python. In *Python in Science Conference*, pages 56–61, Austin, Texas, 2010. doi:10.25080/Majora-92bf1922-00a. (Cited in page 52.)

C. Meisina, F. Zucca, D. Fossati, M. Ceriani, and J. Allievi. Ground deformation monitoring by using the Permanent Scatterers Technique: The example of the Oltrepo Pavese (Lombardia, Italy). *Engineering Geology*, 88(3-4):240–259, Dec. 2006. ISSN 00137952. doi:10.1016/j.enggeo.2006.09.010. (Cited in pages 2, 6, 13 and 25.)

M. Merk, N. Goeppert, and N. Goldscheider. Deep desiccation of soils observed by long-term high-resolution measurements on a large inclined lysimeter. *Hydrology and Earth System Sciences*, 25(6):3519–3538, June 2021. ISSN 1607-7938. doi:10.5194/hess-25-3519-2021. (Cited in page 28.)

F. Micale, J. Vogt, and C. Cammalleri. European Drought Observatory (EDO) Drought News August 2015. Technical report, Publications Office of the European Union, 2015. URL https://edo.jrc.ec.europa.eu/documents/news/EDODroughtNews201508.pdf. (Cited in pages 86 and 115.)

A. K. Mishra and V. P. Singh. A review of drought concepts. *Journal of Hydrology*, 391(1-2): 202–216, Sept. 2010. ISSN 00221694. doi:10.1016/j.jhydrol.2010.07.012. (Cited in pages 1 and 5.)

R. H. Moss, J. A. Edmonds, K. A. Hibbard, M. R. Manning, S. K. Rose, D. P. van Vuuren, T. R. Carter, S. Emori, M. Kainuma, T. Kram, G. A. Meehl, J. F. B. Mitchell, N. Nakicenovic, K. Riahi, S. J. Smith, R. J. Stouffer, A. M. Thomson, J. P. Weyant, and T. J. Wilbanks. The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282):747–756, Feb. 2010. ISSN 0028-0836, 1476-4687. doi:10.1038/nature08823. (Cited in page 27.)

R. B. Mostafiz, C. J. Friedland, R. V. Rohli, N. Bushra, and C. L. Held. Property Risk Assessment for Expansive Soils in Louisiana. *Frontiers in Built Environment*, 7(October): 1–10, 2021. ISSN 22973362. doi:10.3389/fbuil.2021.754761. (Cited in pages 2 and 6.)

MTES. Cartographie de l'exposition des maisons individuelles au retrait-gonflement des argiles. Technical report, 2021. URL https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2021-06/note_methode_croisement_retrait_gonflement_argiles_juin2021v3.pdf. (Cited in pages 19, 39 and 53.)

MTES. Géorisques, 2024. URL https://www.georisques.gouv.fr/accueil-particulier. (Cited in page 30.)

F. Nachtergaele, H. Velthuizen, L. Verelst, and D. Wiberg. Harmonized World Soil Database Version 1.2. Technical report, FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012. URL https://pure.iiasa.ac.at/id/eprint/17595/2/HWSD_Documentation.pdf. (Cited in page 34.)

N. Nakićenović and IPCC, editors. *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge ; New York, 2000. ISBN 978-0-521-80081-5 978-0-521-80493-6. (Cited in page 27.)

J. Noilhan and P. Lacarrère. GCM Grid-Scale Evaporation from Mesoscale Modeling. *Journal of Climate*, 8(2):206–223, Feb. 1995. ISSN 0894-8755. doi:10.1175/1520-0442(1995)008<0206:GGSEFM>2.0.CO;2. (Cited in page 34.)

J. Noilhan and J. F. Mahfouf. The ISBA land surface parameterisation scheme. *Global and Planetary Change*, 13(1-4):145–159, 1996. ISSN 09218181. doi:10.1016/0921-8181(95)00043-7. (Cited in pages 26 and 32.)

J. Noilhan and S. Planton. A Simple Parameterization of Land Surface Processes for Meteorological Models. *Monthly Weather Review*, 117(3):536–549, 1989. doi:10.1175/1520-0493(1989)117%3C0536:ASPOLS%3E2.0.CO;2. (Cited in pages 26, 32 and 33.)

H. Nowamooz. *Retrait/Gonflement Des Sols Argileux Compactés et Naturels*. PhD thesis, Institut National Polytechnique de Lorraine, 2007. URL https://theses.hal.science/tel-01752833/. (Cited in page 11.)

OFB. Explore2 - des futurs de l'eau, 2024. URL https://professionnels.ofb.fr/fr/node/1244. (Cited in page 86.)

T. Oki, D. Entekhabi, and T. I. Harrold. The global water cycle. In R. S. J. Sparks and C. J. Hawkesworth, editors, *Geophysical Monograph Series*, volume 150, pages 225–237. American Geophysical Union, Washington, D. C., 2004. ISBN 978-0-87590-415-3. doi:10.1029/150GM18. (Cited in pages 1 and 5.)

K. Oleson, D. Lawrence, G. Bonan, B. Drewniak, M. Huang, C. Koven, S. Levis, F. Li, W. Riley, Z. Subin, S. Swenson, P. Thornton, A. Bozbiyik, R. Fisher, C. Heald, E. Kluzek, J.-F. Lamarque, P. Lawrence, L. Leung, W. Lipscomb, S. Muszala, D. Ricciuto, W. Sacks, Y. Sun, J. Tang, and Z.-L. Yang. Technical description of version 4.5 of the Community Land Model (CLM). Technical report, UCAR/NCAR, July 2013. (Cited in page 25.)

W. W. Olive, A. F. Chleborad, C. W. Frahme, J. Schlocker, R. R. Schneider, and R. L. Schuster. Swelling clays map of the conterminous United States. Technical report, 1989. URL https://doi.org/10.3133/i1940. (Cited in pages 11, 12, 16 and 17.)

C. Öser. Determining the plasticity properties of high plastic clays: A new empirical approach. *Arabian Journal of Geosciences*, 13(11):394, June 2020. ISSN 1866-7511, 1866-7538. doi:10.1007/s12517-020-05412-9. (Cited in page 22.)

R. Page. Reducing the cost of subsidence damage despite global warming. *Structural Survey*, 16(2):67–75, June 1998. ISSN 0263-080X. doi:10.1108/02630809810219641. (Cited in pages 17, 18, 19 and 20.)

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12: 2825–2830, 2011. doi:10.48550/arXiv.1201.0490. (Cited in page 52.)

J. R. Price, M. A. Velbel, and L. C. Patino. Rates and time scales of clay-mineral formation by weathering in saprolitic regoliths of the southern Appalachians from geochemical mass balance. *Geological Society of America Bulletin*, 117(5):783, 2005. ISSN 0016-7606. doi:10.1130/B25547.1. (Cited in page 12.)

O. G. Pritchard, S. H. Hallett, and T. S. Farewell. Probabilistic soil moisture projections to assess Great Britain's future clay-related subsidence hazard. *Climatic Change*, 133(4): 635–650, Dec. 2015. ISSN 0165-0009, 1573-1480. doi:10.1007/s10584-015-1486-z. (Cited in page 28.)

P. Quintana-Seguí, P. Le Moigne, Y. Durand, E. Martin, F. Habets, M. Baillon, C. Canellas, L. Franchisteguy, and S. Morel. Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France. *Journal of Applied Meteorology and Climatology*, 47 (1):92–107, 2008. ISSN 15588424. doi:10.1175/2007JAMC1636.1. (Cited in page 38.)

I. Ravina. The influence of vegetation on moisture and volume changes. *Géotechnique*, 33(2): 151–157, June 1983. ISSN 0016-8505, 1751-7656. doi:10.1680/geot.1983.33.2.151. (Cited in page 17.)

République Française. Procédure de reconnaissance de l'état de catastrophe naturelle - Révision des critères permettant de caractériser l'intensité des épisodes de sécheresse-réhydratation des sols à l'origine de mouvements de terrain différentiels, May 2019. URL https://www.legifrance.gouv.fr/circulaire/id/44648. (Cited in page 30.)

République Française. Instruction interministérielle relative à la procédure de reconnaissance de l'état de catastrophe naturelle, Apr. 2024. URL https://www.legifrance.gouv.fr/circulaire/id/45549?origin=list. (Cited in page 30.)

L. A. Richards. Capillary conduction of liquids through porous mediums. *Physics*, 1(5): 318–333, 1931. doi:10.1063/1.1745010. (Cited in page 33.)

Y. Robin, L. Corre, P. Marson, S. Bernus, M. Vrac, and S. Thao. Projections climatiques régionalisées: Correction de biais et changements futurs. Technical report, 2023. URL https://entrepot.recherche.data.gouv.fr/file.xhtml?persistentId=doi:10.57745/99X4CD. (Cited in page 86.)

X. Rodó, E. Baert, and F. A. Comín. Variations in seasonal rainfall in Southern Europe during the present century: Relationships with the North Atlantic Oscillation and the El Niño-Southern Oscillation. *Climate Dynamics*, 13(4):275–284, Apr. 1997. ISSN 0930-7575, 1432-0894. doi:10.1007/s003820050165. (Cited in pages 1 and 5.)

O. Rojas-Munoz, J.-C. Calvet, B. Bonan, N. Baghdadi, C. Meurey, A. Napoly, J.-P. Wigneron, and M. Zribi. Soil Moisture Monitoring at Kilometer Scale: Assimilation of Sentinel-1 Products in ISBA. *Remote Sensing*, 15(17):4329, Sept. 2023. ISSN 2072-4292. doi:10.3390/rs15174329. (Cited in pages 140 and 145.)

M. Roman Dobarco, H. Bourennane, D. Arrouays, N. Saby, I. Cousin, and M. P. Manuel. Propriétés de granulométrie (argile, limons, sables) et d'éléments grossiers pour la France métropolitaine au pas de 90 m, 2022. (Cited in pages 141 and 145.)

J. L. Salagnac. Lessons from the 2003 heat wave: A French perspective. *Building Research and Information*, 35(4):450–457, 2007. ISSN 09613218. doi:10.1080/09613210601056554. (Cited in pages 30 and 73.)

L. Samaniego, S. Thober, R. Kumar, N. Wanders, O. Rakovec, M. Pan, M. Zink, J. Sheffield, E. F. Wood, and A. Marx. Anthropogenic warming exacerbates European soil moisture droughts. *Nature Climate Change*, 8(5):421–426, May 2018. ISSN 1758-678X, 1758-6798. doi:10.1038/s41558-018-0138-5. (Cited in page 28.)

R. Seidel, U. Dettmann, and B. Tiemeyer. Reviewing and analyzing shrinkage of peat and other organic soils in relation to selected soil properties. *Vadose Zone Journal*, 22(5):e20264, Sept. 2023. ISSN 1539-1663, 1539-1663. doi:10.1002/vzj2.20264. (Cited in page 11.)

Y. Seity, P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson. The AROME-France Convective-Scale Operational Model. *Monthly Weather Review*, 139 (3):976–991, Mar. 2011. ISSN 1520-0493, 0027-0644. doi:10.1175/2010MWR3425.1. (Cited in pages 140 and 145.)

S. I. Seneviratne, T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3-4):125–161, May 2010. ISSN 00128252. doi:10.1016/j.earscirev.2010.02.004. (Cited in pages 2, 5 and 34.)

A. Sobaga, B. Decharme, F. Habets, C. Delire, N. Enjelvin, P.-O. Redon, P. Faure-Catteloin, and P. Le Moigne. Assessment of the interactions between soil–biosphere–atmosphere (ISBA) land surface model soil hydrology, using four closed-form soil water relationships and several lysimeters. *Hydrology and Earth System Sciences*, 27(13):2437–2461, July 2023. ISSN 1607-7938. doi:10.5194/hess-27-2437-2023. (Cited in pages 26 and 79.)

Soil Observation Institute. Clay Mineralogy, 2024. URL `https://soilobservations.feedsynews.com/clay-mineralogy/`. (Cited in page 11.)

S. Solomon, G.-K. Plattner, R. Knutti, and P. Friedlingstein. Irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Sciences*, 106(6): 1704–1709, Feb. 2009. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.0812721106. (Cited in page 27.)

W.-K. Song and Y.-J. Cui. Modelling of water evaporation from cracked clayey soil. *Engineering Geology*, 266:105465, Mar. 2020. ISSN 00137952. doi:10.1016/j.enggeo.2019.105465. (Cited in page 17.)

M. Soubeyroux, L. Grigis, F. Rousset, L. Corre, and J. Boe. Evolution des sécheresses météorologiques et agricoles en France en contexte de changement climatique. In *36th Conference of the International Association of Climatology*, 2023. URL `https://drive.google.com/file/d/1Fw18eT6FZz756wwla0vh5Rhxy6XwpNYt/view`. (Cited in page 28.)

J. Spinoni, J. V. Vogt, G. Naumann, P. Barbosa, and A. Dosio. Will drought events become more frequent and severe in Europe? *International Journal of Climatology*, 38(4):1718–1736, Mar. 2018. ISSN 0899-8418, 1097-0088. doi:10.1002/joc.5291. (Cited in page 27.)

G. Stoltz, O. Cuisinier, and F. Masrouri. Multi-scale analysis of the swelling and shrinkage of a lime-treated expansive clayey soil. *Applied Clay Science*, 61:44–51, June 2012. ISSN 01691317. doi:10.1016/j.clay.2012.04.001. (Cited in page 13.)

C. Szczypta, J. C. Calvet, F. Maignan, W. Dorigo, F. Baret, and P. Ciais. Suitability of modelled and remotely sensed essential climate variables for monitoring Euro-Mediterranean droughts. *Geoscientific Model Development*, 7(3):931–946, 2014. ISSN 19919603. doi:10.5194/gmd-7-931-2014. (Cited in page 35.)

D. Tessier, F. Blaise, Y. Coquet, C. Courbet, J. Priem, Y. Lemaire, and S. Trillaud. Analyse comparative de politiques de prévention du risque de "sécheresse géotechnique", Modélisation et visualisation du phénomène. Technical report, INRA, 2006. (Cited in pages 11, 12, 16 and 17.)

A. Toreti, D. Bavera, J. Acosta Navarro, C. Cammalleri, A. de Jager, C. Di Ciollo, A. Hrast Essenfelder, W. Maetens, D. Masante, D. Magni, M. Mazzeschi, and J. Spinoni. Drought in Europe - August 2022. Technical Report August, JRC Global Drought Observatory, 2022. (Cited in pages 80 and 86.)

M. Tuller. Water retention and characteristic curve. In *Encyclopedia of Soils in the Environment*. Academic Press, 2005. ISBN 978-0-12-348530-4. (Cited in page 14.)

P. Tzampoglou, D. Loukidis, and N. Koulermou. Seasonal Ground Movement Due to Swelling/Shrinkage of Nicosia Marl. *Remote Sensing*, 14(6):1440, Mar. 2022. ISSN 2072-4292. doi:10.3390/rs14061440. (Cited in pages 2, 6 and 25.)

M. Th. Van Genuchten. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Science Society of America Journal*, 44(5):892–898, Sept. 1980. ISSN 0361-5995, 1435-0661. doi:10.2136/sssaj1980.03615995004400050002x. (Cited in pages 13, 14, 15 and 150.)

D. P. Van Vuuren, J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. C. Hurtt, T. Kram, V. Krey, J.-F. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S. J. Smith, and S. K. Rose. The representative concentration pathways: An overview. *Climatic Change*, 109(1-2):5–31, Nov. 2011. ISSN 0165-0009, 1573-1480. doi:10.1007/s10584-011-0148-z. (Cited in page 27.)

R. Vautard, J. Cattiaux, T. Happé, J. Singh, R. Bonnet, C. Cassou, D. Coumou, F. D'Andrea, D. Faranda, E. Fischer, A. Ribes, S. Sippel, and P. Yiou. Heat extremes in Western Europe increasing faster than simulated due to atmospheric circulation trends. *Nature Communications*, 14(1):6803, Oct. 2023. ISSN 2041-1723. doi:10.1038/s41467-023-42143-3. (Cited in page 86.)

J.-P. Vidal, E. Martin, L. Franchistéguy, M. Baillon, and J.-M. Soubeyroux. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, 30(11):1627–1644, Sept. 2010. ISSN 08998418. doi:10.1002/joc.2003. (Cited in page 38.)

J.-P. Vidal, E. Martin, N. Kitova, J. Najac, and J.-M. Soubeyroux. Evolution of spatio-temporal drought characteristics: Validation, projections and effect of adaptation scenarios. *Hydrology and Earth System Sciences*, 16(8):2935–2955, Aug. 2012. ISSN 1607-7938. doi:10.5194/hess-16-2935-2012. (Cited in page 28.)

M. Vincent, R. Cojean, J.-M. Fleureau, Y. J. Cui, C. Jacquard, J.-B. Kazmierczak, F. Masrouri, D. Tessier, I. Alimi-Ichola, J.-P. Magnan, M. Blanchard, D. Fabre, A. Pantet, M. Audiguier, E. Plat, H. Souli, S. Taibi, A.-M. Tang, C. Morlock, T. Maison, M. Mrad, N. Bréda, I. Djeran-Maigre, M. Duc, J.-M. Soubeyroux, A. Denis, D. Proust, Z. Geremew, S. Le Roy, M. Dumont, S. Hemmati, H. Nowamooz, Y. Coquet, C. Pothier, L. Makki, M. Chrétien, and C. Fontaine. Rapport de synthèse final du projet ARGIC (Analyse du retrait- gonflement et de ses Incidences sur les Constructions). Technical report, 2009. (Cited in pages 12, 13, 16, 17 and 30.)

M. Waskom. Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60): 3021, Apr. 2021. ISSN 2475-9066. doi:10.21105/joss.03021. (Cited in page 52.)

J. Welch, G. Wang, Y. Bao, S. Zhang, G. Huang, and X. Hu. Unveiling the Hidden Threat: Drought-Induced Inelastic Subsidence in Expansive Soils. *Geophysical Research Letters*, 51 (7):e2023GL107549, Apr. 2024. ISSN 0094-8276, 1944-8007. doi:10.1029/2023GL107549. (Cited in pages 13, 18, 19, 24 and 51.)

D. A. Wilhite and M. H. Glantz. Understanding the drought phenomenon: The role of definitions. *Water International*, 10(3):111–120, 1985. (Cited in pages 1 and 5.)

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Université de Toulouse

Titre : Compréhension et modélisation multi-échelle du phénomène de retrait-gonflement des argiles

Mots clés : Sécheresse, Argiles, Risques

Résumé : Le retrait-gonflement des argiles (RGA), aussi appelé sécheresse géotechnique, est un phénomène naturel de rétractation des sols riches en minéraux argileux gonflants lors de périodes de sécheresse. Le bâti peut être affecté si le dessèchement se propage en profondeur dans le sol sous les fondations provoquant la fissuration des façades. En France, les dégâts sont couverts par les assureurs dans le cadre du régime d'indemnisation des catastrophes naturelles (CatNat). Il s'agit du deuxième risque le plus coûteux après les inondations, avec des dommages estimés à plus de 20,8 milliards d'euros depuis 1989. Les coûts sont particulièrement élevés depuis 2017, avec un nouveau record en 2022 estimé à plus de 3 milliards d'euros d'après la Caisse Centrale de Réassurance (CCR). Bien que ce phénomène ait été décrit localement, il reste imprévisible à de grandes échelles spatiales, en partie à cause de l'hétérogénéité inhérente aux dépôts argileux. La thèse vise à améliorer la caractérisation des facteurs d'occurrence du RGA et à quantifier statistiquement le nombre global de sinistres. Il ne s'agit pas ici de déterminer un risque de sinistre pour chaque maison. Différents types de données sont comparés : des simulations de l'humidité du sol (Météo-France), des données géologiques et des acquisitions faites sur un site instrumenté dans le Loiret (BRGM), et une base de données de sinistres (CCR). Le travail est abordé sous l'angle du facteur déclenchant principal : la sécheresse. Un indicateur sur-mesure est développé, la magnitude annuelle de sécheresse (YDMI), basée sur l'humidité du sol simulée par le modèle de surface terrestre ISBA. Le YDMI intègre les fortes anomalies négatives d'humidité du sol sur une année. Ses caractéristiques (définition de la sécheresse, profondeur de la couche de sol et méthode de calcul) ont été ajustées par comparaison avec un échantillon de données de sinistres. L'objectif de cet indicateur est d'identifier les conditions de sécheresse susceptibles d'induire du retrait. Une comparaison des humidités du sol simulées par ISBA et des YDMI associés avec des données du site instrumenté a permis de caractériser les incertitudes liées à l'échelle macroscopique du modèle. Des YDMI sont ensuite calculés de façon spatialisée sur la France, d'abord pour une période historique allant de 2000 à 2022, puis pour une période future allant jusqu'à 2065. Pour cela, ISBA est forcé successivement par la réanalyse SAFRAN, et par les champs atmosphériques simulés par plusieurs modèles de climat sous les scénarios RCP 4.5 et 8.5. L'analyse historique identifie des événements extrêmes qui ont frappé la France en 2003, 2018, 2019, 2020 et particulièrement en 2022. D'autre part, les YDMI projetés montrent que la fréquence et l'intensité médianes des sécheresses favorisant le RGA sont amenées à augmenter dans le futur, surtout dans le scénario le plus pessimiste RCP8.5. Le dernier volet de la thèse met en interaction le YDMI et trois composantes géologiques de susceptibilité des sols au RGA identifiées par le BRGM, qui sont des critères qualifiant leur minéralogie, lithologie et propriétés géotechniques. Un des enjeux pour les assureurs est de prédire le nombre de sinistres causés par une année de sécheresse. C'est pourquoi le YDMI par années et les critères géologiques ont été croisés avec des données assurantielles pour entraîner un modèle ensembliste d'apprentissage random forest. Les résultats soulignent le pouvoir prédictif supérieur des YDMI et du critère minéralogique pour prédire le nombre de sinistres global. Ces travaux constituent une avancée vers une meilleure compréhension du phénomène RGA et ouvrent des perspectives pour la prédiction opérationnelle de la sinistralité assurantielle. La modélisation pourrait être améliorée avec des données d'aléa plus fines et une meilleure caractérisation de la vulnérabilité des bâtiments.

Title: Understanding and multi-scale modelling of clay shrinkage-swelling

Key words: Drought, Clay, Risks

Abstract: Clay shrinkage is the retraction movements of soils rich in expansive minerals during dry periods. Buildings can be affected if the desiccation extends in depth into the ground below the foundations, causing cracks to appear in facades. In France, this damage, known as clay shrinkage-induced subsidence, is covered by insurers in the framework of the national compensation scheme for natural disasters (NatCat regime). It is the second most costly risk after floods, with damages estimated at more than €20.8 billion since 1989. Costs have been particularly high since 2017, with a new record set in 2022 for which damages were estimated at more than €3 billion, according to the Caisse Centrale de Réassurance (CCR). Although this phenomenon has been described locally, it remains unpredictable at large spatial scales, in part due to the inherent heterogeneity of clayey soils. The aim of this PhD work is to improve the characterization of clay shrinkage occurrence factors and to statistically quantify the global number of insurance claims. The aim here is not to determine a risk of damage for each house. Different types of data are compared: soil moisture simulations (Météo-France), geological data and acquisitions made at an instrumented site in the Loiret French department (BRGM), and insurance claims (CCR). To address the problem, we focus first on the main driver of clay shrinkage: drought. A custom drought index, the year drought magnitude index (YDMI), is developed based on soil moisture simulated by the ISBA land surface model. The YDMI integrates strong negative soil moisture anomalies over one year. Its characteristics (drought definition, soil layer depth and calculation method) were adjusted by comparison with a sample of insurance claims data. The purpose of such an index is to identify drought conditions that are likely to trigger clay shrinkage. A comparison of soil moisture simulated by ISBA and associated YDMI with data from the instrumented site enabled us to characterize the uncertainties associated with the macroscopic scale of the model. The YDMI is calculated for the whole country, first for a historical period 2000-2022, and then for a future period until 2065. For this purpose, ISBA is successively forced by the SAFRAN reanalysis, and by atmospheric fields simulated by several downscaled climate models under RCP scenarios 4.5 and 8.5. The historical assessment identifies extreme events that hit France in 2003, 2018, 2019, 2020 and particularly 2022. On the other hand, the projected magnitudes show that the median frequency and intensity of droughts favoring clay shrinkage are bound to increase in the future, especially under the most pessimistic scenario RCP8.5. The last part of the thesis brings YDMI into interaction with three geological components of soil susceptibility to shrinkage identified by BRGM, which are criteria qualifying mineralogy, lithology and geotechnical properties. One of the challenges for insurers is to predict the number of claims caused by a drought episode. For this reason, YDMI per year and geological criteria were cross-tabulated with insurance data to train a random forest supervised learning ensemble model. The results underline the superior predictive power of YDMI and of the mineralogy criterion. This work therefore represents a step towards a better understanding of the clay shrink-swell phenomenon, and opens up new prospects for the operational prediction of the global number of insurance claims. The modeling could be improved with finer hazard data and better characterization of building vulnerability.