Potentiel des séries temporelles d'images satellites multispectrales pour la caractérisation et le suivi dynamique d'une culture : application à la vigne à l'échelle régionale

Eva Lopez Fornieles

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'INSTITUT AGRO MONTPELLER ET DE L'UNIVERSITE DE MONTPELLIER

En Génie des procédés

École doctorale École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau

Portée par l'Université de Montpellier

Unité de recherche ITAP – Technologies et Méthodes pour les Agricultures de demain

Potential of multispectral satellite image time series for the characterisation and dynamic monitoring of a crop: application to vines on a regional scale

Présentée par Eva Lopez Fornieles
Le 23 septembre 2022

Sous la direction de Bruno TISSEYRE

Devant le jury composé de

Anna Maria DE JUAN CAPDEVILA, Professeure associée, Université de Barcelona, Espagne	Rapportrice
José Antonio MARTINEZ CASASNOVAS, Professeur, Université de Lleida, Espagne	Rapporteur
Agnès BEGUE, Directrice de recherche, Cirad, France	Présidente
Jean Baptiste FERET, Chargé de recherche, Maison de la Télédétection, France	Examinateur
Joaquim BELLVERT, Chargé de recherche, IRTA Lleida, Espagne	Examinateur
Bruno TISSEYRE, Professeur, Institut Agro Montpellier, France	Directeur de thèse
Harold CLENET, Professeur associé, École d'ingénieurs de Purpan, France	Invité
James TAYLOR, Directeur de Recherche, INRAE Montpellier, France	Invité

UNIVERSITÉ DE MONTPELLIER

L'INSTITUT agro Montpellier

Acknowledgements

He de donar les gràcies a moltes persones i, de fet, això ja és un regal en si mateix.
J'ai de nombreuses personnes à remercier et en fait, c'est en soi un cadeau.

À mes collègues :

À toutes les personnes (collègues et amis) qui m'ont accompagné dans cette belle expérience, à commencer par les piliers de la thèse que sont les autres doctorants : Maxime, Yulin, Daniel (surtout Daniel, sans qui je ne peux imaginer passer l'année) et le formidable postdoc Baptiste ;-), ainsi que Belal (accompagné de Kahina, '*aleayila*') et Sherif. Le fait d'être tous ensemble dans le même espace de travail, dans un bâtiment ou un autre, a rendu ma thèse beaucoup plus divertissante et enrichissante. Merci de m'avoir donné plus d'heures de conseils et de café que quiconque le mérite. Un salut spécial aux doctorants de l'équipe ELSA qui sont toujours là pour une conversation autour d'une bonne bière (j'ajoute Simon et Yoann au groupe des bières aussi, on ne laisse personne en arrière). Je tiens aussi à remercier mes autres collègues de l'open space : Basile, Thomas, Léo, Simon, Yoann, Guilhem, Jean-Philippe, Sarah et Bruno, car grâce à eux j'ai découvert ce que signifie une bonne ambiance d'équipe et impossible d'oublier de remercier Laure et Hélène pour leur aide précieuse. Je voudrais aussi remercier TOUTES les personnes qui sont sur le campus de Lavalette. Quel plaisir d'avoir partagé avec vous le jogging, la piscine, le potager, les barbecues improvisés et la vie pittoresque. Silvia (‘*Quin plaer haver-la pogut conèixer fa molt de temps*’) et Flo (‘t*engo tierras’*) merci beaucoup. Je suis extrêmement heureuse d'avoir pu rencontrer tant de personnes avec une si grande gentillesse à SupAgro et à Lavalette. Je ne vous remercierai jamais assez pour votre gentillesse, votre accueil, votre expertise et votre enthousiasme.

Un gros câlin à toutes les personnes qui ont été présentes tout au long du parcours professionnel comme Nina, Cécile, Valeria, Miguel et Maxime Metz. J'espère que tout se passe bien pour vous ! ¡Espero que todo os este yendo bien!

À mon encadrement :

Merci à mon infatigable et talentueux encadrement : Bruno, Nicolas, Jean-Michel, James (thanks for the last proofreading!) et Guilhem. Chacun d'entre vous m'a donné les pièces fondamentales pour pouvoir écrire aujourd'hui ces pages de gratitude. Rien de tout cela n'existerait sans vous. Merci pour tous ces échanges, débats, discussions que nous avons eues au cours de ces trois belles années de collaboration. Ces moments m'ont été très précieux pour arriver à mener au bout cette belle aventure. Vous m'avez beaucoup appris. Un merci particulier à Bruno pour m'avoir fait confiance lorsque que je souhaitais donner une nouvelle orientation à mes recherches, à Nicolas pour m'avoir apporté tout le soutien moral par son infatigable humour et à Jean-Michel pour m'avoir permis de sortir de ma zone de confort (le répondeur d'e-mails le plus rapide du monde !).

Als meus amics/gues :

El meu sincer agraïment als amics i amigues que m'han permès mantenir el cap alt durant molts anys i que han responst a les meves queixes amb bones dosis d'humor aixi com bones dosis de cervesa, vi i "Gintonic con Lima": A la Gina, el Joan, la Laura Llobet, la Mònica, el Jordi, la Sandra, la Judit, la Marta Trius, el Sergi, l'Iban, l'Adrià, el Fernando, la Paula, la Laura '*Uni*' la Jana, la Marta de la Mano, l'Ot, la Belén, el Javi i molts altres que m'han omplert d'alegria i felicitat durant tant de temps. Un grand merci à Antoine et Solène qui ont été ceux qui m'ont aidé dans mes premiers pas dans cette jolie ville de Montpellier.

À ma famille française :

Des bisous, des câlins, des rires et des conversations sans fin avec ma grande famille franco-algérienne, qui est brillante, infinie et inspirante. Surtout à vous Louisa Moussaoui et Abdelghani Cheraiet.

A la meva família :

Agraïment infinit a tota la meva família, tiets, cosins i sobretot a la Iaia. Gracias, Papá. Por haber hecho todo lo posible para permitirme estar donde estoy hoy, te quiero mi cabeza-bolo-barandao. Gràcies, mamá. Per tenir la fe més ferotge en mi tot aquest temps, les '*3c*' provenen de tu. Gràcies Daniel, que has enfrontat tota una vida de jo volent estar a prop teu amb gràcia i paciència, no marxaré mai del teu costat. Un agradecimiento especial a Janine y Alan, que te aguantan y, de paso, me aguantan a mí cuando estamos juntos.

À mon présent et future :

Enfin, merci à toi Anice, mon *Bibi*. Tu as joué le plus grand rôle dans la création d'un monde où monbonheur est si complet que je n'aurais jamais pu l'imaginer. Je t'aime (*Aime-moi*).

Table of contents

List of figures

List of tables

Resumé

Les sources d'information basées sur la télédétection présentent des caractéristiques particulièrement intéressantes pour le suivi dynamique des cultures, de l'échelle de la parcelle à l'échelle régionale. L'imagerie provenant de plateformes de télédétection est capable d'être utilisée pour l'aide à la décision opérationnelle (pour l'expertise ou comme entrée de modèle) pour le suivi des cultures à différentes échelles. Malgré la démonstration de cette capacité dans des études précédentes, le développement de ces nouvelles sources d'information (plateformes et capteurs) progresse plus rapidement que le développement de nouvelles technologies de l'information adaptées à la gestion de cette grande quantité de données. En effet, les informations qui caractérisent ce type de données sont non seulement volumineuses (multidimensionnelles), mais aussi très hétérogènes, ce qui reste un défi pour le traitement des données et les interprétations agronomiques.

Afin de contextualiser cette recherche, ce projet de recherche doctorale s'est concentré sur le potentiel d'une série temporelle d'images satellitaires Sentinel-2 pour le suivi des vignobles à l'échelle régionale dans la région Occitanie (France). Ce jeu de données spatio-temporel Sentinel-2 présente des caractéristiques uniques en termes de temps de revisite, de résolution spatiale, d'informations attributaires fournies et de coût. De plus, le choix de la couverture spatiale est intéressant en soi, car la région viticole de Languedoc-Roussillon représente une grande diversité de conditions agro-environnementales qui se traduit par un grand nombre de cépages différents cultivés ainsi qu'une grande diversité dans les pratiques de gestion des viticulteurs. L'ensemble de ces facteurs introduit des niveaux supplémentaires de variabilité dans l'analyse des données viticoles à l'échelle régionale. Ce travail de thèse est basé sur l'hypothèse que l'évaluation de la variabilité temporelle de l'imagerie satellitaire, en plus des variations spectrales, permettrait une analyse plus complète pour dériver des informations nouvelles et pertinentes sur la variabilité de la production des vignobles individuels à l'échelle régionale. Dans cette optique, l'objectif principal de cette thèse a été d'intégrer des analyses temporelles, en tant que descripteur supplémentaire de la variabilité des vignobles, afin de prendre en compte, d'une manière plus holistique, toutes les dimensions spécifiques des données de télédétection (spectrales, temporelles et spatiales). Différentes méthodes d'analyse multivoie supervisées et non supervisées, dérivées du domaine de la chimiométrie, ont été utilisées, capables de générer des informations à l'échelle régionale à partir de séries temporelles d'images multi-spectrales. Les approches non supervisées ont démontré la possibilité d'extraire des connaissances agronomiques dans le temps (par exemple, différentes dynamiques végétatives) sans prérequis préalables. Les méthodes supervisées ont permis, d'une part, l'évaluation spectrale, temporelle et spatiale d'un événement climatique extrême (par exemple une vague de chaleur) et, d'autre part, la sélection de variables multidimensionnelles

(spectrales-temporelles) pour approfondir la compréhension agronomique de l'impact d'un événement climatique extrême sur la vigne à une échelle régionale.

Ce travail démontre que les méthodes d'analyse exploitant les signatures temporelles et spectrales pour extraire des informations sur les variations de la croissance végétative à l'échelle régionale offrent des informations précieuses pour évaluer la performance des cultures individuelles. En tenant compte de la haute dimensionnalité des données, qui inclut la dimension temporelle, les besoins ainsi que les limites de l'analyse des séries temporelles sont explorés dans le contexte de la fourniture d'informations pertinentes pour aider à la connaissance à grande échelle d'une culture, telle que la vigne.

Abstract

Information sources based on remote sensing have particularly interesting characteristics for dynamic crop monitoring, from the plot scale to the regional scale. Imagery from sensing platforms is capable of being used for operational decision support (for expertise or as model input) for crop monitoring at different scales. Despite the demonstration of this capability in previous studies, the development of these new sources of information (platforms and sensors) is progressing more rapidly than the development of new information technologies adapted to the management of this vast quantity of data. Indeed, the information that characterises this type of data is not only large (multidimensional), but also very heterogeneous, which remains a challenge for data processing and agronomic interpretations.

In order to contextualise this research, this doctoral research project has focused on the potential of a time-series of Sentinel-2 satellite images for monitoring vineyards at the regional scale across the Occitanie region (France). This spatio-temporal Sentinel-2 dataset presents unique characteristics in terms of revisit time, spatial resolution, attribute information provided and cost. Moreover, the choice of spatial coverage is interesting in itself, as the Languedoc-Roussillon wine region represents a great diversity of agri-environmental conditions resulting in a large number of different grape varieties being cultivated as well as a large diversity in the management practices of the wine growers. Collectively these factors introduce additional levels of variability into the analysis of regional-scale viticulture data. This PhD work is based on the assumption that assessing the temporal variability in the satellite imagery, in addition to spectral variations, would allow a more complete analysis to derive new and relevant information about production variability of individual vineyards at the regional scale. With this in mind, the principal objective of this thesis is to integrate temporal analyses, as an additional descriptor of vineyard variability, in order to take into account, in a better and more holistic way, all the specific dimensions of remote sensing data (spectral, temporal and spatial). Different supervised and unsupervised multi-way analysis methods, derived from the field of chemometrics, were used, capable of generating information at the regional scale from time series of multispectral images. Unsupervised approaches demonstrated the possibility of extracting agronomic knowledge over time (e.g. different vegetative dynamics) without *a priori* knowledge. The supervised methods allowed, firstly, the spectral, temporal and spatial assessment of an extreme climatic event (e.g. a heat wave) and, secondly, the selection of multidimensional (spectral-temporal) variables to deepen the agronomic understanding of the impact of an extreme climatic event on grapevines at a regional scale.

This work demonstrates that analysis methods exploiting temporal and spectral signatures to extract information on regional-scale variations in vegetative growth offer valuable information for assessing individual crop performance. Taking into account the high dimensionality of the data, which includes the temporal dimension, the needs as well as the

limitations of time series analysis are explored in the context of providing relevant information to aid large-scale knowledge of a crop, such as grapevines.

Foreword

Successive advances in remote sensing platforms, sensors and computing techniques are leading to a significant growth in remotely sensed data and applications. Many projects dedicated to environmental and crops monitoring are exploiting multi-temporal and multi-sensor remote sensing data at different spatial scales for processing high-dimensional data. Indeed, large-scale remote sensing applications, i.e. covering large geographical areas being dominated by massive remotely sensed data, are considered as data-intensive problems. In agriculture, the potential of remotely sensed time series to support management practices has never been greater; however, the complexity of the remote sensing data as well as the large datasets involved often limits this potential. How plants interact with sunlight is fundamental to the existence of life, it also provides a window into the functioning of agricultural ecosystems. The basic properties of vegetation spectra have been known for decades, but interpretation of the spectra remains difficult because, in agriculture, canopy response (reflectance) is affected by multiple factors that can range from spatial variability of soil-climate conditions to agricultural practices. Based on the assumption that the spectral response of the crop makes it possible to detect the onset of processes, such as the beginning of crop growth, as well as the occurrence of incidents (e.g. extreme weather events, diseases, etc.), remote sensing is a way to design effective and objective methods for assessing crop characteristics. Furthermore, crops have a temporal and seasonal behaviour, i.e. their phenology, which makes phenological diversity a factor related to the diversity of spectral responses. Therefore, temporal variations can be used in addition to spectral variations through time series of satellite images to monitor agriculture production over large areas. In this context, the Sentinel-2 satellite mission, which provides free multispectral time series for continuous vegetation monitoring, offers great promises.

Time series analysis of imagery is becoming increasingly important for agricultural monitoring. Increasing the use and usefulness of time series data is in the interest of understanding agricultural crop dynamics and factors that may affect this dynamic. Understanding the latter in relation to the changes taking place on our planet could generate value-added products relevant for informed decision-making by stakeholders, for example, by providing interpretative insights into the effects of climate on annual and perennial crops. Therefore, the importance of properly considering the temporal dimension in the analysis of remote sensing data is of paramount importance. In this thesis, the proposition is made to study the construction of high-dimensional spaces using all values of the time series, together with advanced multidirectional chemometric methods, as a robust and efficient approach for crop monitoring. Considering a regional decision scale, the object of this thesis will be to propose a methodological framework to assess the relevance of using the temporal information provided

from satellite image time series to identify, describe and predict patterns with specific spectro-temporal signatures that contribute to improve crop knowledge at large scales. Integration of the time dimension should be seen as a key step in facilitating the progress of the research and potential future services for the agriculture sector.

In order to answer these questions, this manuscript is divided into six chapters. After a first chapter specifying the general context and the scientific question to address, there are four chapters detailing the research undertaken in the thesis. Each of these four chapters consists of one or two scientific articles (4 published and 1 submitted). Each article (or pair of articles) is preceded by an introduction and followed by a conclusion that summarises its position in the proposed scientific approach. The last chapter is a general synthesis proposing different perspectives arising from this research work.

Chapter 1. Introduction

1.1 How to monitor crops with remote sensing: state of the art

1.1.1 Generalities: remote sensing development for agriculture

In the last decades, remote sensing (RS) techniques have provided valuable insights into agriculture management by determining physiological and phenological status of the crops (Thenkabail et al., 2011). Beginning with Gates et al. (1965), Allen et al. (1969), Gausman et al. (1969) and Woolley (1971), who pioneered the fundamentals of how optical properties change as a function of plant morphological characteristics (e.g. canopy architecture, leaf thickness or water status). Research followed in the 1970s, particularly linked to the first Earth Observation satellite systems, with Tucker (1979) demonstrating that vegetation can be monitored from spectral properties of reflectance. The exponential growth of RS studies applied to agriculture between 2000 and 2019 suggests progress in the technologies used, including (Khanal et al., 2020):

- unprecedented combinations of spatial, temporal and spectral capacities depending on the platform or type of sensor used;
- advent of new platforms such as Unnamed Aerial Vehicles (UAV) in addition to satellites or aircraft;
- the development of cloud computing and data storage techniques (Khanal et al., 2020).

In particular, the increase in agricultural application studies had an important focus on Unnamed Aircraft Vehicle (UAV) as well as satellite/airborne-based RS studies. Increase in RS studies is mainly explained by the free access to a large amount of historical satellite imagery, such as Landsat, and more recently the Sentinel constellations as well as by the accessibility of powerful data platforms and services, such as Google Earth Engine (GEE) (Bégué et al., 2018). Khanal et al. (2020) identified that the number of RS-related studies for agriculture is now 20 times higher than it was in the early 2000s, with the majority of studies related to multispectral and hyperspectral applications, especially since 2017. Thus, in view of the trends in RS research, the main historical limitations associated with the inadequate availability of time series that offer sufficient temporal and spectral resolution as well as data storage, distribution and computing issues, seems to be gradually fading away, meeting the long-lasting expectations of RS applied to agriculture.

As RS allows the acquisition of information on objects or phenomena from a distance and in a non-destructive way (Gitelson et al., 2001), it is considered as an essential tool for agriculture that permits the deepening of knowledge on crops and their temporal evolution. However, its application is related to several specificities that involve different spatial scales (e.g. local, plot or regional), different temporal scales from real-time acquisitions to decades

later, as well as the different levels of spectral accuracy required for a proper knowledge of crop status (Weiss et al., 2020). Therefore, monitoring of agriculture using RS is a broad topic that has been widely approached from multiple perspectives. RS is discussed, for example, for specific agricultural applications (e.g. yield prediction, irrigation monitoring, etc.), on specific RS platforms (e.g. satellites, UAV), sensors (e.g. active or passive sensing systems) or specific locations and climatic contexts (e.g. intra-plot, or region, wetlands or drylands) (Weiss et al., 2020).

1.1.2 Remote sensing for agriculture applications

1.1.2.1 Specific agriculture applications

Remote sensing, together with other advanced techniques, such as global navigation satellite systems and geographic information systems, is playing an important role in the assessment and management of agricultural activities (Shanmugapriya et al., 2019). The number of applications of RS in agriculture is vast, as it can serve multiple purposes: making agricultural activities economically viable, securing agricultural production to feed a growing population, reducing negative environmental impacts by minimising resource depletion and contributing to climate mitigation at a local and global scale. Weiss et al. (2020) highlighted a number of general current trends that address the multiple objectives outlined above. The applicability of RS ranges from selecting varieties better adapted to challenging contexts (e.g. climate change), to monitoring land use or crop growth, forecasting in-season crop yields, optimising short-term production and providing ecosystem services related to soil or water resources, as well as monitoring biodiversity. In recent years, work in RS for agriculture has increasingly focused more on characterising the biophysical properties of plants to provide valuable information on various agronomic parameters, such as crop phenology stages, and on the detection of stress situations and other disturbances (Atzberger, 2013).

Remote sensing relies on how radiation reflectances are affected by the properties of observed surfaces to extract information of different kinds (Soudani, 2005). The information of interest is based on traits or features of the crops or agricultural systems and, above all, how these later vary in space and time. Weiss et al. (2020) established six categories of agronomic features that can be studied with RS: chemical (e.g. leaf nitrogen content), physical (e.g. soil moisture), typological (e.g.; crop type and variety), biological (e.g. phenological stages and growth), structural (e.g. leaf inclination and position) and geometrical (e.g. leaf density and shape). The relationship between what is directly measured by RS instruments and the agronomic features themselves needs to be modelled in some way (more or less substantially) to infer the latter from the former (Weiss et al., 2020). This definition of the relationship involves at first, the choice of an instrument or a sensor mounted on a distal platform, such as a satellite, an aircraft, an UAV/Unnamed Ground Vehicle (UGV) or a mobile or static terrestrial

platform or a probe (located on the plant or on the soil). It should be noted that although all of the above platforms could be labelled as remote sensing, the UGV/terrestrial platforms and the probe are considered as proximal detection sensors, and not as sensor-based remote sensing. As a result, although some strong similarities may be found, terrestrial and probe platforms will not be considered as RS in the following discussion.

Since the sensor measures electromagnetic radiation that is reflected or emitted by the target, the type of information that can be accessed depends on the specific properties of the instrument and the specific properties of the target (Navalgund et al., 2007). Thus, from Weiss et al. (2020), spectral, directional and polarization capabilities, spatial resolution, revisit frequencies (temporal resolution), as well as signal-to-noise ratios are defined according to the sensor/platform properties.

The understanding of how to relate an agronomic trait to a signal response, has been used to quantify various agronomic parameters. Thus, it is also important to know what information one wants to obtain from reflectance measurements in order to know the type of possible instruments to use (Thenkabail et al., 2011). Bridging the gap between the physical RS measurement and agronomic variables, Baret et al. (2007) determined that depending on the spectral domain under consideration, different physical processes may be involved in:

- the visible to shortwave infrared domains, the physical quantity considered is the reflectance, meaning the fraction of radiation reflected by the surface that relates to canopy structure variables;
- the thermal infrared, the canopy is characterised by its brightness temperature according to the emitted radiation flux and reflects leaf surface temperature;
- the active microwave, the backscatter coefficient from the canopy is measured by radar systems and relates to physical conditions such as water content);
- passive microwaves, the canopy is characterised (similarly to the thermal infrared domain) by its brightness temperature, but here, emissivity plays a significant contribution and can be related to physical-chemical properties such as soil organic carbon.

Once an appropriate instrument (platform and sensor) for agricultural data has been defined by the nature of the problem to be studied (what to measure, why, how and where) (Khanal et al., 2020), the second step concerns deciding on the type of approach to be used to model the relationship between RS instruments and the agronomic features. There are many approaches to retrieve agricultural variables from remotely sensed data but they basically fall into three categories (Weiss et al., 2020):

- purely empirical methods, which entails directly calibrating a numerical relationship between the measured signal and one or several variables of interests. Basic examples would be linear and non-linear regressions with either classical statistical approaches or machine learning algorithms;

- mechanistic methods consisting of the inversion of models based on radiative transfer theory, Maxwell's equations or optical and projective geometry. The application domains are solar energy and microwaves, radar interferometry and polarimetry, and LIDAR and photogrammetry;
- 'contextual methods' that exploit the spatial and/or temporal contrasts over thermal and solar spectral domains of images, mainly to assess plant water status with regard to the evapotranspiration process (Stefan et al., 2015).

1.1.2.2 Specific platform and sensor types

Although a large number of RS instruments are already now available, new sensors are still being developed to provide greater spectral resolution with more targeted information that is compatible with the essential requirements for agricultural applications. Improvements deal with the revisit time of less than one week and the spatial resolution of less than 10 metres (Qiu et al., 2019). For a proper monitoring of agricultural systems, and considering crop evolution and changes over time, it is important to take into account the different instruments, as well as their measurement conditions and possible disturbances, in order to properly characterise the signal provided by the target (Khanal et al., 2020). Broadly, RS platforms are divided into satellite, aerial and Unnamed Aerial Vehicle (UAV), and RS sensors into visual, thermal, multispectral, hyperspectral and microwaves (radar). The first four of these fall into the category of passive sensors and microwave radar into the category of active sensors. Thermal sensors have proved to detect soil moisture and plant water stress (Hassan-Esfahani et al., 2017). Multispectral and hyperspectral optical (visible and near infrared) sensors are mainly used to detect crop health patterns beyond the visible spectral domain using several narrow spectral bands, thus improving the characterisation of crop growth (Yao et al., 2017). It should be noted that hyperspectral sensors have a greater number of spectral bands in close vicinity to each other compared to multispectral sensors. Finally, microwave sensors have the advantage of being independent from atmospheric conditions. They allow the biophysical characteristics of the crop to be determined both day and night (Valcarce-Diñeiro et al., 2019).

Currently, multispectral, hyperspectral and visual sensors are the most common form of sensor available from satellites used for agricultural applications (Khanal et al., 2020). Compared to satellite platforms, UAVs remain less widely used but are gaining popularity due to their flexibility in terms of revisit frequency. However, their use presents some drawbacks, such as their dependence on weather conditions and their operational limits related to their autonomy (e.g. battery power) (Herrero-Huerta et al., 2014). Moreover, as the optical domain is already considerably exploited, new research with other types of sensors such as thermal infrared is starting to gain popularity especially for water resources management (Lagouarde et al., 2019).

Operational constraints

Currently, the vast majority of RS technologies, i.e. platforms and sensors, are accessible and relatively affordable for the agricultural sector. However, technological advances are not keeping pace with the development of the sector concerned, so it has not yet been possible to fully adapt these useful technologies for the following three reasons (Khanal et al., 2020):

- Efficacy: the choice of the instrument to be used is crucial in order to adequately address the nature of the problem. As Srivastava et al. (2020) showed, the use of hyperspectral sensors has made it possible to determine and discriminate specific features using hundreds of narrow spectral bands. However, the increase in spectral resolution (bands) has caused an increase in the size of the dataset, leading to a higher level of complexity related to data storage and subsequent pre- and post-processing. In addition, hyperspectral sensors are still considered expensive compared to other sensors on the market. Thus, their efficiency for the characterisation of agronomical variables in a commercial context may not compensate for the difficulty of data management as well as their price.
- Economics: There are different platforms for RS data collection and each has arguments for and against. In agriculture, the two most common platforms are satellites and UAV. Regarding satellites, there are a considerable number of open-access satellite data streams with medium resolution (i.e. $\geq$ 10 m pixel size), such as Landsat and Shuttle Radar Topography (SRTM) among others. These types of satellites are used to cover large areas for free but, as the spatial resolution increases, so does the cost; for instance, imagery with higher resolution (5 m pixels) from the RapidEye satellite costs 1.28 USD/km^2. Regarding UAVs, they offer high-frequency data at a much smaller scale, allowing the monitoring of crop growth with a high spatial resolution at plant level. However, in addition to the limitations already mentioned, their average price has been estimated to be around 7.4 to 12.4 USD/ha (Khanal et al., 2020), which implies that if the study area is large, the cost can be a major constraint to their use. Therefore, the majority of studies related to RS applications in agriculture continue to be limited to a low or medium spatial/temporal scale.
- Availability of RS data, cloud computing and processing: the amount of data available today is greater than ever before. Although having large amounts of data can only be seen as a positive factor, there are several challenges associated with the limits of computer storage and computation as well as different resolutions (spectral-spatial-temporal resolutions), which have to be taken into account to study underlying processes of dynamic phenomenon like crops. For these reasons, the exploitation of remotely sensed data is still a challenge to reach its full potential for applications in agriculture.

Data acquisition and methods: the impact of data resolution

The main sensor/target characteristics, which facilitate sensor/target discrimination, are spectral, spatial, temporal, and polarisation signatures. While polarisation is a specificity of the radar sensors, with sensitivity to identified variables such as cloud particle size distribution, soil texture, agricultural crops (Egan, 1992), the other characteristics are common for most of the other sensors. The spectral resolution determines the amount and size of spectral sampling intervals provided from the electromagnetic region. Spatial resolution refers to the pixel size of an image that conditions the ability of the sensor to detect the target, while temporal resolution is defined by the revisiting time between two successive acquisitions (Navalgund et al., 2007).

Depending on the resolution of sensor properties, exploration and knowledge creation using search, mining and modelling techniques will be adapted to different volumes and dimensions of data, hence the importance of prior knowledge of the phenomena to be investigated. When using high-resolution images, either spectral (e.g. hyperspectral), spatial (e.g. 50 cm for GeoEye-1 satellite platform) or temporal (e.g. UAV) resolution, it has been shown that studies tend to focus on a local context, i.e. small areas, due to the large processing needs in terms of time, cost and skills (Khanal et al., 2020). However, while the adaptation of some agricultural practices for sustainable development requires working at a spatial resolution of at least the decimetre, other variables of interest, such as yield estimation, can be done at larger scales (e.g. hectometre or kilometre). Similar reasoning also applies to the temporal and spectral resolution scale. Regarding temporal resolution, an example of monitoring that requires specific (but weather-dependent) revisit times to obtain reliable predictions is that of nitrogen fertilisation, as imaging needs to be performed at specific growth stages (Verrelst et al., 2015). In terms of spectral resolution, a sensor that only provides information in the visible spectrum is limited to applications such as crop emergence detection and crop classification. However, if the sensor is also capable of capturing spectral regions such as the Near-InfraRed (NIR) and/or thermal bands, it is useful for other applications, such as detecting crop stresses that are invisible to the human eye. Furthermore, as the spatial scale increases (e.g. from local to regional to global), the link between the measured variables and the agronomic information of interest becomes more complex to establish. This is explained because the factors linking the spatial, temporal and spectral dimensions are variable and dependent on the environment (e.g. soil type and climate), the cropping system (e.g. field size, variety and planting pattern) and the conditions during image acquisition (e.g. shade/sunlight), which implies that RS data for large-scale agricultural monitoring needs to overcome these issues when the goal is to provide wide-scale timely information on crop production, health or nutrient status and yield prediction.

1.1.2.3 Specificities related to location and climatic context

Experiences acquired from monitoring specific crops in specific locations and climatic conditions can be very informative to assist in adapting cultural practices in the same locations

under different climatic conditions or in other locations with comparable climatic conditions (Weiss et al., 2020). As an example, monitoring the phenological development of crops not only helps to identify different types of land cover and crops (and crop growing conditions), but also provides evidence of ongoing global/climate change (Atzberger 2013). According to Atzberger (2013), the variables of interest where RS can be relevant to respond to the above-mentioned insight at regional scale are: biomass and yield, vegetation vigour and drought (stress monitoring), crop phenological development, crop area estimation and cropland mapping and disturbances and land use/land cover changes. To obtain this information, a large number of methods and tools have been developed, first focusing on spectral analysis but later evolving to obtain useful:

- information by analysing the temporal signature (Wardlow et al., 2007);
- knowledge of the directional reflectance properties of vegetation (Clevers et al., 1994);
- information from the spatial arrangement of pixels (e.g. image texture) (Irons and Petersen, 1981).

The FAO (2010) reports that agricultural production follows strong seasonal trends linked to the biological life cycle of crops, the field environment (e.g. soil type), climatic variables and agricultural management practices. Moreover, as productivity can change over short periods of time, due to unfavourable growing conditions, the location and climatic context are of utmost importance for relevant crop monitoring.

1.1.3 Remote sensing for agriculture: challenges and opportunities

Agriculture is currently in a challenging context given the strong need to monitor crop growth and conditions in various locations and environmental contexts, with various temporal resolutions for different purposes (Bégué et al., 2018). As the number of satellite instruments and the quality and scope of information collected keeps on increasing, providing greater crop coverage with shorter revisit times and with the availability of archival data, the need for efficient algorithms able to cope with such an amount of data will increasingly require new methodological approaches to be used in agricultural monitoring (Dalla Mura et al., 2015). There are several common challenges in RS related to the increasing degree of diversity and complexity of RS data. Non-exhaustively, the three main challenges that the literature raises are (Ma et al., 2015):

- the large volume and rate of RS data;
- the diversity of RS data;
- the complexity of RS data, especially when considering high dimensionality (either spectral, temporal or spatial).

Regarding the first challenge, although the interest of applying remote sensing techniques in agriculture has been demonstrated since the 1970s, the operational use of remote

sensing data in this field has recently intensified with the collection of higher resolution data that are able to perceive a scene with more precision (i.e. provide more detail), thus considerably increasing the amount of information to be collected. Furthermore, the inherent high dimensionality of satellite data implies more complex data processing algorithms to use these information sources to their full potential (e.g. deep learning methodologies) (Defourny et al., 2019).

Regarding the second challenge, one of the main issues is to take advantage of all available information by combining remote sensing observations of different natures in terms of temporal, spatial and spectral dimensions, as well as ancillary data from other sources (e.g. other sensors or ground measurements), in order to be able to better monitor and characterise crops (Huang et al., 2018). However, combing remote sensing observations of different natures gives rise to new problems in high-dimensional data processing. The main question is how to organise multidimensional remote sensing data in a 1-D, 2-D or 3-D array to best process/analyse the data. In particular, the organisational structure of the data has to be chosen in such a way as to improve its availability and the consequent extraction of useful information.

The third challenge is to preserve the semantics, character and shape of the original RS time series when considering high dimensionality. To address this challenge and thus exploit all dimensions of remote sensing imagery in a meaningful way, dimensionality reduction is one of the main approaches to process RS data (Liu, 2015). By reducing the dimensions and size of the data, it is possible to extract the core of the dataset. In an example applied in agriculture, it was shown that the semantics, character and shape of the data can be lost if the appropriate sampling time window was not precisely specified (Schneider and Xhafa, 2022). This example of an appropriate time window for a specific agricultural application also highlights the need for the correct use of technical knowledge (e.g. crop types and crop calendars) to support the choices between the wide range of RS data sources and data processing methods. This would reduce processing time, while producing relevant results (Weiss et al., 2020).

These three main challenges highlight some of the limitations of current processing and analysis methods (outlined above). Nevertheless, opportunities are emerging for future technological development in remote sensing.

Firstly, the increase in the spatial and temporal coverage and resolution of satellite observations through regular acquisitions has made data fusion a very active research area. Zhang (2010) defined the data fusion approach as combining data from multiple sources to: (i) improve the potential value and interpretation performance of the data sources, and (ii) produce a high quality visible representation of the data. For example, to improve the spatial and temporal resolution of multispectral imagery, Wu et al. (2018) proposed a data fusion approach that combined 250 m Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) and 10 m Sentinel-2 NDVI data to generate a synthetic Sentinel-2 NDVI time series for monitoring disease in cotton crops. Data fusion practices are now widely used in many applied remote sensing tasks, such as estimating physiological stress

(Sagan et al., 2019) or monitoring site-specific crop parameters, such as nitrogen, chlorophyll, leaf area index and aboveground biomass, either from space (e.g. satellite) or from aerial (e.g. UAV) platforms (Ahmad et al., 2022). These types of techniques have shown the potential to automatically extract spatio-temporal relationships to obtain more useful information to improve the prediction of observed physical phenomena at multiple spatial and temporal scales. These methods are very attractive for satellite data analysis, especially if the versatility of data-driven machine learning methods are combined with physical process models (Reichstein et al., 2019). However, the evaluation of these advanced models for crop monitoring is still in the development phase, as accurate inter-calibration between instruments is an ongoing challenge for the RS community (Dubovik et al., 2021).

Secondly, although there has been progress in addressing the increasing degree of diversity and complexity of data over the last decade, studies have mainly focused on demonstrating the feasibility and operability of techniques and methods that had previously been developed for near-real time and operational needs (Weiss et al., 2020). The development of a near real-time decision process is the main opportunity from a near-future perspective for operational demands. This will require the implementation of sufficient temporal coverage (revisit time) and rapid data processing to reduce the time between acquisition and the output of the results necessary for the decision maker (farmer) to make decisions at the field scale. In this way, it would be possible to take full advantage of the characteristic properties of RS sensors to provide a comprehensive understanding of the factors affecting agricultural crops (Hatfield et al., 2019). Consequently, a lot of effort has been devoted to machine learning and deep learning approaches that address operational and near real-time needs, either to solve classification problems (Ray, 2019) or to establish complex empirical relationships to estimate crop variables using RS data (Terliksiz and Altýlar, 2019). However, exploiting the full potential of the high dimensionality of RS data with ML or DL approaches for agricultural applications is still in its infancy (Khanal et al., 2020).

Thirdly, according to Liu (2015), the high dimensionality of remote sensing data for agriculture is mainly reflected in the spectral and temporal dimensions. Therefore, integrating the spectral and temporal component into the analysis is a challenge both to generate new knowledge and to identify the potential of RS for crop monitoring. Nowadays, the short revisit time of new satellite systems offers a new potential for the use of multispectral and multi-temporal data in crop monitoring. Thus, given these short temporal revisit times, multi-temporal and multi-spectral satellite data can offer additional opportunities for crop monitoring.

In conclusion, the development of massive data storage and the generalisation and the standardisation of remote sensing data is the inevitable trend in the future technological development of remote sensing.

1.1.4 Multitemporal multispectral data

As a consequence of the increased amount of data available, particularly from satellite platforms, methodological approaches have shifted towards a greater consideration of the temporal and spectral dimensions. Concerning the temporal dimension, initial remote sensing approaches only considered a single date (Woodcock et al., 2020), i.e. using single date images that record certain phenological events in a period that is empirically considered to be the best time for identification. However, it has been shown that this type of single-date approach does not guarantee having the 'best' possible images (e.g. cloud cover is likely to prevent or delay image acquisition in many locations) and, additionally, single date imagery is often uninformative for capturing crop dynamics (e.g. phenology). Adding to these limitations is the fact that in recent decades the cost of storing RS data has decreased dramatically, resulting in an overwhelming increase in computing power, which explains the growing importance of time series analysis for crop monitoring (Zhu, 2017). Currently, the focus has shifted to multi-date image analysis instead of single-day image analysis (Ma et al., 2019). This transition started with the realisation that the use of satellite images collected by the same sensor on different dates, but covering the same geographical area, is advantageous for finding either natural (e.g. growth rate variation) or human-induced (e.g. crop management) changes related to the crop (Woodcock et al., 2020). Therefore, change detection analysis was the first methodological approach applied to exploit multi-temporal imagery. In the late 1980s, Singh (1989) defined the concept of change detection as the process of identifying differences between images at different points in time (Jianya et al., 2008). However, according to Woodcock et al. (2020) there is currently a transition from change detection to remote sensing monitoring. As the time interval between observations used in the time series becomes smaller (e.g. theoretical Sentinel-2 satellites have a revisit time of 5 days), data processing approaches have begun to move from simple change detection to more continuous monitoring (Yang et al., 2019). At present, most change detection methods correspond to the bi-temporal approach, i.e. the comparison between two dates (Woodcock et al. 2020). In contrast, time series analysis is based on a 'continuous' time scale, i.e. the focus of the analysis is not only on what has changed between dates, but also on the progress of change over a considered period (Jianya et al., 2008). Since time series analysis emphasises the uncovering of the trend of change by constructing multi-temporal data 'profiles', this shift in temporal focus has several implications: (i) it allows for a more accurate characterisation of the timing of change; (ii) it provides a better determination of the drivers of change since the characterisation of the timing of change is improved; and (iii) it opens the way to enable near real-time monitoring, i.e. changes can be detected quickly after new satellite imagery becomes available (Jianya et al., 2008). These aforementioned functionalities, which are addressed by time series analysis, are essential for accurate crop monitoring, as each crop type has a well-defined crop calendar with specific crop operation times and unique seasonal growth patterns (Loew et al., 2013). Consequently, there is a strong need for a global

understanding of the optimal temporal windows and the impact of the quality and quantity of temporal features on crop monitoring (Hu et al., 2017), for example to improve estimates of crop state variables (Roznik et al., 2022) or to improve modelling of vegetation functioning (Zhang et al., 2003).

Concerning the spectral dimension, it is the main factor that needs to be thoroughly considered for crop monitoring. It has been demonstrated that the spectral reflectance of crops is highly related to leaf pigment, leaf water and canopy structure, and varies with the growing season (Hu et al., 2017) and can exhibit different biochemical and structural properties at each growth stage. Moreover, it can undergo different cultural operations (e.g. soil preparation, weeding, pruning or trimming), which may be very different from one crop to another, and sometimes from one pedo-climatic context to another for the same crop throughout the growing season. Therefore, it is essential to assess the usefulness of spectral features for crop monitoring and to understand which spectral features are or are not important over time. Early remote sensing time series studies have used a limited number of variables or Vegetation Indices (VIs). Given their ability to estimate vegetation cover, vigour and growth, more than 40 VIs have been developed and used in a wide range of studies over the past 50+ years (Xue and Su, 2017). As the analysis of the spectral dimension has become a fundamental scientific tool in many subfields that use time series data, the use of multiple indices for a more holistic understanding or the modelling of spectral RS data to develop time series of vegetation-related variables has been a critical subject of study (Southworth & Muir, 2021). However, although the use of VIs is widespread due to the simplicity of their applications, it has been shown that relying on VIs, rather than on ‘continuous’ spectra, can lead to a potential misidentification of multiple and crucial characteristics when dealing with crop monitoring (Hu et al., 2017). There are several types of approaches that allow for a consideration of the whole spectral resolution of a multispectral sensor, such as multivariate regression methods based on chemometric techniques that extract information from chemical systems (Li et al., 2020) or physics-based reflectance methods (Jacquemoud et al., 2009). The latter methods are used to study possible changes in the ecophysiological and biophysical state of crops, for example with the PROSAIL model, which simulates the reflectance of the canopy as a function of various ecophysiological variables (Berger et al., 2018).

Having seen the potential of multi-temporal and multi-spectral data, it is important to reposition this potential in the face of the limitations related to high dimensionality data. The complexity of the vast majority of existing algorithms for time series analysis is exponential with respect to the number of dimensions, i.e. as dimensionality increases, these algorithms soon become computationally more difficult and, hence, inapplicable in many real applications (Zhou et al., 2017). Furthermore, as noted by Woodcock et al. 2020, the context is important in assessing the relevance or significance of any particular change, which means that the number of spectral channels and the lengthening of the data time series does not necessarily mean that the amount of effective information for crop monitoring increases proportionally. The

information gain depends on the mutual independence of co-occurring measurements (Carrão et al., 2008). These drawbacks explain why some authors still focus on both temporal and spectral low-dimensional feature space studies using, for example, bi-temporal change methods to analyse crop changes (Hussain et al., 2013) or a single vegetation index to characterise the whole crop dynamics in the growing season (Tian et al., 2015).

1.1.5 Time series analysis: remote sensing methods for crop monitoring

Time series data from remotely sensed imagery with high temporal and medium or low spatial resolution can assist crop monitoring by providing key information on crop development during a growing season over large areas (Rembold et al., 2015). Remote sensing time series studies for crop monitoring mainly concern the study of VIs (e.g. Enhanced Vegetation Index, EVI) and biophysical variables (e.g. Leaf Area Index, LAI) (Zeng et al., 2020) enabled by gap-filling, cloud removal and shadow algorithms (Liu, 2015). Data models have also been used to develop time series of vegetation-related variables, such as the reconstruction of LAI time series by data fusion (Zhou et al., 2020). However, due to the increased availability of temporal data from satellites, greater emphasis has been placed on the development of new approaches, which allow researchers to examine temporal trends (Southworth & Muir, 2021). There are several types of time domain methods used for understanding the relationship between remotely sensed time series and agronomic variables, including parametric and non-parametric methods, which can be linear or non-linear and these in turn can be univariate and multivariate (Mishra et al., 2016). Furthermore, the introduction of Artificial Intelligence (AI) has accelerated the progress in time series processing. AI advances have taken time series research in two main directions: (i) new methods for constructing time series datasets and, consequently, (ii) new methods for the data mining of these time series (Southworth & Muir, 2021). The construction of new time series refers to the fact that AI has provided better data fusion and integration and analysis of data from multiple sources and dimensions (Liu et al., 2019; Peng et al., 2020; Chen et al., 2021). This is exemplified by methods such as the multi-temporal fusion approach applied to time series collected from ENVISAT/MERIS and Landsat/TM instruments that obtained coherent time series at high spatial resolution for crop monitoring (Amorós-López et al., 2013), as well as the spatio-temporal fusion of Sentinel-2 and Sentinel-3 satellites images to create daily pseudo-Sentinel-2 imagery to obtain the desired spatial and temporal resolution needed for local monitoring (Wang and Atkinson, 2018). Moreover, new data mining techniques have arisen from the need to better organise the initially fragmented information of the original high-dimensional data. This started with Machine Learning (ML) methods, such as Support Vector Machines (Wu et al., 2011), and the possibilities continued to expand through Deep Learning (DL) methods, characterised by Neural Networks (NN) which usually involved data with two or more layers (Southworth & Muir, 2021). However, such analyses with ML or DL methods

have typically addressed issues associated with the increased temporal resolution (and associated high dimensionality) but have not addressed the increased spectral resolution for classification or transformation of reflectance into variables of interest (selection/extraction). In fact, the vast majority of ML/DL approaches that manage the high dimensionality of data for crop monitoring using time series usually refer to spatio-temporal characteristics use a VI, rather than the actual spectra, to represent crop phenology dynamics (Zheng et al., 2016; Wu et al., 2018; Sisheber et al., 2022). This could be explained by the fact that sensors with higher temporal resolution acquire data with low-medium spatial resolution and often also lack spectral bands in essential wavelength domains (e.g. shortwave infrared). Therefore, multi-temporal information extraction methodologies associated with the dynamics of measured variables (e.g. reflectance) linked to changes between successive acquisitions, such as multi-temporal classification and trend analyses of data time series (forecasting/prediction), either with: i) the same sensor; ii) different sensors with similar properties and iii) different sensors with different properties, are still mainly part of the land cover monitoring domain at present (Bovolo and Bruzzone, 2015).

However, as the spectral dimension of modern satellites are expanding, the integration of multi-spectral features in multi-temporal analysis has the potential to provide new insights in crop monitoring for applications such as crop yield estimation, crop conditions assessment and crops stress detection. Therefore, there is a gap in information and the development of new techniques for these temporal dimensions in relation to crop monitoring applications in order to perceive and detect dynamic phenomena of crop function, evolution and change, as well as on the adaptation of models to these new data sources.

In order to integrate the use of these new mutli-temporal, multi-spectral data sources in agriculture, one of the interesting approaches to explore is the use of chemometric methods. These methods offer the possibility of being able to provide a quantitative description of agricultural measurements, while also having the capacity to reveal previously overlooked trends in high dimension datasets. Applying chemometrics to agricultural data allows the identification and description of the inter-relationship of environmental drivers, and their potential impact upon crops. The following part (1.2) is intended to justify and explain the methodological approach that has been used in this research. However, it should be noted that, as noted above, in terms of crop monitoring analysis through time series, the current dominant high-dimensional paradigms are related to spatio-temporal modelling and spatio-temporal decision-making (Ge et al., 2022). Therefore, the possibility of integrating the spatial dimension, usually *a posterior* to the main spectral-temporal analysis, is shown through geostatistical solutions in a punctual way, throughout the manuscript.

1.2 Multispectral time series coupled with chemometric methods for crop monitoring

1.2.1 Definition – generalities of chemometrics

Chemometrics, first introduced by Swedish scientist Svante Wold in 1971, is simply the application of mathematical and statistical techniques to derive more information from multivariate chemical data (Sarker and Nahar, 2015). It is generally applied for one or more of these three purposes:

- Explore the associative patterns in the data.
- Track the properties of elements on a continuous basis.
- Prepare and use multivariate classification models.

Although chemometrics is widely used in spectrometry and chromatography, it has the ability to analyse and model a wide variety of data types for an even more diverse set of applications, such as hyperspectral imagery (Amigo et al., 2013)

As described by Héberger (2008), the chemometric approach uses multivariate methods, i.e. all variables are considered simultaneously. In this way, the model is fitted to the data in such a way that the conclusions must be in harmony with the information present in the data. This is very different from a classical approach where the model is derived from theory and data is sought to prove the validity of the model. As a result, the classical approach determines new (causal) relationships and discovers new natural laws, while the chemometric approach usually finds a formal relationship, which has the elements of causality. The advantage of the chemometric approach is that correlations between variables can be considered as informative to identify new knowledge, but the disadvantage is that the resulting models do not necessarily have physical relevance (Héberger, 2008). This sometimes leads to difficulties in interpreting the output of the models. Although chemometrics seems to be completely outside the trend of crop monitoring methodologies, the methods used in chemometrics often also fall into the categories of data mining and machine learning. Chemometric methods are generally classified as either supervised or unsupervised (Zielinski et al., 2014). In supervised methods, both the input (X) and output (Y) are known, and the goal is therefore to determine the function that best approximates the relationship between X and Y. Typically, supervised learning is used to define a relationship based on data where both X and Y are known (test data). The resulting relationship can then be applied to new data X to determine the output Y, e.g. for classification problems when an object is assigned to a discrete category, or for regression problems when the input is assigned to a continuous output. Some of the most common supervised learning techniques are: regression methods (e.g. Partial Least Squares, PLS) and Discriminant Analysis

(e.g. Linear Discriminant Analysis, LDA) (Messai et al., 2016). In unsupervised learning, no prior knowledge of the data is required as the aim is to identify the underlying structure of the data without the user having to intervene. Unsupervised methods are typically used for exploratory data analysis and dimensionality reduction, i.e. to represent a dataset using fewer features than the original data. Among the most common unsupervised methods are clustering and principal component analysis (PCA) to analyse unlabelled datasets (Rácz et al., 2018).

In addition to the above techniques, there are a wide range of chemometric methods that can be applied to complex datasets (e.g. high dimensional) beyond traditional univariate approaches. The suitability of a method for analysing a given data set will depend on a wide range of factors and it is important to consider the limitations of a method when applying it (Héberger, 2008). For instance, highly accurate methods that are often computationally expensive (e.g. SVM) may not be suitable for a given analysis if a simpler model achieves similar accuracy. Furthermore, if the aim of the analysis is to understand the underlying relationships between variables, 'black box' methods (e.g. ANN) may not be as useful as those where relationships between variables can be interpreted, such as decision trees or multi-linear regression. It is important to note that the structure of the dataset to be analysed also determines the best methods of analysis; hyperspectral images typically contain tens to hundreds of contiguous narrow bands (Padoan et al., 2008). A hyperspectral image could be expressed as a hypercube in which each pixel contains dozens to hundreds of spectral bands, so that for any pixel in the hypercube, a full spectral reflectance curve is approximated. In several situations, it is necessary to extend these methods to account for the multi-way nature of the data. Thus, one can analyse data by methods such as N-way Partial Least Squares (supervised method) or PARAllel FACtor analysis (unsupervised method) that allow for the processing of high-dimensional data. An understanding of the strengths and limitations of each method, as well as the structure and properties of the dataset, is necessary to obtain the most accurate and relevant results from any chemometric analysis (Héberger, 2008).

1.2.2 Chemometrics for agricultural monitoring: beyond univariate analysis for multispectral time series data

Crops monitoring is complex and involves a multitude of processes arising from both natural and human-induced sources (Gremillion and Piperno, 2009). For example, the complexity of the response of the vegetation spectrum may be due to a complex mix of vegetation, soil brightness and environmental effects, among other factors (Bannari et al., 1995). Consequently, crop data obtained by RS are often multivariate due to the relationships between these multiple factors, which makes accurate monitoring difficult. To search for statistical relationships between biochemical characteristics of crop properties and vegetation reflectance, from the visible to the shortwave infrared region (not just a specific part, as is the

case with VIs), without being hampered by conventional univariate methodologies that may mask the underlying information within a high-dimensional space, an interesting option to consider is the use of multivariate chemometrics regression models.

The chemometric tools typically used in the high-dimensional domain to process hyperspectral imagery are based on 4 steps (Amigo et al., 2013):

- the reduction of data dimensions;
- the selection of the most significant spectral features;
- the extraction of spectral key features;
- the development of classification and prediction models.

The lack of knowledge of chemometric methods and their applications often prevents their dissemination in the scientific community and beyond (Héberger, 2008). However, in the context of multispectral time series, these methods may provide unique insights by the quantitative retrieval of the chemical composition of crop vegetation (Xue and Baofeng, 2017), as well as being able to adapt to the high-dimensional requirements of the spectral-temporal characteristics of the data. Therefore, these techniques have the potential to be applied in crop monitoring to mine crop knowledge through high-dimensionality analysis.

1.3 Research positioning

1.3.1 Research aims

1.3.1.1 General research problem

There are now technological developments in the field of remote sensing applied to agriculture in line with the development of platforms and sensors that characterise vegetation at multiple spectral, temporal and spatial resolution. As an example, the recent launch of optical remote sensing (RS) satellites, such as the Sentinel 2 constellation, guarantees there is a constant monitoring the Earth's surface at a high temporal resolution. Monitoring for agriculture can be considered at different spatial and temporal scales of decision by the farmer or the advisor, e.g. from a plot to a region, from one sole date to a complete cropping season. Among the avenues regularly evoked to deal with these possible different resolutions are new methodologies that provide multi-dimensional and efficient data retrieval capabilities. The main assumption of these methodologies is that when the multi-dimensional attributes of the RS database, such as the spectro-temporal responses, are lost, it is difficult to satisfy the needs of applications other than data visualisation (Zhu et al., 2021). Based on this assumption, it has been shown in the agricultural domain that the temporal dynamics of crop surface reflectance is a key factor in crop monitoring (Soudani et al., 2012). However, analysing the large amount of available

satellite data in terms of time series observations is often a difficult task, which leads to the formulation of the general research problem of the thesis:

Does time series information from multispectral remote sensing imagery offer new potentially valuable knowledge for agriculture?

1.3.1.2 Scientific objectives and questions

In view of the general question posed above, the application of more advanced methods than those currently proposed in the literature was considered necessary. In particular, because this hypothesis is predicated on the assumption that if the spectral and temporal dimensions are simultaneously taken into account in the analysis methods and, above all, if these dimensions are preserved in the results produced, the results themselves can be interpreted/analysed/correlated with respect to these two dimensions. Consequently, these new results will more completely take into account the data characteristics from the sensors currently available.

Following this reasoning, the specific scientific questions become:

- **How to better integrate and consider the temporal dimension, together with spectral and spatial dimensions, in the analysis of multispectral time series images for crop monitoring in agriculture?**
- **Does the simultaneous consideration of the spectral and temporal dimension in the analysis of time series provide a more relevant and detailed way of capturing crop changes?**

The decisions taken in advance of this thesis to address these issues have been:

Firstly, that chemometric methods can be used for the treatments and evaluation of RS data, for the extraction of useful information, and for decision-making in crop monitoring. These chemometric methods allow for the analysis of the high-dimensional, complex datasets obtained from RS data sources. The adaptation of these chemometric methods to the field of agriculture will allow the integration of spectral and temporal information to identify and describe the interrelationship of environmental and/or human factors on crop production and thus their potential impact on agriculture.

Secondly, the crop research undertaken will focus on viticulture systems on a regional scale by considering the situation of vineyards in Occitanie (South of France). The Occitanie region is known for having the largest area of vineyards in France (Filippi, 2012), with approximately 250 000 ha of vineyards in the target area (Languedoc-Roussillon) for this work, but these vineyards are located across a great diversity of landscapes and contrasting pedo-climatic conditions. The topography of this wine-growing region mixes

mountains and plateaus, foothills and intermediate plateaus and the coastal plain, which determines a multitude of different soil types for cultivation. The climate is Mediterranean, with Atlantic influences in the west and continental influences in the east. Such regional diversity of conditions results in a large number of different grape varieties and a large difference in management practices by the winegrowers across Occitanie. This regional diversity is what makes it an interesting case study.

Thirdly, taking into account the large spatial scale of the study, the Sentinel-2 satellites (A/B) were identified as the preferred RS data source because of their interesting and representative features in terms of temporal (global revisit frequency of 5 days), spectral (13 spectral bands from the visible to the shortwave infrared) and spatial (4 spectral bands at 10 m, 6 bands at 20 m and 3 bands at 60 m) resolution. The European Space Agency (ESA) also provides free and open data from the Sentinel-2A (launched in 2015) and Sentinel-2B (launched in 2017) satellites, which carry exactly the same multispectral instrument (MSI) on board and that were specifically designed to obtain spatially extended multispectral and multitemporal data to assist international agricultural development (Segarra et al., 2020). Quality spectral resolution is the main element for determining the physiological properties of plants, and the improvement of this dimension in Sentinel-2, compared to Landsat, is a leap towards a more accurate and reliable characterisation of crops. In situations where crops interact with any aspect of their environment (seasonal climate variations, extreme weather events, soil properties, etc.) or as crops grow and go through different phenological stages, interactions between plants and light reflectance will lead to changes in plant signalling patterns that can be monitored and interpreted using these Sentinel-2 satellite time series data (Segarra et al., 2020).

1.3.2 Organisation of the manuscript

After this first chapter, which presents the general context and the scientific question of the thesis, the manuscript is organised in five distinct chapters:

- The second chapter (Chapter 2) is devoted to a discussion on the limits of classical approaches to spectral analysis (i.e. based on vegetation indices) to be able to fully exploit the potential of time series for agricultural monitoring. To this end, an article (Article 1) is presented detailing the potential of time series for assessing water stress at the regional scale, without neglecting the limitations of using VIs.
- The third chapter (Chapter 3) is dedicated to the exploration of multispectral time series images at a regional scale for the extraction of agronomic knowledge by taking into account simultaneously the spectral and temporal dimensionality of the data. This chapter is divided into two parts presenting two different unsupervised approaches. The first part (3.1) highlights how exploratory studies can be enriched by the inclusion of the temporal dimension for an improved understanding of crop

monitoring (Article 2), while the second part (3.2) presents a second, alternative, unsupervised, exploratory approach (compared to 3.1) with a high potential to fit the temporal characteristics of the RS dataset.

- The fourth chapter (Chapter 4) focuses on the implementation of a supervised approach to predict the effect of a disruptive weather event (heatwave) on individual vineyards at the regional scale. This chapter is again divided into two parts. The first part (4.1) presents a scientific article (Article 3) that shows the potential of a multi-way supervised approach to remote sensing time series management. The second part (4.2), contains an article (Article 4) that assess the impact of an extreme weather episode with spectro-temporal and spatial information from the multi-way approach presented in the previous part (4.1).
- The fifth chapter (Chapter 5) is based on the last article (Article 5) published within this thesis and is dedicated to the selection of multidimensional variables to identify the most discriminating spectral domains, as well as the most discriminating time periods, to characterise the effect of heat wave event on vineyards in the Occitanie region.
- The sixth and final chapter (Chapter 6) consists of a summary of the conclusions of the various outcomes and issues of the chapters in the thesis and provides a perspective for future and on-going work in this area of multi-temporal, multi-spectral analysis of RS data for agriculture.

This manuscript is therefore structured in 6 chapters, 4 of which are in the form of articles (Chapter 2, 3, 4, and 5). The articles in Chapters 2, 4, 5 have already been published, while the article in the first part of Chapter 3 is currently under review. Each of these chapters has an introduction and a conclusion section to set the scene for the published articles and to provide a positioning for the articles within the overall scientific framework of the thesis.

Chapter 2. Major limitations of relying exclusively on the spectral dimension of multispectral time series data for crop monitoring

Introduction to Chapter 2

Remote sensing from space satellites plays a key role in understanding the dynamics of the vegetated land surface. Satellite remote sensing of vegetation is mainly performed by obtaining the electromagnetic reflectance information at different wavebands from the vegetation canopy, in the knowledge that the reflectance of light spectra from plants is affected by various distinct elements, such as chlorophyll levels in the vegetation, tissue water content, phenological stage, as well as plant species and other intrinsic factors (Chang et al., 2016). These different wavebands can then be combined in various ways to construct Vegetative Indices (VIs), with the type and number of indices possible dependent on the wavebands collected by the sensor. Such an approach has been the background of investigations in vegetation remote sensing research over the past four decades (Xue & Baofeng, 2017). Much of this research has attempted to formulate specific VIs that can be related to percent vegetation cover, biomass, leaf area index (LAI) and the fraction of photosynthetically active radiation absorbed by the canopy (Silleos et al., 2006). Therefore, spectral indices of vegetation have been proposed, designed and used as indicators of temporal and spatial variations in vegetation structure, biophysical properties and vegetation density (Xue & Baofeng, 2017), thereby presenting themselves as a means to study and monitor vegetation and its dynamics (Silleos et al., 2006). Although a large number of indices have been proposed over time[1], one of the most commonly used and widely implemented index calculated from multispectral information is the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973). It is calculated as the normalized ratio between the red (600 – 700 nm) and near infrared (700 – 1200 nm) bands (Rouse et al., 1973). A direct use of NDVI is to characterise canopy size (growth or vigour), enabling the assessment and the monitoring of changes in canopy biophysical properties (Junges et al., 2017). Considerable success has been achieved with the use of VIs, such as NDVI, as they are simple and effective tools that have been designed to find a functional relationship between crop characteristics and remote spatial observation (Basso et al., 2004). Therefore, it is understandable that vegetation indices have been recurrently used in agriculture to qualitatively and quantitatively assess the vegetation cover of different crops, such as wheat (Aparicio et al., 2000), corn and soybeans (Chen et al., 2005) olives (Solano et al., 2019) and grapevines (Brunori et al., 2020) among others. One of the key factors to consider for a successful application of remotely-sensed imagery to cropping systems is the final choice of the specific VI to use. Different VIs have different sensitivities to various factors, hence the choice of VIs needs to be made with caution by comprehensively considering and analysing the advantages and limitations of each existing VI that could be applied in a specific environment (Xue & Baofeng, 2017).

1. https://www.indexdatabase.de/

Given this context, it is natural that time series analysis has focused on the analysis of the evolution of VIs over time (Hatfield et al., 2019; Li et al., 2021). In the light of this, the objective of this section is to implement a classical approach based on the analysis of a vegetation index over time in viticulture and on a regional scale, with the purpose of evaluating its potential uses but also its limitations.

This chapter examines the potential of temporal series of remotely sensed data for the assessment of vine water restriction using a conventional approach such as VIs, particularly the NDVI. Due to the temporal and spatial variability of the vine water status, the classical sampling procedures based on reference methods currently used are not suited for large scale (regional) monitoring (Ojeda et al., 2002). To overcome the aforementioned limitations, the assessment of grapevine water status with NDVI time series during the vegetative period is proposed. Although Laroche-Pinel et al. (2021b) recently demonstrated that other indices such as Red-Edge Position (REP), Normalized Difference Red-Edge (NDRE2), Red-Edge Chlorophyll Absorption Index (RECAI), Normalized difference Infrared Index (NDII) and Moisture Stress Index (MSI) are more effective (than NDVI) in monitoring water status specifically for vineyards. Other studies have highlighted the potential of NDVI to characterise the spatial variability of vine water status in non-irrigated conditions (Acevedo-Opazo et al., 2007).

In the conditions of Southern French wine-growing region, water availability is the main factor, which limits plant growth. Therefore vigour is considered as strongly related to soil water availability, which implies that NDVI can provide relevant information to zone the vineyard according to water restriction (Acevedo-Opazo et al., 2007). Moreover, since water stress conditions are very different depending on the time of occurrence, NDVI time series may provide an added value to properly analyse the relevance of the temporal variability of the phenomenon. It is worth noting that most studies, such as Acevedo-Opazo et al. 2008, demonstrate the usefulness of NDVI for characterising vigour differences exclusively at field level. The field level corresponds to a production unit that has the same characteristics such as age, variety and rootstock, and is usually managed uniformly in terms of weeding, fertilization, pruning, irrigation, etc. (Santesteban et al., 2013). However, as scale resolution considered increases, new sources of variability are likely to emerge and affect NDVI values, making this approach more difficult at larger scales.

Moreover, as Lawrence and Ripple (1998) stated at the time, several problems have continued to plague remote sensing scientists concerning NDVI. Three problems that continue to be present and are also relevant to the article that follows in this second chapter are that (i) for sparse canopy, vegetation reflectance signals become increasingly contaminated by soil reflectance signature; (ii) soil heterogeneity can therefore influence spectral responses; (iii)

atmospheric conditions and sun-sensor-surface can impact Satellite-derived NDVI time series (Maynard et al., 2007; Hird & Mcdermid, 2009). Notwithstanding these limitations that can be applicable to a large part of VIs, the NDVI and its functional equivalents continue to be widely applied in remote sensing studies of vegetation attributes (Yengoh et al., 2015).

In this following chapter, there is no intention to evaluate the usefulness of the different VIs, but the basic limitations of such approaches linked to the visualisation and compression of data for dissemination and analysis using the most commonly used vegetation index as a reference; it is a first step to understanding the current restrictions (spatial and temporal) of such approaches and why there is a demand for a change of paradigm in vegetation monitoring using time series.

Article 1: Potential of temporal series of Sentinel-2 images to define zones of vine water restriction

Published in the Special issue of *Precision Agriculture*: In *Precision agriculture'21* (pp. 77-88). Wageningen Academic Publishers, 202. DOI: 10.3920/978-90-8686-916-9_65

E. Fornieles-Lopez[1], G. Brunel[1], N. Devaux[2], F. Rancon[1], L. Pichon[1] and B. Tisseyre[1]

[1]ITAP, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France,

[2]LISAH, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France

Abstract: This paper investigates the potential of time series of Sentinel-2 satellite images and derived NDVI index to characterise vine water status. It assumes that in non-irrigated conditions, growth is the main factor affected by water restriction. Therefore, NDVI time series may be relevant to identify sites presenting differences in vine water restriction. The study was carried out on 7 plots and 2 reference plots across the Mediterranean rim. It shows that raw NDVI temporal series present some limitations. Therefore, proposes an original methodology based on Empirical Cumulative Distribution Function of NDVI values observed over the whole season to account for the whole temporal dynamics of NDVI. Results showed that an approach based on cumulative distribution of NDVI values is relevant to classify plots in relation to their water restriction.

Keywords: water restriction; time series Sentinel-2 images; vine growth; ECDF

2.1 Introduction

Many authors have shown that grapevine water status has a direct effect on grape quality through its influence on vegetative and fruit growth (Dry & Loveys, 1998; Ojeda et al., 2002). Viticulture in the southern region of France is mainly carried out using non-irrigated practices, which adds to the challenges faced by wine producers when trying to maintain the high quality requirements in wine production and highlights the necessity to develop tools and methods to better characterise spatial and temporal changes in vine water status (Ojeda et al., 2005). Different approaches and tools have been proposed in the literature to enable the water status of the vine to be characterised and monitored (Rienth & Scholasch, 2019). Among these solutions, remote sensing is an interesting sensor because the resolution and the spatial footprint of the data collected is likely to meet the challenges of monitoring and characterising the vine water status at different spatial scales, from the field to the region scale. Until now, remote sensing has mainly been dedicated to defining zones of vine water status in order, for example, to consider differential harvest of different potential quality grapes (Acevedo-Opazo et al., 2008). At larger scales (territory), remote sensing has also been proposed as a decision support tool to highlight large zones of water availability (Montero et al., 1999).

In the south of France, and to a lesser extent in other Mediterranean vineyards, it has been shown that estimating biomass through a conventional index such as the NDVI was a particularly simple and relevant approach to characterise zones of vine water status (Acevedo-Opazo et al., 2008, Montero et al., 1999). Indeed, as the vineyard is mostly non-irrigated, access to water is the main factor affecting vine vegetative growth (Pellegrino et al., 2005). As a result, one NDVI map at the end of the growth cycle (at veraison) makes it possible to highlight zones of differentiated growth corresponding to zones with differentiated water restriction. Moreover, recent work has shown the relevance of time series of sentinel-2 images for characterising vine growth dynamics (Devaux et al., 2019). This new source of observation is therefore an opportunity to investigate the interest of NDVI in identifying zones of different water status in viticulture by considering the whole dynamics of vine growth. Such an approach could make it possible to characterise in greater detail how vine growth dynamic is affected by the occurrence of a water stress. It could therefore be a relevant tool for highlighting zones whose growth is likely to be affected by access to water. To our knowledge such an approach was never tested in the literature.

The objectives of this work are therefore to investigate the potential of time series of free Sentinel-2 satellite images to define vine water status zones. As a first approach, it will investigate the potential and the limitations of raw NDVI time series and then propose an original methodology to account for the whole temporal dynamics of NDVI while limiting drawbacks of raw data.

2.2 Materials and Methods

2.2.1 Description of the study area

The study was carried out in 7 plots and 2 reference plots across the Mediterranean rim in northern Spain and southern France (Figure 2.1). All these plots are commercial productive vineyards with characteristic representative of Mediterranean vineyards in terms of density and management practices. The vineyard plots' areas were ranging from 0.1 to 1 ha. They were trained in a vertical shoot positioning system with 2.5 m inter-row. All plots were non-irrigated. The climate was Mediterranean with hot and dry summer leading to high evaporative demand and to significant water restrictions during summer (Fernandes-Silva et al., 2019).

The reference plots Ref1 and Ref2 were chosen for their extreme pedo-climatic conditions leading to high and low water restrictions respectively. The 7 study plots were chosen because they were representative of the range of pedo-climatic conditions for vineyards in a Mediterranean context.

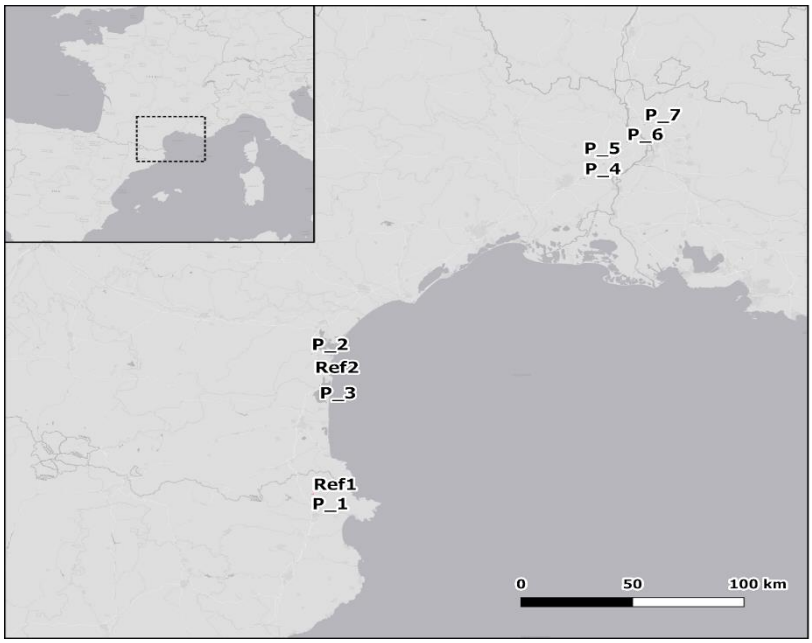


FIGURE 2. 1 - Location of the study area for the 7 vineyards of interest and the 2 reference plots Ref1 and Ref2, with respectively high and low water restriction.

2.2.2 Acquisition of images

Images were selected via Google Earth Engine (GEE) web portal providing Sentinel- 2 L2A (Sentinel-2A and Sentinel-2B). As suggested by Devaux et al. (2019), time series with dates ranging from May 16th to September 3rd were considered for being the most relevant period to monitor vine growth vegetation in the considered region. Only images containing the study plots were selected, and to avoid mixed pixels with no relevant information, a 10 m buffer was designed over the border of each plot. As the revisit period of the Sentinel-2 (A/B) satellite is 5 days, 30 images should have been potentially available over each plot for the chosen time period. However, following the methodology proposed by Hollstein et al. (2016) images containing clouds or shadows were removed from the time series. Table 2.1 presents the resulting available images for each plot showing that filtering out images possibly affected by climatic conditions, leads to more appropriate information but also to a great heterogeneity in terms of images availability for each plot. The number of images availability over the growing season ranged from 9 to 27 depending on the considered plot. This result was considered.

Red band (640 nm - 690 nm) and NIR band (780 nm - 910 nm) were used to compute the Normalised Difference Vegetation Index (NDVI) (Rouse et al., 1973) at 10m resolution.

TABLE 2. 1 - Number of images available to compute the NDVI temporal spectrum, area and the spatial coordinates of each of the 7 vine plots of interest and the 2 reference plots Ref1 and Ref2.

ID Plot	Images Sentinel-2	Area (ha)	Coordinates (WGS84)	
			Latitude (degree)	Longitude (degree)
Ref1	23	0.56	42.365	2.952
Ref2	9	0.12	42.996	2.958
Plot_1	23	0.62	42.364	2.947
Plot_2	9	0.35	42.998	2.965
Plot_3	9	0.35	42.997	2.960
Plot_4	27	0.32	43.859	4.466
Plot_5	27	0.30	43.955	4.643
Plot_6	26	0.33	44.032	4.706
Plot_7	27	0.43	44.107	4.817

2.2.3 Ground truth observations

Water restriction was estimated by measuring the vine shoot growth as proposed by Martinez de Toda et al. (2010). According to Lovelle et al. (2009), vine shoot growth was measured by observing the apexes of 50 branches spread over 10 different vines and classified into 3 categories: full growth apex, moderate growth apex or stopped growth apex. These 50 observations were summarised by an index of Growing Apex (iG-Apex) ranging from 0 to 1. An iG-Apex equal to 1 was corresponding to a high vine shoot growth with all apexes growing. An iG-Apex equal to 0 was corresponding to a field where all apexes have stopped growing, which indicates an appearance of water stress. Values ranging from 0 to 1 thus represented a gradation between full growth and complete cessation of growth.

Observations were collected using the Apex-Vigne application (Brunel et al., 2019) by the manager of each vineyard. Between 6 and 8 observations were made on each plot during the season. The observations' dates were adapted to local operational constraints.

2.2.4 Data analysis

Hird and Mcdermid (2009) have demonstrated the impact of varying atmospheric conditions and sun-sensor-surface viewing geometries on Satellite-derived NDVI time series. In order to partially avoid these effects, this study proposes to study plots by calculating the distribution of all NDVI values observed during the season. As a first approach, this study proposes to use the Empirical Cumulative Distribution Function (ECDF) (Dowd, 2020) of the NDVI values (Equation 2.1).

$$\mathrm{ECDF}(x) = \mathrm{P}\,(X \leq x) \qquad \text{(EQ. 2.1)}$$

ECDF stands for Empirical Cumulative Distribution Function, x corresponds to an observed NDVI value and X is a sample of normally distributed random variables. For a given NDVI value, ECDF is the frequency of observed NDVI values falling below that value. It is thus a step function jumping at the end of each data point, converging with the frequency 1. This transformation was chosen for its ability to account for the dispersion, shift and tails of empirical datasets that can't be modelled with a classical statistical model. Moreover, the "standardisation" made it possible to better consider field temporal data comparison.

In a first step, in order to verify the ability of the method to arrange the plots according to their level of water restriction, a comparison between the observed ECDFs with those of reference plots was carried out. The Wasserstein statistic (Dowd, 2020) was used to assess how similar an observed ECDF was from the ones of the reference plots by a pairwise comparison. The interest of this statistic was to consider the whole curves to perform this comparison. The assumptions of the statistic followed: H_0 if the data follows a specified distribution (one of the

reference plots) or H_1 if the data do not follow a specified distribution. The null hypothesis was rejected for *p*-values of 0.1 ($p < 0.1$). In a second step, ground truth data (iG-Apex) were used to validate the classification as an indicator of the level of water restriction experienced by plant vines.

2.3 Results and Discussion

Figure 2.2 show changes in NDVI values observed for the nine fields under study, over the vine vegetative season. NDVI values of both reference fields, Ref1 and Ref2, that respectively experienced a high and a low water restriction are plotted as a thick solid line in red and blue, respectively. Focusing first on reference fields, Ref2 shows expected evolution of vine canopy change in NDVI as described by Devaux et al. (2019). The dynamics of NDVI during the growing season is divided into two phases with first an increase in NDVI values corresponding to the growth of vegetation (April-June) and second, a plateau over July-September corresponding to the growth stop and fruits maturation. Regarding the Ref1 field, this expected trend is not observed. NDVI values of Ref1 field remained very low over the season highlighting a very low growth of the vegetation. This result confirms the relevance of NDVI temporal series to assess how vine growth is affected by water restriction in non-irrigated conditions. Regarding the NDVI temporal series of other fields, Figure 2.2 also highlights the difficulty to directly use this information to assess how water restriction impacts vine growth.

Two plots (Plot_1 and Plot_5) have NDVI temporal dynamics similar to Ref1 (Figure 2.2). The other plots (Plot_2, Plot_3, Plot_4, Plot_6 and Plot_7 to a lesser extent) present dynamics similar to Ref2 (Figure 2.2). However, the temporal evolution of NDVI remains locally disturbed, probably due to cultivation operations such as weeding at the beginning of the season or to conventional canopy management operations that cause sudden variations in NDVI values. Note also that curves may be affected by missing images due to local climatic conditions during acquisition or even climatic conditions when images are not filtered out properly by the cloud detection algorithm.

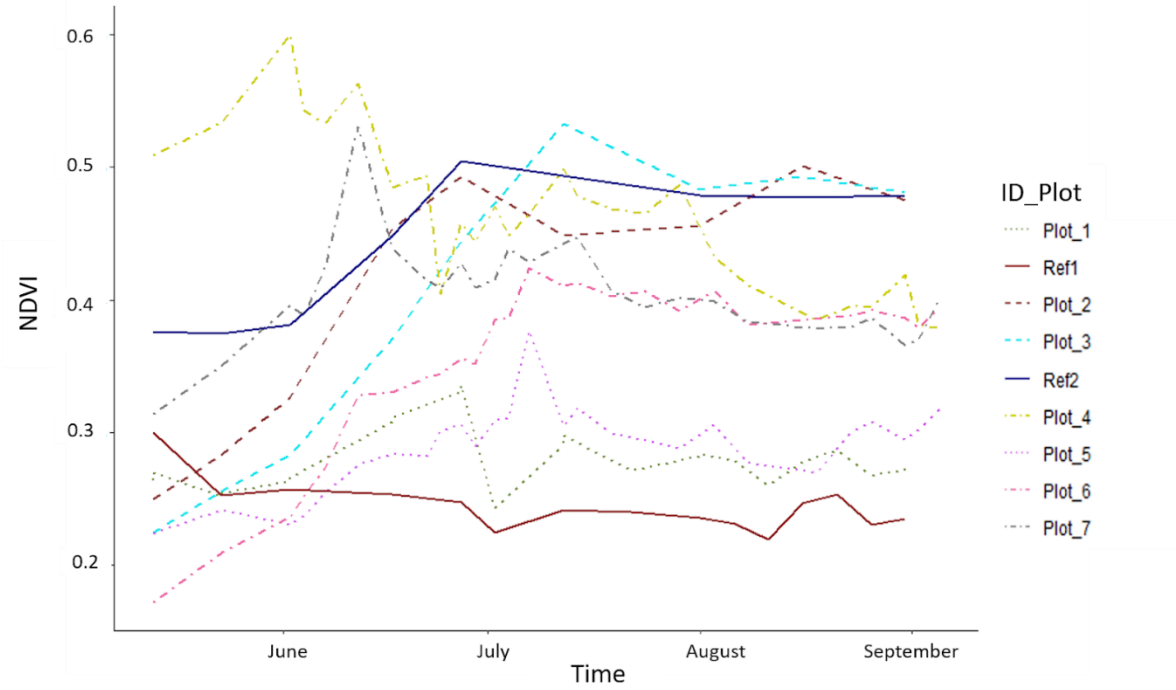


FIGURE 2. 2 - NDVI temporal dynamics for the nine fields under study over the vegetative season (from May 16th to September 3rd) for the 2019 season.

Figure 2.2 shows the potential interest of NDVI time series to identify plots with different growth dynamics possibly to differences in water restriction experienced by plots over the season. However, it also shows the difficulty to define an objective index to classify plots according to a level of water restriction. These results highlight the necessity to propose a more comprehensive data transformation to simultaneously account for the whole NDVI time series, noise reduction and standardisation. As a first approach, this study proposes to investigate the potentiality of a cumulative function (ECDF) to answer these issues.

When analysing the NDVI time series with ECDF curves (Figure 2.3), the range of NDVI values over the season offers the explanation of the particular growth conditions of each plot. The slope of each plot gives information on the variance of NDVI during the season, which provides information on vine growth dynamics. On the one side, the Ref2 curve has a steep slope starting at 0.2 and ending at 0.3. This stagnation of NDVI can be related to the appearance of a significant water restriction since 100 % of NDVI values do not exceed 0.3 for this field. On the other side, the curve of the Ref1 has a gentle slope indicating high variance in NDVI values. This high variance is understood as a good development of the vegetation which is not constrained by water restriction. As already seen in Figure 2.2, the NDVI dynamics of Ref1 during the growing season is divided into two phases: one of growth and one of standstill.

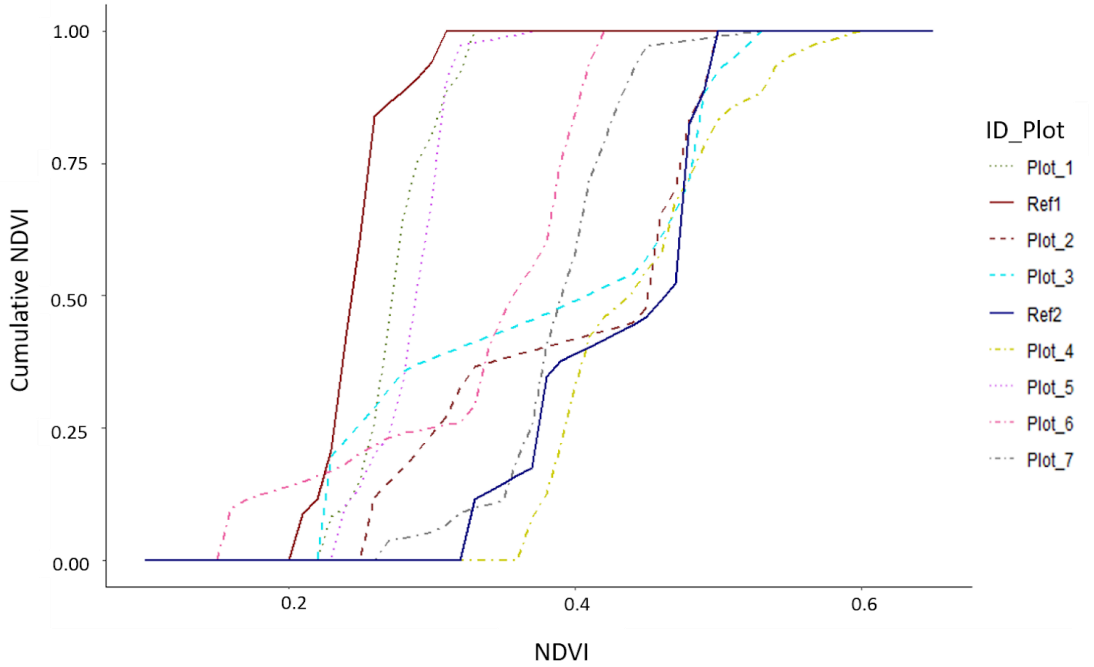


FIGURE 2. 3 - ECDF curves derived from NDVI time series for the nine plots under study from May 16th to September 3rd for the 2019 season.

Like Ref1 plot, Plot_1 and Plot_5 present a very strong slope as Ref1. Although some variability appears for NDVI values lower than 0.3, the rest of the plots present gentle slopes similar to that of Ref2 plot. Figure 2.3 shows that ECDF better summarises the global behaviour of plots over the growing season and f permits a better distinction between the dynamics observed on the different plots. After having defined two reference plots (Ref1 and Ref2), it allows to visually define two groups of plots that can be distinguished by their water status.

To validate the possibility to classify the plot according to ECDF, the Wassertein Statistic has been used to analyse whether plot behaviour was more similar to a water stressed plot (Ref1) or to a low water restriction plot (Ref2). Table 2.2 presents the results of the analysis.

TABLE 2. 2 -Values of Wassertein Statistic Distance between empirical cumulative distribution function of NDVI time series of plots and plot of reference Ref1 and Ref2 corresponding respectively to a high water restriction and low restriction, and ground truth reference – iG-Apex observed on 30th July). Significance threshold, *p*-values <0.1*.

Wasserstein Statistic Classification between ECDF of NDVI				**Reference iG-Apex**
ID_Plot	Ref1	Ref2	Vine water restriction	Value at 2019-07-30
Plot_1	**0.046**	0.250*	High	0.11
Plot_5	**0.063**	0.230*	High	0.26
Plot_6	0.160*	0.120*	Not Classified	0.33
Plot_7	0.230*	**0.060**	Low	0.42
Plot_4	0.320*	**0.040**	Low	0.45
Plot_3	0.208*	**0.098**	Low	0.55
Plot_2	0.240*	**0.052**	Low	0.86

According to the Wasserstein test, Plot_1 and Plot_5 are closer to the Ref1 plot (Table 2.1.2). This can be interpreted by the fact that these plots have experienced strong water restriction during the season. Plot_2, Plot_3, Plot_4 and Plot_7 are closer to the Ref2 reference plot (Table 2.2) meaning that they may have experienced low water restriction. Plot_6 does not seem to be closer to Ref1 neither to Ref2 with *p*-value < 0.1. This may be interpreted by the fact that this plot has experienced an intermediate situation regarding water restriction.

The ground truth iG-Apex observed on 30 July 2019 is consistent with these results. Plot_1 and Plot_5 have lower iG-Apex than the other plots and Plot_8 presents intermediate values. In a Mediterranean context, low values of vegetative growth at the end of July are related to high levels of water restriction. These results show that the proposed approach is a relevant tool to classify plots regarding their experienced water restriction during the season.

NDVI ECDF shows to be a relevant approach to differentiate vine plot populations in terms of experienced water restriction account for the whole NDVI time series. This evidence is confirmed by ground truth iG-Apex. Transformation NDVI time series into NDVI ECDF an interesting approach to account for plot vine growth in relation to water. Note that this approach could also be applied at the within field level considering the limitation of sentinel-2 image resolution (10 m size pixels). By increasing the density of iG-Apex measurements, it will be possible to retrieve in-situ data more representative of the diversity of viticulture conditions in the Mediterranean basin. The coupling of this iG-Apex index data with the possibility of

generating temporal profiles of NDVI offers the opportunity to characterise growth dynamics to better anticipate water needs in viticulture.

2.4 Conclusion

This paper highlighted the potential of NDVI time series of Sentinel-2 for identifying zones with different water restrictions dynamics. These results showed that an approach based on empirical cumulative distribution function of NDVI values is relevant to classify plots in relation to their water restriction. These results open up further investigation, especially regarding the identification of which part of the spectrum is the most discriminating in the classification process in order to determine how many images are needed to identify the water constraint trends. This method would permit the development of a decision support tool allowing identification of zones with different water restrictions dynamics to refine strategies for monitoring water status providing inputs to early warning systems.

Conclusion of Chapter 2

The objective of this second chapter was to verify the potential of multispectral satellite imagery while presenting at the same time its limitations in terms of temporal analysis and interpretation for crop monitoring through vegetation indices, such as NDVI. The latter considerations demonstrated that the quantitative and qualitative interpretation of remote sensing information from VIs is a complex task. Article 1 intended to test an original but simple approach with NDVI time series from Sentinel-2 satellites to first identify the shortcomings and then overcome the difficulties of interpreting the raw data of a case study linked to the dynamics of vine water status. Although using NDVI data for water restrictions dynamics as shown in Article 1 might be possible and, in fact, there are already fully operational systems for other crops for drought monitoring using NDVI (Yengoh et al., 2015); Maynard et al. (2007) defined three major issues when only using one sole information source:

(i) when using a single or limited group of bands to compute the indices (usually in the red and near infrared domains) there is no consideration of whether other bands might add and improve the interpretation;
(ii) it restricts the ability to model the effects of different interactions between vegetation and energy in different portions of the spectrum;
(iii) the spectral responses may potentially be strongly affected by soil heterogeneity.

These limitations are especially evident when it comes to applying different VIs to discontinuous canopy structures, such as grapevines, over time. In such situations, canopy closure is never achieved and imagery always contains a combination of spectral information from the target plants that is mixed with the spectral information from soil, weeds and cover crops in the inter-row. The presence of mixed pixels makes discrimination of regions of interest and the extraction of a simple VI challenging when operating at the regional scale (Xue & Baofeng, 2017). These issues have significant implications for identifying changes and monitoring the dynamics of vegetated surfaces. Since the early days of satellite remote sensing, researchers have sought to use multispectral imagery and derived VIs to measure assorted variables related to vegetative biomass (Lawrence & Ripple, 1998). On the basis that a change from the VI approach is necessary to correctly assimilate all the spectral information that multispectral imagery offers, an approach using non-indexed spectral data should be considered as an interesting alternative to using VIs and to avoid the issues defined above. It is important to contextualise the rapid change in the amount and variety of spectral data that have been become available through recent evolution in remote sensing systems. In addition to providing higher resolution in the spectral dimension, new satellites offer increasingly better temporal and spatial resolution information within the imagery. As data structures change, it is important to also adapt and change of methodological approaches to extract and process all relevant

information that may be present within the large data flow originating from modern earth observation satellite systems. Throughout this manuscript, the aim is to highlight the potential of more holistic approaches to processing multispectral image time series following the hypothesis that crops differ in their phenology. The analysis of temporal variations, in addition to spectral variations, will provide a more complete analysis of the information within the imagery, compared to a classical VI approach, and will result in improved knowledge of crop performance.

Chapter 3. Potential of non-supervised multi-way methods for regional crop data exploration from multispectral time series data

Introduction of Chapter 3

Chapter 2 has shown a case study representative of the relevance of considering the temporal dimension of satellite images to understand a dynamic phenomenon such as the occurrence of water restriction in vine fields at the regional scale. By showing some limits of common approaches that only account for changes in vegetation indexes (VIs) over time, it highlighted the necessity to use more complete approaches aiming to consider both spectral and temporal dimensions simultaneously to account for all the potential information provided by time series of multispectral images. When considering only changes in VIs over the time, the knowledge that remote sensing images can provide is thus diminished from its full potential. The temporal ordering of remotely sensed data to provide another dimension to the data space creates not only a challenge in terms of mining and analysis, but also in terms of structuring these high-dimensional data sets (Chi et al., 2016). In the following, it was hypothesised that organising data from multispectral instruments (MSI) as multidimensional data cubes was a relevant proposal to extract the best potential of time series of multispectral images. This proposal follows domain-based examples in the field of Earth Observation (EO) (Ferreira et al., 2020) or in the field of hyperspectral data processing (Shaw and Manolakis, 2002). This original way of structuring data allows the temporal dimension to be considered as an inherent factor in the analysis (Ferreira et al., 2020). The relevance of including temporal analysis from crop monitoring can be approached in several ways, e.g. depending on the information available to achieve this objective. To this end, two different strategies, unsupervised and supervised are respectively proposed in Chapter 3 and Chapter 4 of this manuscript.

The main objective of the Chapter 3 is to explore the potential of multispectral time series imagery for crop monitoring at a regional scale. This means dealing with the simultaneous consideration of the temporal and spectral dimensions in data processing. At this spatial scale, different climatic and soil conditions may lead to different crop management practices or differences in crop canopy responses. Therefore, the variability at this spatial scale may reveal very useful information for crop monitoring. However, although the study of the regional scale is interesting because of its diversity of information, it is very difficult to have a complete ground truth data set representative of all the factors affecting and/or explaining observed heterogeneity. This scarcity of available samples (in small numbers) is often not sufficient to allow a sensible and adequate learning from ground data. In this context, unsupervised methods were considered as a first attempt to data exploration as they only involve the intrinsic properties of the time series collected in the multispectral images. Therefore, taking into account the variability observed at the regional based on RS information captured simultaneously in the temporal and spectral dimensions, this Chapter 3 proposes a new approach to deal with the large volumes of information we are facing.

Concerning this Chapter 3, the exploration of temporal analysis to denote phenological variability induced by climatic (e.g. seasonality, weather variability, extreme weather events) or non-climatic (e.g. human cultivation and soil variability) factors is addressed with two of the most representative unsupervised methods in the field of chemometrics. It is approached with two groups of methods, depending on whether they assume a trilinear structure in the data set or not (de Juan and Tauler, 2001). Assuming a trilinear structure requires that the matrices composing the data set can be arranged as a cube, i.e. the matrices forming the cube must have the same size and their rows and columns must represent the same type (nature) of variable, which provides a powerful and robust model (Blanchet, 2008). Alternatively, in order not to restrict the model to the natural structure of the data, it is possible to use data that share certain commonalities, where the resulting data matrices can be assembled in different ways, e.g. in so-called multi-set structures. In other words, multiset analysis does not require trilinearity, but only bilinearity of the data which implies that these different matrix arrangements are more flexible than the construction of a cube because they allow the simultaneous analysis of matrices that do not have the same structure (Blanchet, 2008).

Two different three-way resolution methods (one from each family mentioned) are applied to address the possibility of showing time-sensitive approaches by considering different sequential relationships of multi-temporal observation. Hence, Part 3.1 presents the PARAllel FACtor analysis (PARAFAC) that handles a folded cube, i.e. it assumes a trilinear structure in the data by imposing a restriction to the temporal dimension. It should be noted that the spatial dimension is not taken into account in this analysis but that a geostatistical analysis of the data is produced for validation purposes. Part 3.2 presents the Multivariate Curve-Resolution Alternated Least Squares method (MCR-ALS) to propose an alternative time perspective to the above by non-assuming a trilinear structure in the data set.

To explore the potential of time series for crop monitoring at regional level, a case study focusing on Sentinel-2 satellite time series and a grapevine crop is presented. Given its characteristics with respect to its: (i) high temporal resolution of the satellite images (5 days), (ii) medium spectral resolution (13 spectral bands) and medium spatial resolution (10 m, 20 m and 60 m), it allows an almost continuous monitoring of crops. As for grapevine cultivation, this is an interesting case as the vegetation cover of the grapevine is not continuous (plants are mainly organised in rows and soil is more or less visible depending on the phenological stages) and requires cultivation operations (e.g. chemical or mechanical weeding or canopy thinning). This soil-vegetation interaction, peculiar to vine cultivation, is very complex, as it includes a more or less important edaphic component depending on the phenological stage and on weeding and canopy management. This complex interaction starts at small spatial scales (plot level) and evolves as the spatial scale increases. The interesting complexity and variability inherent to vineyards in the Occitanie region was seen as a challenge but also as an opportunity to access new knowledge on crop monitoring through unsupervised approaches since the above

characteristics are likely to provide intelligible evidence linking environmental factors and cropping operations to RS data.

3.1 Application of Parallel Factor Analysis (PARAFAC) to the regional characterisation of vineyard blocks using remote sensing time series

Article 2: Application of Parallel Factor Analysis (PARAFAC) to the regional characterisation of vineyard blocks using remote sensing time series

Submitted to the "Use of Satellite Imagery in Agriculture" special issue in *Agronomy.*

E. Fornieles-Lopez[1], G. Brunel[1], N. Devaux[2], J.M. Roger[1], J. Taylor[1] and B. Tisseyre[1]
[1]ITAP, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France,
[2]LISAH, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France

Abstract: Monitoring wine-growing regions and maximising the value of production based on their region/local specificities requires accurate spatial and temporal monitoring. The increasing amount and variability of information from remote sensing data is a potential tool to assess this challenge for the grape and wine industry. This article provides a first insight into the capacity of a multiway analysis method applied to Sentinel-2 time-series to assess the value of simultaneously considering spectral and temporal information to highlight site specific canopy evolution in relation to environmental factors and management practices which present a large diversity at this regional scale. PARAllel FACtor analysis (PARAFAC) was used as an unsupervised technique to recover pure spectra and temporal signatures from multi-way spectral imagery of vineyards in the Languedoc-Roussillon region in the south of France. The model was developed using a time series of Sentinel-2 satellite imagery collected over 4978 vineyard blocks between May 2019 and August 2020. From the Sentinel-2 (spectral and temporal) signal, the PARAFAC analysis allows the identification of spectral and temporal profiles in the form of pure components, which corresponded to vegetation and soil. The PARAFAC analysis also identified that two of the pure spectra are strongly related to characteristics and dynamics of vineyard cultivation at a regional scale. A conceptual framework was proposed in order to simultaneous consider both vegetation and soil profile and to summarise the mass of data accordingly. This methodology allowed to compute a concentration index that characterises how close is a field to a vegetation or a soil profile over the season. The concentration indices were validated for the vegetation and the soil over two growing seasons (2019 and 2020) first with a geostatistical analysis. Indeed, a non-random distribution of concentration index at the regional scale was assumed to highlight strongly spatially organised phenomenon related to spatially organised factors of environment (soil,

climate, training system, etc.). In a second step, spatial patterns of indices were subjected to the expertise of a panel of advisors of the wine industry in order to validate them in relation to vine growing conditions. Results show that the introduction of PARAFAC method opens up the possibility to identify spectro-temporal profiles of vineyard blocks relevant for understanding and characterising them on a regional scale.

Keywords: time-series; multispectral imagery; remote sensing; folded methods, unsupervised methods; expertise gathering

3.1.1 Introduction

The ability to use Remote Sensing (RS) time series information for landscape monitoring allows the study of large-scale evolutionary phenomena to be considered (Griffiths and Lee, 2000). It opens up the possibility to improve the understanding of dynamic processes, such as changes in forests, agricultural crops, land use, etc. Of all the fields of application, agriculture is of primary interest. In particular, crop monitoring may show temporal variability (different phenological growth patterns) due to either natural development variation or diverse crop-management decisions made by farmers (Gilbertson and van Niekerk, 2017). Recent relevant technologies that have emerged are compatible with the requirements of agricultural applications (spatial resolution should at least be decametric and revisit frequency lower than a week) (Weiss et al., 2020). There are various specific remote sensing platforms (e.g. satellites, Unnamed Aerial Vehicles, UAVs) that offer interesting options in terms of image resolution and flight agility for crop monitoring. In the case of agriculture-focused studies using RS, satellites remained dominant in agriculture in recent years (Weiss et al., 2020). This could be explained by the presence of cost-free satellites such as Sentinel-2 that are able to measure a sample (e.g. a crop field), under various information sources (i.e. a combination of 13 spectral bands) with a temporal repetition of the measurement (i.e. every 5 days) that allow obtaining information of interest of the field and its temporal evolution (e.g. physiological, structural and phenological traits) and how these latter vary in space (Weiss et al., 2020).

However, all these technological developments also entail the need to develop methodologies to deal with the large volume and complexity of data (Huang et al., 2018). More specifically, the assumption behind crop monitoring is that remotely sensed data represent the seasonal vegetation signal in a meaningful way, and that the underlying vegetation variation comprehensively explores the relationship of spatio-temporal-spectral dimensions. According to Kroonenberg et al. (2009), the data resulting from this high-dimensional approach, e.g. considering spatial, temporal and spectral information, has a three-way structure. In previous work, the common approach was to analyse three-way data after aggregating them one after the other or to handle them as only a two-way data set (Kiers and Mechelen, 2001). For example, a widely used approach to mitigate high dimensionality is to carry out feature extraction and/or feature selection (Gilbertson and van Niekerk, 2017), such as vegetation indices (VIs) (e.g. normalised difference vegetation index, NDVI or enhanced vegetation index, EVI). In fact, in this example and in the vast majority of studies that reduce the high dimensionality of RS data for agricultural applications, the analysis of dynamic phenomena such as vegetation development is limited to the study of the evolution of an index over time. However, going beyond this kind of classical approach can be interesting, not only from a temporal point of view to identify temporal variations that may indicate seasonal changes in reflectance, but also to explore spectral richness through changes in reflectance as a function of wavelength. The

risk of using only vegetation indices to define the spectral characteristics of crops is that other wavelengths provided by the sensor that could lead directly to the identification of an object and/or its status in a more complete way are ignored. Moreover, this type of feature extraction methods are typically calibrated over experimental observations and thus are constrained by the representativeness of the calibration dataset (Khanal et al., 2020). Having complete ground truth data sets that are representative of the large number of agronomic variables that need to be taken into account to validate the relevance and interest of these new sources of information for agriculture are rarely available due to the complexity of their acquisition. In light of this, Esbensen and Geladi (1989) determined that unsupervised, i.e. exploratory, data analysis methods are very useful in situations characterised by little or no domain-specific knowledge, e.g. when analysing a scene without any prior knowledge of the 'ground truth'.

Within the current alternatives proposed in the literature to address this challenge for high-dimensional data, factorisation techniques are an important class (Verbeeck et al., 2020). One of several decomposition methods for multi-way data known to be the simplest and most restricted unsupervised model available for higher order arrays is Parallel Factor Analysis (PARAFAC) (Harshman, 1970). Although PARAFAC is referred to as a generalisation of the classical Principal Component Analysis (PCA), in the former, there is no rotation problem, i.e. pure spectra can be recovered from multi-way spectral data (Bro, 1997). In practice, such methods take a high-dimensional dataset, and decompose it into a (typically reduced) number of underlying trends in the observed data (Verbeeck et al., 2020). As this reduced dimension representation allows for the visualisation of temporal and spectral signatures, which opens up the possibility to discover the underlying principles that guided the sample's (crops) responses over time (Kroonenberg et al., 2009).

It is important to note that this type of unsupervised approach has certain limitations. Assuming that there is no access to a large number of agronomic variables to explain the observed phenomena appropriately, it is important to contextualise the application of the PARAFAC method on a sufficiently large spatial scale, such as a regional scale. The wide heterogeneity of the environments that constitute the region, e.g. different soils, different climates and different cropping practices, can help to provide large variations either spatially and temporally in relation to agronomic traits. In addition, given the scarcity of ground truth data, it is necessary to proceed by other means to validate the relevance and potential of the information extracted. To this end, as shown by Pichon et al. (2019), the evaluation of results through expert observations, as a different alternative to what is normally considered in scientific literature, allows for a more integrative and systemic validation approach by taking into account crop characteristics, climatic and meteorological conditions of the year, local specificities, etc.

Given that it is a new paradigm for agricultural monitoring to have so much data and in near real time, it is important to explore, as comprehensively as possible, all the information that RS can provide in terms of spectral, temporal and spatial dimensions. To answer this question, a holistic exploratory analysis has been carried out using the PARAFAC methodology applied to a time series of Sentinel-2 data concerning viticulture in Southern France. To our knowledge, this is the first time that this type of method from the field of chemometrics is applied to remote sensing multispectral images. The specific objectives of this study are:

i) to propose a multi-way approach aiming to identify spectro-temporal profiles from a time-series of images, to assess the value of the approach for its potential to generate knowledge in a viticulture case study, to determine whether the spatial patterns highlighted can be considered relevant with regard to the experts' observations and;
ii) to address the possible limitations of the approach when dealing with large-scale time series of multispectral images without ground truth data.

3.1.2 Materials and Methods

3.1.2.1 Notations

For N-way arrays, capitalised bold and underlined characters will be used, e.g. $\underline{\mathbf{X}}(I,J,K)$ indicates a 3-way array with I samples at J times described by K wavelengths. For matrices, bold and capitalised characters will be used, e.g. **X**, and for vectors, a lower case bold character will be used, e.g. **a**. Upper case and italics characters will be used for scalars, e.g. the number of wavelengths, K and lower case characters will be used for running indexes, e.g. a_i is the i^{th} element of the vector a.

3.1.2.2 PARAFAC method

PARAFAC is a decomposition method used to decompose N-way arrays into distinct components. It is based on an alternating least square (ALS) algorithm where the data signal is decomposed into a set of trilinear terms and a residual array (Yang et al., 2018). Following Ouertani (2014), the PARAFAC decomposition of a three-way array $\underline{\mathbf{X}}$ is the decomposition in the form of the sum of a minimum number of three-way arrays of rank one (Equation 3.1):

$$\underline{\mathbf{X}}_{ijk} = \sum_{f=1}^{F} \mathbf{a}_{if}\mathbf{b}_{jf}\mathbf{c}_{kf} + \underline{\mathbf{E}}_{ijk}\ i = 1,\ldots, I;\ j = 1,\ldots, J;\ k = 1,\ldots, K;\ f = 1,\ldots, F \qquad \text{(EQ.3.1)}$$

where $\underline{\mathbf{X}}_{ijk}$ is the reflectance value of the *i*th sample at the *j*th variable (temporal mode) and at the *k*th variable (spectral mode). Each *F* corresponds to a PARAFAC component and each such component has *I* **a**-values (scores); one for each sample. Each component also has *J* **b**-values as well as *K* **c**-values; one for each loading. The $\underline{\mathbf{E}}_{ijk}$ parameter is the residual noise, representing the variability not accounted for by the model. In addition, a valid PARAFAC model will provide an estimate of the pure spectra of the constituents of the sample under study. It will also provide an information of the concentration of the constituents along the time, i.e. the components resulting from a valid PARAFAC model will allow a direct interpretation of spectral dynamics. For example, the PARAFAC method is widely used today in the analysis of multivariate data (structured in a three-dimensional excitation emission matrix, EEM) to quantitatively compare the content of fluorescent organic compounds in samples (Xu et al., 2021). The EMM-PARAFAC combination allows dividing fluorescent organic compounds into several independent components according to their unique properties and structures, which provides a basis for better understanding the dynamic changes of fluorescent compounds (Xu et al., 2021).

A schema of the PARAFAC decomposition of a three-way array is given in Figure 3.1.

The PARAFAC model was run and validated following Andersson and Bro (2000) using the *N-way* Toolbox in Matlab version R2015a.

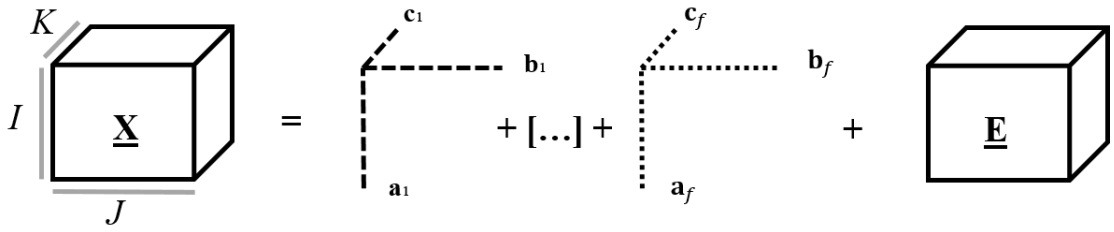


FIGURE 3. 1 - Schema of the PARAFAC decomposition of a three-way array (Ouertani, 2014).

3.1.2.3 Case-study

Remote sensing data are characterised by several sets of information that are collected as a time-series. This type of multi-way data could provide valuable spectral-temporal-spatial information to characterise crops. In order to have a large variability of conditions, which will be favourable to highlight the traits of a particular crop, it has been chosen to focus on the study at a regional scale. Research at the regional scale allows, among other issues, the monitoring of how different environmental and management conditions affect vegetation development (Giovos et al., 2021). It also may provide insights on how an extreme climate event may affect the vegetation activity (Zhou et al., 2019). Moreover, the crop selected for the study-case is

grapevine, as it is a major crop in the South of France at this scale (see the following section, study area).

For remote sensing applications in agriculture at regional scale, recent studies have demonstrated the potential of no-cost satellite imagery, such as Sentinel-2 satellites (Sozzi et al., 2020). In terms of revisiting time and spatial and temporal resolution, Sentinel-2 satellite imagery is well suited to the requirements of monitoring agricultural fields, with a high potential to extract relevant agronomic information.

In this study, the decomposition PARAFAC method was therefore applied to vine cultivation at a regional scale using multispectral times series from the Sentinel- 2 satellites.

Study area

The study area corresponded to 4978 samples (vineyard blocks) extracted from the graphical parcel register of France (RPG) from a large wine-growing region located in the south of France, the Languedoc-Roussillon (LR). The LR vineyards extend over approximately 27 400 km², covering four French administrative sectors: Gard (A), Hérault (B), Aude (C) and Pyrénées-Orientales (D) (Figure 3.2). The study area encompasses a great variability of pedo-climatic conditions and a great diversity of varieties, training systems, etc. (Fernández-Mena et al., 2021). It is assumed the changes over time of the reflectance values from Sentinel-2 satellites over 4978 vineyard block samples will provide agronomic insights of interest. Assuming that a unique solution can be expected if the Sentinel-2 data are trilinear, i.e. that there is a relationship between the three-way structure of these data (samples × times × wavelengths) this would imply that the true pure spectra could be found if the correct number of components is used and the signal-to-noise ratio is appropriate (Bro, 1997).

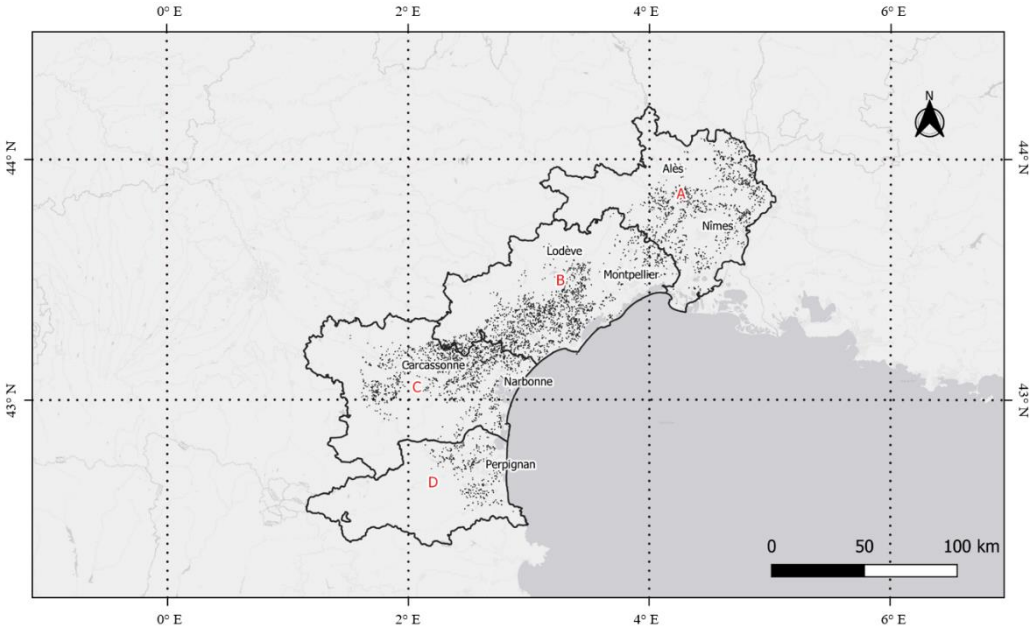


FIGURE 3. 2 - Location of the 4978 vineyard blocks within the study area in Southern France for the four administrative sectors: Gard (A), Hérault (B), Aude (C) and Pyrénées-Orientales (D).

Remote sensing data

The Sentinel-2 (A/B) satellites provide 13 spectral bands from the Visible (Vis) and Near InfraRed (NIR) to the shortwave infrared (SWIR). They provide coverage of land surfaces on a global scale with a revisit frequency of 5 days in cloud-free conditions with a spatial resolution of 10 m, 20 m and 60 m depending on the spectral band (Devaux et al., 2019).

Sentinel-2 L2A data containing the study vineyards (Figure 3.2) were selected and processed via the Google Earth Engine (GEE) platform. Prior to the calculation of the average pixel values for each block, each date and each waveband, a 10 m inner-buffer was imposed over each vineyard block boundary extracted from the graphical parcel register of France (RPG). This 10 m inner-buffer was used to prevent information outside the block from being integrated into the analysis (Lopez-Fornieles et al., 2022). Images from the years 2019 and 2020 from 1st May to 31st August were selected. The time period considered (from May to August) for both years of the study was considered as being the most relevant for monitoring vine vegetation in the study region (Devaux et al., 2019). Images containing clouds or shadows altering the visibility of the blocks were removed from the database. For this purpose, the spectral band 10 at 1380 nm was used for the detection of visible and sub-visible cirrus clouds (Hollstein et al., 2016). About 25 images should have been potentially available on each vineyard block, however, the local atmospheric conditions over each block for each acquisition date altered the number of images available (Lopez-Fornieles et al., 2022). The final Sentinel-

2 database consisted of 4978 vineyard blocks containing 12 spectral bands and an average of 11 images (dates) per block.

3.1.2.4 Modelling

Data array construction

Two PARAFAC models were derived from the data set of 4978 samples (vineyard blocks), one for each year of the study, 2019 and 2020. To overcome the challenge of heterogeneity in the number of images per block, an interpolation was performed to obtain two continuous data cubes; $\underline{\mathbf{X}}_{19}$ and $\underline{\mathbf{X}}_{20}$. Interpolation using the Gaussian filter by Alam et al. (2008) was applied in order to have a consistent time step dimension (J) between the 1st May and 31st August for both years. The parameters involved in the interpolation setting were fixed to the Gaussian filter width (P) = 30 days and date interval (N) = 5 days to simulate the revisit time of the Sentinel-2 satellites (Lopez-Fornieles et al., 2022). At the end of the interpolation step, the data set was meaningfully arranged in two three-way arrays $\underline{\mathbf{X}}$ ($\underline{\mathbf{X}}_{19}$ and $\underline{\mathbf{X}}_{20}$) of dimensionality 4978 (samples, I) × 25 (times, J) × 12 (wavelengths, K).

PARAFAC model

A separate decomposition of the two arrays; $\underline{\mathbf{X}}_{19}$ and $\underline{\mathbf{X}}_{20}$ into trilinear components (F) were performed by PARAFAC method. The trilinear model is found to minimize the sum of squares of the residuals, $\underline{\mathbf{E}}_{ijk}$ in the model. Each component consisted of one score vector and two loading vectors, i.e. temporal and spectral loadings. A PARAFAC model of a three-way array, e.g. $\underline{\mathbf{X}}_{19}$, was given by one score matrix **A** and two loading matrices, **B** and **C,** with its respective $\mathbf{a}_{if}$, $\mathbf{b}_{jf}$ and $\mathbf{c}_{kf}$ (Figure 3.3). The parameter $\mathbf{a}_{if}$ is the score of the ith sample of the fth component; $\mathbf{b}_{jf}$ is the loading specific to reflectance intensity at the time j of the fth component; $\mathbf{c}_{kf}$ is the loading estimate of the reflectance spectrum k of the fth component. Essentially, the PARAFAC model provides an estimation of the relative concentration, i.e. the relative amount of the fth component in each sample (matrix **A**) and temporal and spectral properties of the components loadings matrices (**B** and **C** respectively), which can be used to interpret the spectral dynamics of the constituents of the samples. For 2020 ($\underline{\mathbf{X}}_{20}$), the same PARAFAC decomposition scheme was used as for 2019.

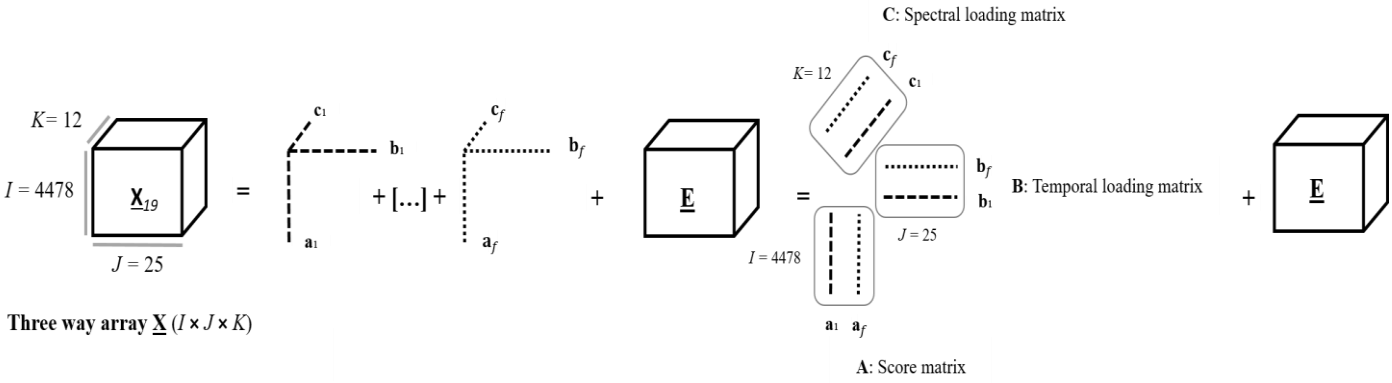


FIGURE 3. 3 - Schema of the PARAFAC decomposition model of three-way array for the year 2019 ($\underline{\mathbf{X}}_{19}$). For the record, $\underline{\mathbf{E}}$ corresponds to the residuals.

Component selection

The PARAFAC models with two to five components were fitted to the data in order to determine the correct number of components. A *F*-component model was validated following the CORe CONsistency DIAgnostic (CORCONDIA) of Bro (1997). The PARAFAC model is considered to be valid if the value of the CORCONDIA is close to 100%. If the CORCONDIA value is around 50%, the model is considered unstable and if the value is close to 0 (even negative), then it is considered that the data cannot be described by a tri-linear model (Ouertani, 2014). The CORCONDIA value will mostly decrease with the number of components. It decreases very sharply after the correct number of components is exceeded. Therefore, the appropriate number of components is the model with the highest number of components and a valid CORCONDIA value.

3.1.2.5 Validation of the model

The PARAFAC models for both the years 2019 and 2020 were validated in two steps:

1. A geostatistical analysis (see next section) to test whether the score values were spatially autocorrelated. The underlying hypothesis is that the environmental variables (soil, climate and associated training practices, etc.) likely to explain the differences between blocks at the regional scale are spatially organised (not random). Consequently, if score values resulting from the PARAFAC models are not randomly distributed at the regional level, they were assumed to be related to environmental variables.
2. An expert judgement to interpret the relevance of the values of the scores in relation to their knowledge of the diversity of soil, climate and practices in the

region. The objective of this second step was to identify the relevance of the relationship between the score and the agronomic information. In order to do so, an original approach based on placing the human, i.e. domain experts, at the centre of the study was conducted. Following Pichon et al. (2019) methodological framework, a global approach, so-called 'scenario simulation' (Wacheux, 1996), was used. Thus, external reliability can be considered valid if the conclusions reached by different experts (at least two experts) coincide (Shadish et al., 2002; Drost, 2011). In order to facilitate the expert's interpretation, maps of score values were produced at the regional level. These maps were created by interpolating (see next section) the scores observed on the 4978 vineyard blocks for each year (2019 and 2020) and for each component (*F*) when spatial autocorrelation was observed.

3.1.2.6 Spatial analysis and mapping

The spatial analysis was based on the modelling of semivariograms from which the following featured parameters were derived: C_0 (nugget effect), C_1 (sill) and A_1 (range). The modelling of the semivariograms and the calculation of their parameters was used for (i) the calculation of the Cambardella Index to determine the spatial dependence and (ii) kriging to obtain the score maps. The characterisation of the spatial structure via semivariograms, as well as the realisation of the score maps were performed using GeoFis 1.0 software (Leroux et al., 2018).

Spatial auto-correlation of the score values of each component was assessed with Cambardella Index (I_c) (Cambardella et al., 1994) (Equation 3.2):

$$I_c = \frac{C_0}{C_0 + C_1} \quad \text{(EQ. 3.2)}$$

where C_0 is the nugget effect and C_1 is the sill of the semivariogram model. According to Martínez and Gomez-Miguel (2017) the following thresholds were used:

1. if I_c is less than or equal to 25%, the distribution is considered strongly spatially organised (high auto correlation),
2. if I_c is between 25 and 75%, the distribution is considered moderately spatially organised and,
3. if I_c is higher than 75%, the distribution is considered weakly spatially organised.

Score maps were obtained using point kriging interpolation (Oliver and Webster, 2014). The latter was performed on a grid of points regularly spaced 1000 m apart within the geographical boundary of the LR region (Lopez-Fornieles et al., 2022a).

3.1.2.7 Experts evaluation

Following Pichon et al. (2019), 6 experts were selected. These 6 experts were chosen in order to benefit from an integrated assessment of the results obtained over the whole region, taking into account simultaneously their knowledge of plant development, the cultural practices of the winegrowers in each part of the region, the soil and its variability, etc.

In order to validate the experts' consistency, they conducted a self-assessment on the criteria considered important by the authors' perspective for the study (Table 3.1). This self-assessment consisted of asking the experts about their level of knowledge in each of the domains, as well as their knowledge of the sectors within the LR region. This information was also used to analyse the opinions expressed by the experts during the working session.

TABLE 3. 1 - Description of the interviewed domain experts. Rating range from 0 to 5, with 0 being the minimum and 5 of the maximum knowledge/expertise. The term Vegetation Indexes is abbreviated as VIs.

	Experience grapevine cultivation				**Knowledge of viticulture of the LR region**				**Knowledge of pedo-climatic characteristics of the LR region**		**Expertise with remote sensing images**		
Expert	Vine Physiology	Irrigation	Technical itineraries	Climate impact	Aude (A)	Gard (B)	Hérault (C)	Pyrénées-Orientales (D)	Soil type	Micro-climate	Data processing	VIs	Resolution scale
A1	5	5	4	4	4	2	4	3	2	2	3	4	3
A2	4	3	3	3	3	3	4	4	3	3	4	4	4
A3	4	4	4	4	1	3	2	1	3	3	1	2	1
A4	4	5	4	4	4	4	4	4	4	4	4	4	1
A5	2	1	3	2	3	1	5	1	5	5	1	1	3
A6	4	5	5	4	4	4	4	3	4	3	3	5	4

A ‘scenario simulation’ session was conducted with experts on 31st May 2022. The general workflow of the session is presented in Table 3.2 following Pichon et al. (2019) methodological framework. During the session, the score maps were hand-delivered to each expert and projected on a screen managed by an animator (zoom in and out) according to requests from the experts. These score maps were previously elaborated by adding

characteristic elements (cities, watercourses, roads, etc.) that would allow the experts to orientate themselves on the maps. It should be noted that an explanation session of the scores values meaning was necessary so that the experts could understand the maps and interpret them in the best possible way using a colour coding defined by the authors. Maps were displayed in QGIS 3.2.3 (Open Source Geospatial Foundation, http://qgis.osgeo.org). Experts were asked to first write down their opinions before oral group discussions. In the oral group discussion, the animator wrote down the experts' comments and reactions.

TABLE 3. 2 - Description of the general workflow of the 'scenario simulation' session.

	Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**
Duration	40 minutes	20 minutes	40 minutes	20 minutes	40 minutes
Type of session	Individual (written)	Individual (written)	Collective (oral)	Collective (oral)	Collective (oral)
Presentation of maps	Once at time for the year 2020	All together for the 2020 year	Once at time for the year 2020	All together for the 2020 year	All together for the years 2019 and 2020

It was agreed that the experts would react mainly to score maps derived from the 2020 years, because it was considered to be most representative of common crop-soil-climate interactions in the region. Indeed, the year 2019 was assumed to present specific behaviour (although interesting) as a heatwave which strongly affected vine growth hit the region at the end of June. Score maps of 2019 were then presented in a second step as a particular case-study.

3.1.3 Results

3.1.3.1 Component selection

PARAFAC modelling was done by fitting a series of models, i.e. increasing the number of components from two to five to look for the correct number of components that fit the data. Table 3.3 presents the results obtained for $\underline{\mathbf{X}}_{19}$ and $\underline{\mathbf{X}}_{20}$. With two components, the observed CORCONDIA and explained variance values were 100 % - 99.62 % and 100 % - 99.57 %) for the years 2019 and 2020 respectively. For a high number of components, the CORCONDIA values drop significantly, showing that the two-component model is the best for both years. Since a 2-component model was validated according to CORCONDIA, the results in Table 3.3 also justified the suitability of applying the PARAFAC methodology assuming a trilinear structure in the presented dataset.

TABLE 3. 3 - CORe CONsistency DIAgnostic (CORCONDIA) by fitting a series of models using two to five components for $\underline{\mathbf{X}}_{19}$ and $\underline{\mathbf{X}}_{20}$ databases.

Year	Number of components (*F*)	CORCONDIA (%)	Variance explained (%)
2019	1	100	99.25
2019	**2**	**100**	**99.62**
2019	3	15.65	99.74
2019	4	-6.34	99.82
2029	5	-0.72	99.88
2020	1	100	99.20
2020	**2**	**100**	**99.57**
2020	3	11.15	99.68
2020	4	-0.23	99.79
2020	5	-0.44	99.85

3.1.3.2 Data signal decomposition

A two-component PARAFAC corresponds to two pure spectra of constituents. For both PARAFAC models (years 2019 and 2020), the score in the matrix **A** (4978 × 2) contains estimated relative concentrations of the two components in the 4978 samples. The matrix **B** (25 × 2) estimates temporal loadings and the matrix **C** (12 × 2) estimates spectral loadings.

The spectral loadings (**C**) and the temporal loadings (**B**) for the two components selected (Co1 and Co2) from the PARAFAC model of the years 2019 and 2020 are presented in Figure 3.4 and Figure 3.5 respectively.

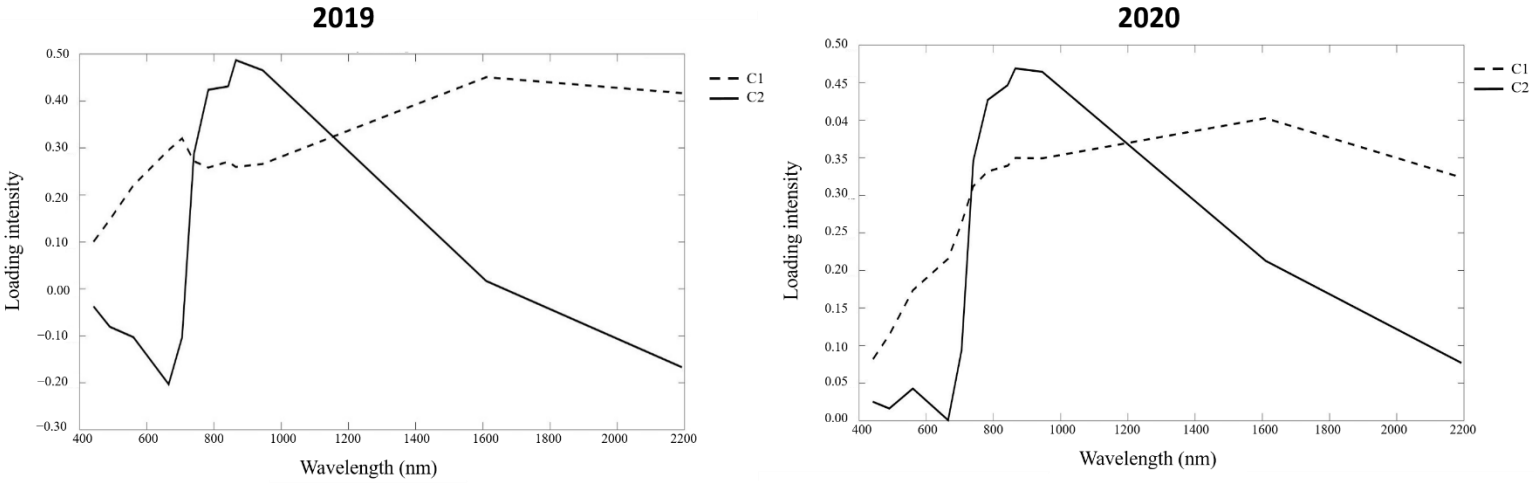


FIGURE 3. 4 - Two-component PARAFAC models for 2019 and 2020 years including spectral loadings (matrix **C**) for both components (Co1 and Co2).

Regarding Co1 for both years, it can be observed that the loading intensity is characterised by an increase from the Visible (Vis) to the Near InfraRed region (NIR). It should be noted, though, that in the year 2019, there is a slight drop in intensity between the wavelengths ranging from 700 nm to 1000 nm. Concerning Co2 for both years, the spectral loading is characterised by: i) low intensity in red wavelengths (650 - 680 nm), ii) the highest intensity in the NIR range (785 - 875 nm) and iii) a decrease in intensity in the short-wave infrared region (1360 - 2200 nm). Taking into account that matrix **C** represents the pure spectra and considering the study case, the components were identified as soil (Co1) and vegetation (Co2). Indeed, typically, soil spectral reflectance is characterised by a steady increase with wavelengths from the Visible (Vis) to the Near InfraRed region (NIR), except at 950 nm, 1200 nm and also in 1350 nm, where reflectance decreases (Khadse, 2012). As for the typical vegetation spectral reflectance, it exhibits the following characteristics: i) low reflection in the blue (458 - 523 nm) and red wavelengths (650 - 680 nm), ii) relatively more reflection in green wavelengths (543 - 578 nm), iii) reflectance in the NIR range (785 - 875 nm) is the highest and iv) the short-wave infrared region (1360 - 2200 nm) is mainly determined by less reflectance (Roman and Ursu, 2016). Given that grapevine cultivation is seasonal and is a row crop, the PARAFAC model clearly identifies the mixture of these two components. In fact, for both years, the Co2 spectral loadings represent a ‘reference’ vegetation profile (which refers to the notion of the pure spectra in chemometrics) that probably includes vines and grass reflecting the typical signature of healthy vegetation while the Co1 spectral loadings represent a ‘reference’ soil profile at the region level. It actually summarises different soil types of the region, probably with different spectral characteristics. This could explain why the spectral loading identified as soil followed a slightly different general trend than that described as a typical soil profile by Khadse (2012).

Figure 3.5 presents the temporal loadings (**B**) for both years 2019 and 2020 and both components (Co1 and Co2). It was observed that the two components showed large differences between them in their temporal loadings in the two consecutive years. Regarding Co1, in the year 2019 its intensity in its time profile decreased over time, while in the year 2020 it remained more stable. Regarding Co2, its spectral profile in 2019 peaked between 17 and 27 July with a slow decrease afterward, while for 2020, Co2 shows a maximum around 11 and 26 June followed by a very steep decline later on in the season. Indeed, the temporal loadings of both years seem to show that Co2 intensity (vegetation) increases with the vegetative development of the vine (June-July) and, conversely, Co1 intensity (soil) decreases with the growth of the vine. In fact, the decreasing temporal profile of Co1 observed throughout 2019 reaffirms the hypothesis that it may be a compensatory profile of the vegetation at certain times during the time period considered. For the record, 2019 was marked by a heatwave late June which affected a significant part of the region which may explain a compensatory effect afterwards in the season. With regard to Co2, the maximum peak of intensity shifted between 2019 and 2020, moving forward by almost one month in 2020. This could be explained if the time of vegetative development of the vineyards had been different between the two years studied.

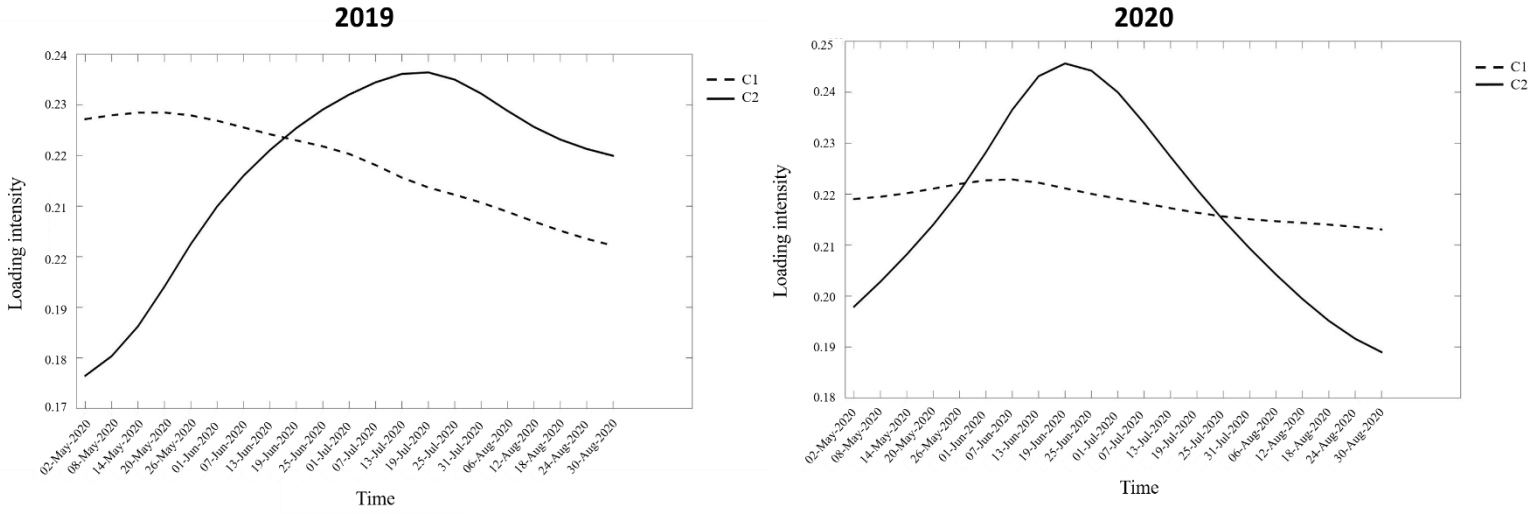


FIGURE 3. 5 - Two-component PARAFAC models for 2019 and 2020 years including temporal loadings (matrix **B**) for both components (Co1 and Co2).

Given that the samples studied were blocks of vineyards, the spectral and temporal profiles indicate ‘reference’ behaviour for each component (soil and vegetation) at the regional scale, but each block may differ more or less from this ‘reference’ behaviour. Thus, once the models have been validated, the data can be examined with respect to the variability of each component found, i.e. with respect to the values of the score matrix (**A**) for both years.

3.1.3.3 Spatial analysis and characterisation

The score value maps (Figure 3.6) represent the spectro-temporal profiles of each block in relation to each component and its corresponding year, e.g. for Figure 3.6.a, the estimated concentration value of the Co1 (soil) component of each block is plotted in relation to its spectro-temporal profile in 2019 at the regional LR level. Therefore, the score map for Co1 was supposed to highlight blocks that were directly related (or counter -related) to the identified spectro-temporal signature for the soil in 2019.

The spatial organisation of the score maps for all the years and all the selected components is confirmed by the semivariogram model (Table 3.4), which showed that at least 40 - 50 % of the variability is explained by a spatial phenomenon. The year and component assessed with the strongest spatial structure were the soil component (Co1) for the year 2019 with an I_c of 45 %. These results support the assumption that scores values are not randomly organised, but spatially structured and likely to vary according to environmental variables.

The score maps obtained after kriging (Figure 3.6) show the spatial organisation of the blocks that have: i) similar spectro-temporal profile (high scores) and ii) different spectro-temporal profile (low scores) to the 'reference' spectral profile for each component and year. Therefore, the more different the time spectral profiles are, the lower the scores will be, becoming negative if they are opposite to the 'reference' spectro-temporal profile. The spatial patterns for soil (Co1) between the two years are very similar, showing that the same vineyard blocks had both high (light colour) and low (dark colour) score values in the same areas for both years. Regarding the spatial patterns of vegetation (Co2) for both years, a general trend of higher values (light colour) of the scores in the northern part of the region was highlighted. For other sectors, the spatial patterns are less similar between both years. It should be noted that for the Co2, the spectral loadings (Figure 3.4) followed the same profile in both years but, in contrast, the temporal loadings (Figure 3.5) showed a significant temporal shift. This implies that the temporal dimension may have a major role in the difference in score values for each vineyard block for both years, i.e. in the differences in spatial patterns for each year represented.

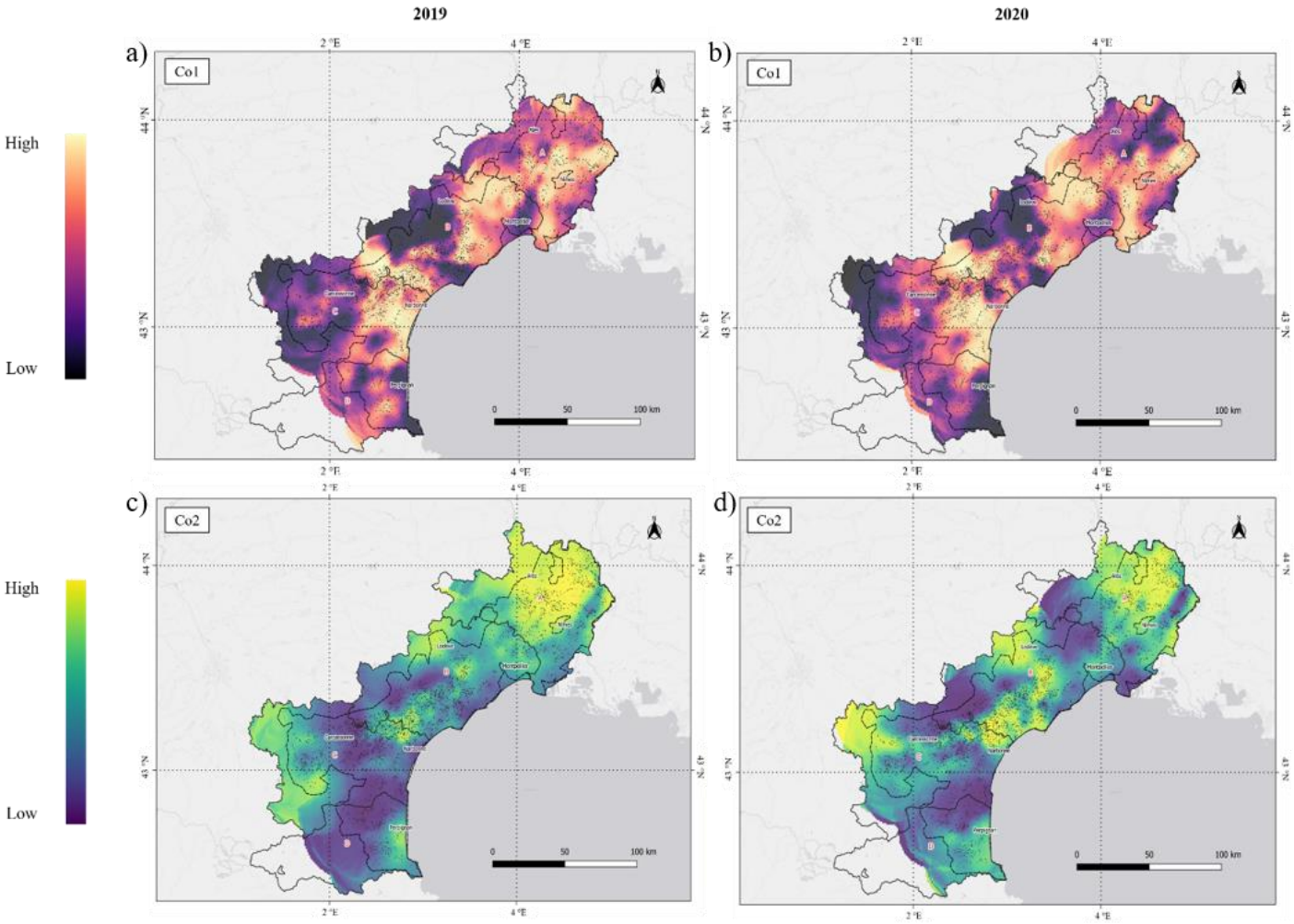


FIGURE 3. 6 - Kriged maps of the score values of each vineyard blocks at the regional scale for a) component 1 (Co1) identified as the soil spectro-temporal profile in the years 2019 and b) 2020 and component 2 (Co2) identified as the vegetation spectro-temporal profile in the years c) 2019 and d) 2020. Light colours represent high score values and dark colours represent low score values for both components.

TABLE 3. 4 - Semivariogram parameters and spatial variability index for score values. A_1 (Range), C_0 (Nugget), C_1 (Sill) and I_c (Cambardella Index).

Year	Components	Semivariogram model	A_1 (Km)	C_0	C_1	Ic (%)
2019	Co1	Exponential	6	0.067	0.079	45
	Co2	Exponential	6	0.026	0.022	53
2020	Co1	Exponential	5	0.205	0.113	64
	Co2	Exponential	5	0.145	0.086	62

3.1.3.4 Comparison of cross-observations by experts

As the spatial patterns for Co1 between 2019 and 2020 were almost identical (Figure 3.6.a and 3.6.b), it was considered that the score map of the Co1 for the year 2019 would not provide any relevant information to the study for the experts. Therefore, the workflow of cross-validation by experts focused first on the two 2020 components (Co1 and Co2) and then the vegetation component (Co2) of the year 2019 was presented only for comparison with 2020. Table 3.5 presents the extent to which each type of observation was made by the expert group for the 2020 year.

TABLE 3. 5 - Number of experts who have made similar observation in each sector of the LR region for the Co1 and Co2 score maps for the year 2020.

Component	Figure Number	Observations on	Sector A	Sector B	Sector C	Sector D
Co1 (Soil)	3.6.a	Variation of pedological units	5	5	2	1
	3.6.b	Variation of geological units	5	4	4	0
Co2 (Vegetation)	3.7.a	Difference in grape varieties (phenology, yield, etc.)	4	2	3	1
	3.7.b	Variation of vigour	2	1	3	1
	3.7.c	Presence/absence of irrigation	4	4	1	1

It clearly appears that a significant number of experts made similar observations on Sectors A and B of LR region. Moreover, more expert observations were related to Co1 compared to those of Co2. This result can be explained by the scale of the study where vegetation heterogeneity may be more difficult to analyse properly by the experts given the large diversity of situations. Experts were obviously more comfortable with identifying and explaining differences in spatial patterns within sectors that they were familiar with. Figure 3.7 shows a summary map of each type of observation for the year 2020 made by the group of experts related to Co1 for each sector.

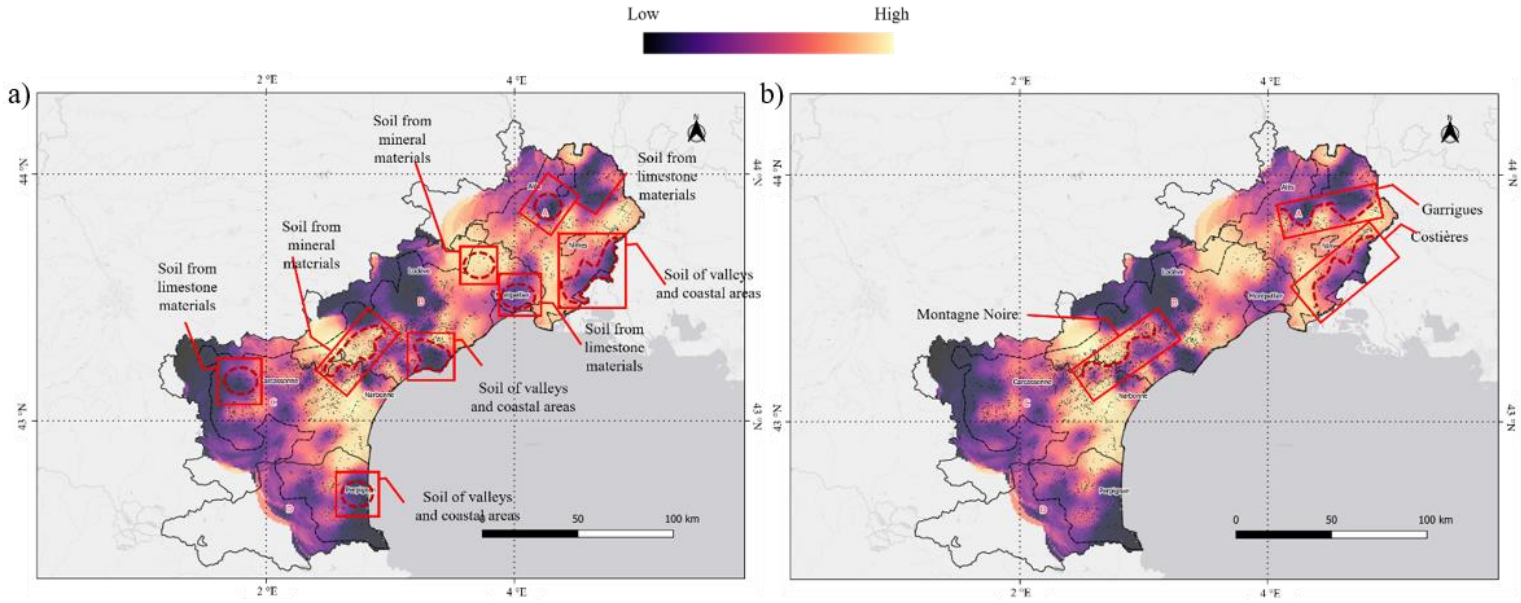


FIGURE 3. 7 - Summary map of each type of Co1 observation being: a) the variation of the pedological units and b) the variation of the geological units for the year 2020 made by the group of experts for each sector. Light colours represent high score values and dark colours represent low score values for Co1. Areas of the experts' observations are highlighted in red squares.

The observations made by individual experts coincide in its vast majority with at least one other expert. Therefore, if at least two experts are in agreement, the relevance of the identified area is valid according Wacheux (1996). Concerning the soil units (Figure 3.7.a), experts identified that spatial patterns are related to the depth of the soil. In terms of colour range, it was perceived that the darker areas could represent valley and coastal soils, e.g. fluviosols, as well as calcareous soils. The former are soils containing mostly coarse elements (gravel, pebbles, stones…) with a thickness greater than 50 cm and the latter are defined as medium to thick soils (over 35 cm thick), developed from calcareous materials. On the other hand, the lighter zones (less distinct from a visual point of view) could represent mineral soils, e.g. rankosols or lithosols, which are characterised by a thickness of less than 30 cm. It should be noted that this spatialisation of the different geological units was directly related to the information on the soil available water capacity provided by 4 experts. Regarding Figure 3.7.b, some major geological formations from LR region were largely commented on by 5 of the experts. Three clear boundaries between the two colours were observed and outlined in sectors A-B-C of the map.

Regarding Co2 (vegetative factor), it should be noted that the colour coding does not always represent the same phenomenon but rather differences when compared to the ‘reference’ spectro-temporal profile for the 2020 year. That is, the yellow vineyard blocks represent a spectro-temporal profile very similar to the ‘reference’ vegetation profile extracted by the

PARAFAC method. In contrast, the blue vineyard blocks represent plots that differ from the 'reference' spectro-temporal profile of the vegetation. Therefore, as these colour variations may be due to differences in the spectral and/or temporal profiles in relation to the 'reference' profile, there were some inconsistencies in the experts' interpretation of the meaning of the colour codes. A clear example was the comments on the phenological state of the vineyards by sectors in Figure 3.8.a. Each of the areas highlighted in Figure 3.8.a were selected by the experts as characteristic of late growth vineyard development but the colours are different in each area. These inconsistencies could be explained not by the late phenological stage but by the different grape varieties selected and adapted to this late growth in each sector, e.g. a predominant presence of Chardonnay variety in sector A and a predominance presence of Mourvèdre variety in sector C. Since the grape varieties are different, different management practices, different training systems, etc., could be involved and thus explain why on some maps the observed trends and the resulting agronomic interpretations are difficult to interpret.

Concerning Figure 3.8.b, certain spatial structures which are characterised by high/low vigour were commented on by a minority of the experts. A possible explanation given by one of the experts focusing on sector A was the presence of the distinct denominations areas that characterise the different vineyards. In particular, the yellow area corresponds to the Protected Geographical Indication (PGI) label and the blue area to the Protected Designation of Origin (PDO) label. Two different identification implies e.g. different yield objectives, different soil cultivation, grape varieties and management practices. The latest observations made by the experts are shown in Figure 3.8.c concerning zones where irrigation is available through regional facilities. The experts largely addressed these observations for sectors A and B.

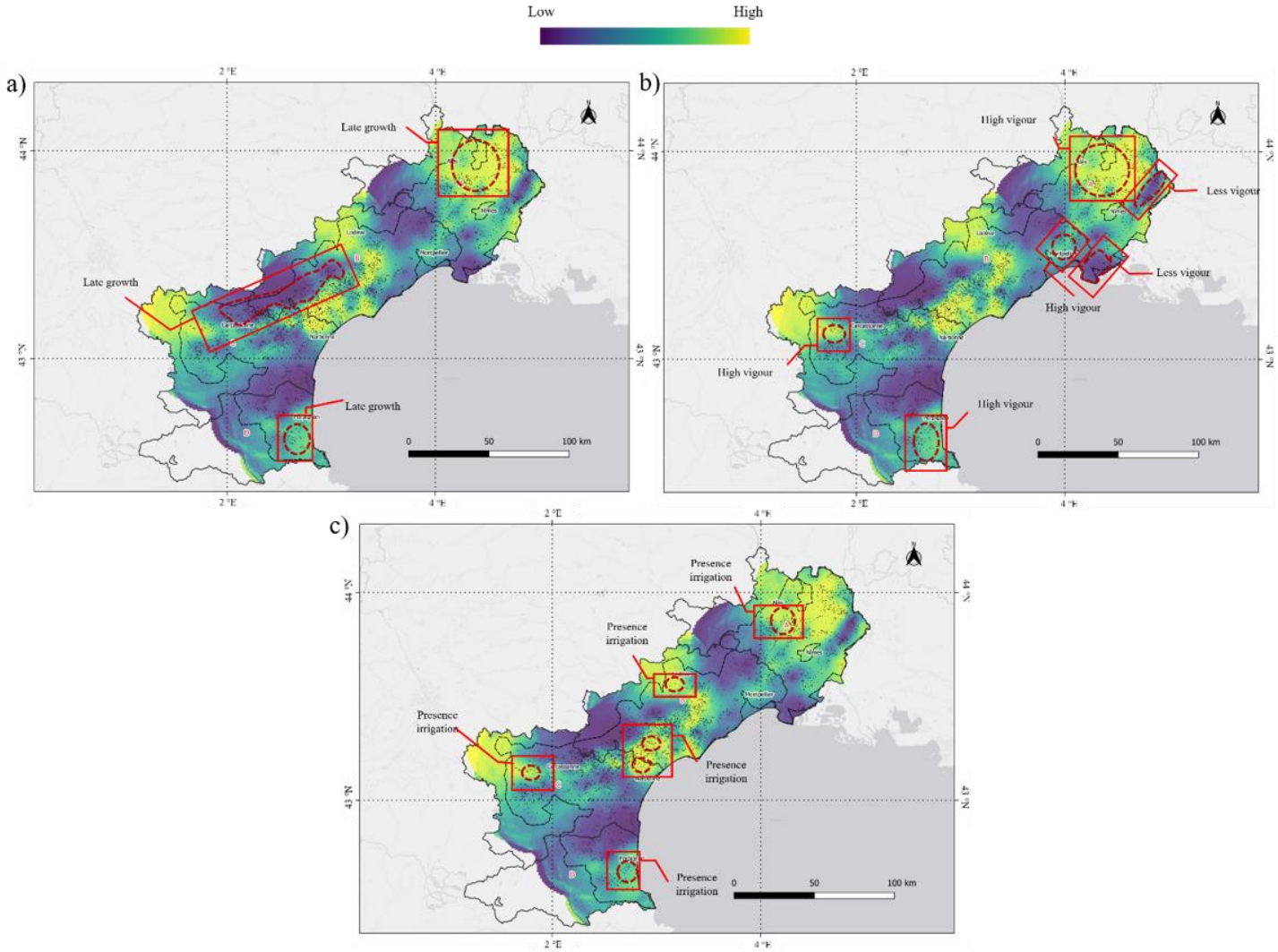


FIGURE 3. 8 - Summary map of each type of Co2 observation being: a) the different grape variety (phenology, yield, etc.), b) the variation of vigour and c) the presence/absence of irrigation for the year 2020 made by the group of experts for each sector. Light colours represent high score values and dark colours represent low score values for Co2. Areas of the experts' observations are highlighted in red squares.

For the final step of the workflow (Table 3.2) it was decided to present only the score map of the year 2019 for the vegetation component (Co2) as the score map of the Co1 confirms the stability of this component from one year to another and the temporal stable factors (mainly related to soil) that it is related to. Therefore, when comparing the two years for the Co2 component, the experts observed the same spatial patterns but less pronounced for the year 2019 (Figure 3.9.a). This attenuation of the yellow colour related to higher score values could be justified by the heatwave episode from the 23rd of June to the 8th of July of 2019 experienced

by the LR wine-growing region. The heatwave would have homogenised the differences between vineyards and therefore the corresponding spectro-temporal profiles.

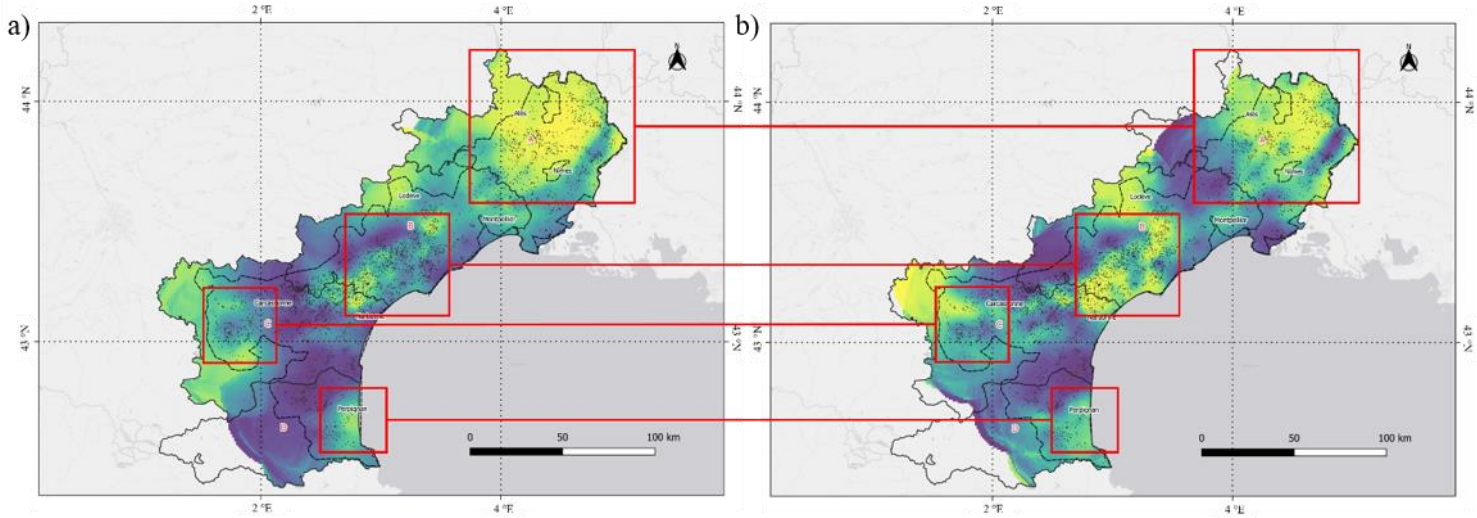


FIGURE 3. 9 - Visual comparison of the main spatial structures of the Co2 score maps for a) the year 2019 and b) the year 2020. Light colours represent high score values and dark colours represent low score values for Co2. Observations made by experts in 2020 that these same experts found in 2019 are highlighted in red squares.

3.1.4 Discussion

A remote sensing time series study for the regional characterisation of vineyard blocks was provided by the application of the PARAFAC algorithm. Results showed the potential as well as the limitations provided by the application of an unsupervised complex data analysis method focusing on grapevine production at the regional scale. The validated application with a practical framework of expert winegrowers' opinions demonstrated the complexity as well as the added value of considering the feature approach in the temporal and spectral dimensions for interpretation purposes to identify relevant region/local specificities in a grapevine context.

In order to analyse a time series multispectral images to assess the value of simultaneously considering spectral and temporal information over the LR wine growing region, the PARAFAC method was used to tackle the issue of analysing the three-way Sentinel-2 datasets (year 2019 and year 2020 dataset). It should be noted that a trilinear structure was assumed *a priori* in the analysis of the datasets. This assumption could be partly accepted because the fit of the data obtained was satisfactory according to the CORCONDIA criterion (100% for both years) and because the shape of the spectro-temporal profiles obtained was coherent from an agronomic point of view. Several authors, including De Juan and Tauler

(2001) conclude that methods based on the trilinear structure assumption will always fit the data less efficiently than others that do not, such as Tucker3 or Multivariate Curve-Resolution Alternative Least Squares (MCR-ALS). However, the PARAFAC decomposition accurately modelled soil dynamics and vegetation dynamics for 2019 and 2020 years. The interpretability of the spectro-temporal profiles was useful to understand the variation of the spectral response of crops (soil-vegetation dynamics) over time.

Since PARAFAC decomposition allows for unique solutions to be obtained, i.e. pure spectra, the modelling was considered relevant for the study of the temporal dynamic of vineyard blocks at the scale of the region. However, it is essential to place the results presented in this paper within the reality of large-scale unsupervised case study data. It is apparent that the study of such variable samples representative of a large geographical area has limitations related to the reliability of the validation. To overcome this challenge, Wacheux's (1996) 'scenario simulation' was used, where if at least two experts confirm that there is some logic in the observations, then these observations considered valid. The proportion of different experts who made the same observation for each score map for the year 2020 can be considered a reliable indicator of the representativeness of the findings and may validate the relevance of the approach to highlight relevant agronomic information. The observations were not made in the same proportion or in the same detail for the different sectors of the LR region. Table 3.5 clearly shows that almost all the observations in sector D were made by only one expert. This does not question the approach, but given the observations of the selected experts, it is to be expected that the conclusions from sectors A, B and C were more consistent than in sector D, where only one expert had in-depth knowledge. However, it should be highlighted that the 'external reliability' of this study refers to the LR region and that the same validation framework would have been approached differently, i.e. with experts from other regions or different countries, if a different scale of study had been defined.

It is essential to place the results presented in this paper within the limits of the approach used. The main limitation of the PARAFAC methodology in relation to its application with Sentinel-2 satellite data at a regional scale is that it requires a temporal interpolation step prior to the analysis. This temporal interpolation is required in order to create a continuous data cube. However, this step necessarily involves smoothing of the spectral data, which could lead to the removal of relevant information for a proper crop monitoring. In addition to the methodological limitation mentioned above, there is also an important limitation within the expert's validation framework used. Indeed, the spatial representation of a single score value summarising the whole growing season (the same colour code could represent a variety of phenomena), probably added complexity to the interpretation by experts to identify relevant areas related to 'real' soil or vegetation dynamics. However, according to the experts, the PARAFAC analysis highlighted different relevant zones according to their knowledge. From the differences in the score value maps, different spatial patterns were visually highlighted that could potentially reveal some

interpretative clues depending on the situation of the vineyard (geographical area, soil characteristics and soil water capacity), the different phenological stages of the vineyards (different denominations area, grape variety and management practices) and the impact of the use of irrigation in certain areas.

Results validation focused on the regional scale with a group of experts able to integrate a wide range of knowledge (soils, crops, climates) to analyse the relevance of the information obtained. It is likely that similar analyses conducted at finer scales will identify new information and perhaps even interesting links to new agronomic information.

Concerning the year 2019, the validation by the expert group focused only on the vegetation component, highlighting its similarity, i.e. same spatial patterns with the subsequent year. The main observation was the attenuation of the high-value zones in the year 2019, which raised the hypothesis that the crop-heatwave interaction caused the homogenisation of the reflectance signal, creating a less differentiated spatial distribution of vine vigour at the regional scale, i.e. fewer strong concentration values (yellow colour). However, no sudden variations of the profiles were observed, as would be expected as a consequence of the impact of an extreme weather event. This could also be attributed to the fact that the temporal interpolation performed may have masked this punctual variability.

The unsupervised approach (PARAFAC) presented here represents a specific application case for the LR wine-growing region in 2019 and 2020. In view of the results, it is a type of approach that can be effective to spatialize and characterise phenomena with a temporal evolution, e.g. the spectral response of the vine canopy, providing spectro-temporal ‘fingerprints’ able of highlighting differences in behaviour. However, in this particular case study, probably as a consequence of the resolution scale, its application requires *a posteriori* expert knowledge of the observed phenomenon, thus limiting its applicability.

3.1.5 Conclusion

The work conducted in this study showed the potential of an appropriate three-way data resolution methodology such as PARAFAC in the analysis of remote sensing images time series. The results obtained from data collected over two cropping years (2019-2020), on 4978 vineyard blocks showed that the PARAFAC method provided relevant information on temporal spectral profiles, which, in turn, allowed the spatial characterisation of the LR wine region. The validation of the results was based on expert observations. Although this can be seen as a limitation, the study showed that the feasibility that expert knowledge could be useful to validate the interest of applying this to explore the data without a priori.

The practical approach of experts demonstrated that the application of specific methodologies for the resolution of complex three-way data can be optimised at various levels and be potentially useful for understanding and characterising viticulture at the regional scale. Elements such as e.g. the relationship with the landscape, the irrigation areas in relation to the soil characteristics, the spatial footprint of the soil water capacity seen as an indicator of biomass production over the season, etc., have given the authors new insights into elements to be investigated in more detail. However, the requirement to have a continuous cube can be a limiting factor in characterising isolated episodes that affect the crop growth throughout the year.

The proposed methodology is potentially transferable to other LR resolution scales. In fact, for the effective characterisation of the specificity of the spectral temporal response of agricultural-related factors at other scales, the spatial segmentation elements (different spectral time zones) highlighted by the PARAFAC methodology and identified by expert observations would be appropriate as a starting point.

3.2 Potential of the Multivariate Curve-Resolution Alternative Least Squares (MCR-ALS) method to identify temporal variations from multispectral data

3.2.1 Introduction

The practical expert approach presented in Part 3.1 validated the results of the PARAFAC method. However, it raised some questions concerning the relevant choice of the resolution method of three-way data sets to be adapted to this type of approach. A key factor for the proper interpretation of the maps (Figure 3.6) was the understanding of the meaning of the score values, i.e. that the score values integrated both spectral and temporal variation over the entire growing season for each component (soil and vegetation). The idea that a single value could represent all variability was considered for the authors' opinion to be a challenge for the interpretation. Therefore, it was decided to test another method of solving three-way datasets that would allow the temporal dimension to be squeezed in a different way. In an attempt to address this issue, the Multivariate Curve-Resolution Alternated Least Squares (MCR-ALS) was considered as a promising methodology to reconfigure the temporal dimension in the analysis. Multi-way analysis based methods such as PARAFAC offer the advantage of single solutions, while the MCR-ALS method offers multiple set analysis, i.e. it allows the flexible use of many subsets of data containing diverse information in a single dataset structure (de Juan and Tauler, 2021). Although this type of analysis still suffers from a certain degree of ambiguity, it allows working with more flexible dataset structures (no need to build a continuous data cube). Thus, the main differences between multi-way analysis and multiple set analysis algorithms are: (i) the requirements of the initial data structure and (ii) the underlying model, multilinear in multi-way analysis or bilinear for multiple set analysis. Therefore, unlike the previous method presented in Part 3.1, which dealt with folded three-way array, the MCR-ALS methodology could work with the original time series Sentinel-2 data (without any temporal interpolation to compensate loss of information due to acquisition conditions during the season). By working with the original temporal data, it opens up the possibility of obtaining actual date-by-date information since no interpolation process is necessary.

The current Part 3.2 focuses on 4978 vineyard blocks in the LR region during the period from May to September 2019. The aim of this part is to determine whether the MCR-ALS method is a potential alternative to fit the temporal characteristics of the dataset, thus comparing and extending the knowledge gained in the previous part. Indeed, the year 2019 was characterised by the impact of an extreme weather event (heatwave) that affected the LR wine region at the end of June 2019, specifically from the 23rd of June until the 8th of July. The

unusual and extremely early summer heat that affected the LR wine region during the growing season may result in an alteration of most of the phenological stages of vineyards. Therefore, the second purpose of this part was to see if an MCR-ALS model may faithfully represent the study-case, i.e. the assessment of the temporal alteration related to a heatwave episode.

3.2.2 Materials and Methods

3.2.2.1 MCR-ALS method

MCR-ALS method (de Juan et al., 2019) has been chosen as the main strategy for a time series of multispectral images when no prior information is available (de Juan and Tauler, 2021). The MCR-ALS method is a resolution method oriented to recover the underlying concentration profiles, as well as the pure spectra profiles (signatures) of the constituents of the analysed samples and it is based on a simple bilinear model which assumes that the constituents of the analysed sample weighted according to their relevance are expressed by a simple matrix equation (Zhang and Tauler, 2013). Following de Juan and Tauler (2021) the equation could be written down for a mixture with components I = 1, 2, …, *n*, being *n* the sources of variation coexisting in the analysed samples (system) as (Equation 3.2):

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E} = \sum_{\mathbf{i}=\mathbf{1}}^{\boldsymbol{n}} \mathbf{c}_{\mathbf{i}}\mathbf{s}_{\mathbf{i}}^{\mathbf{T}} + \mathbf{E} \qquad \text{(EQ 3.2)}$$

where **D** is the original data set measurements that come from *n* sources of variation. Data **matrix C** (*n* columns) contains the related proportion profiles (concentrations) of the **matrix** $\mathbf{S}^{\mathbf{T}}$ (*n* rows) that contains the qualitative profiles (pure spectra of the constituents) of the individual source of variation. **E** is the matrix associated with noise or experimental error, i.e. variation unexplained by the model. The MCR model using a constrained Alternating Least Squares algorithm (ALS) is explained by de Juan and Tauler (2021) as a method that works with initial estimates of the full matrix **C** or $\mathbf{S}^{\mathbf{T}}$ and, at each iteration, calculate both matrices of the bilinear model under the action of appropriate constraints. When the reproduction of the original data set **D** by the model $\mathbf{C}\mathbf{S}^{\mathbf{T}}$ is good enough, the optimisation stops.

3.2.2.2 Case-study and Remote sensing data

MCR-ALS was applied to a data set of 4978 samples (vineyard blocks) extracted from the graphical parcel register of France (RPG) from the large wine-growing region of LR as in Section 3.1.2.3 but only for the 2019 year. Sentinel-2 L2A data containing the study vineyards

were selected and processed via the Google Earth Engine (GEE) platform from May to August (Section 3.1.2.3, remote sensing data).

3.2.2.3 Modelling

Data construction: multiset structure

The simultaneous analysis of multiple data sets with information in common is possible with MCR-ALS (Tauler et al., 2020). The possibility to simultaneously analyse multiple datasets implies that they share at least some parts of their data variance, e.g. some chemical components (Tauler et al., 2020). Due to the nature of the Sentinel-2 satellite images, a multiset structure that has the spectral dimension (12 wavelengths) as shared information between the different time series data has been proposed. Given this multiset structure, the MCR-ALS methodology allows handling the temporal information gaps, which implies that the temporal interpolation process as in PARAFAC method, is not necessary. This is because by sharing the spectral dimension (12 wavelengths), the actual reflectance data can be used for each of the individual vine blocks according to their satellite revisit times. Therefore, the measured reflectance intensity is not structured as a cube, but as data matrices **D** (x × y, λ), as a function of two variables: pixels position x-y and spectral wavelengths for each revisit time of Sentinel-2 satellites. As a result, there is a 2-D matrix for each date. In order to adapt the data to the application of the method, it has been necessary to implement the following steps:

- Create a grid of 1000 × 1000 with the boundary of the LR region. This grid allows to emulate pixels as in a real 'image' containing two spatial dimensions (x and y pixel position). The size of the grid was decided in advance by considering the optimal size of the kriging interpolation grid (as applied with the PARAFAC methodology).
- For each pixel, link the spectral and temporal data by vineyard blocks for the year 2019 from their available geographic coordinates with the pixel's position. The number of blocks for each pixel was 2.8 on average, being 1.23 the standard deviation of the established values;
- Calculate the average of the spectral values from vineyard blocks per pixel and per date. That is, each pixel contains only the information of several vineyard blocks (2.8 block on average). It should be noted that with this third step, it is assumed that the information of nearby vine blocks in space will be similar.

At the end of these steps, the dataset was meaningfully organised into a multiple dataset (25 matrices representing satellite revisit date) sharing the spectral dimension (12 wavelengths, λ) for 3213 pixels. Therefore, each of these matrices represented the conceptual idea of an image

(x-y pixel position) according to the satellite revisit time. Figure 3.10 summarises the workflow of the database reconstruction.

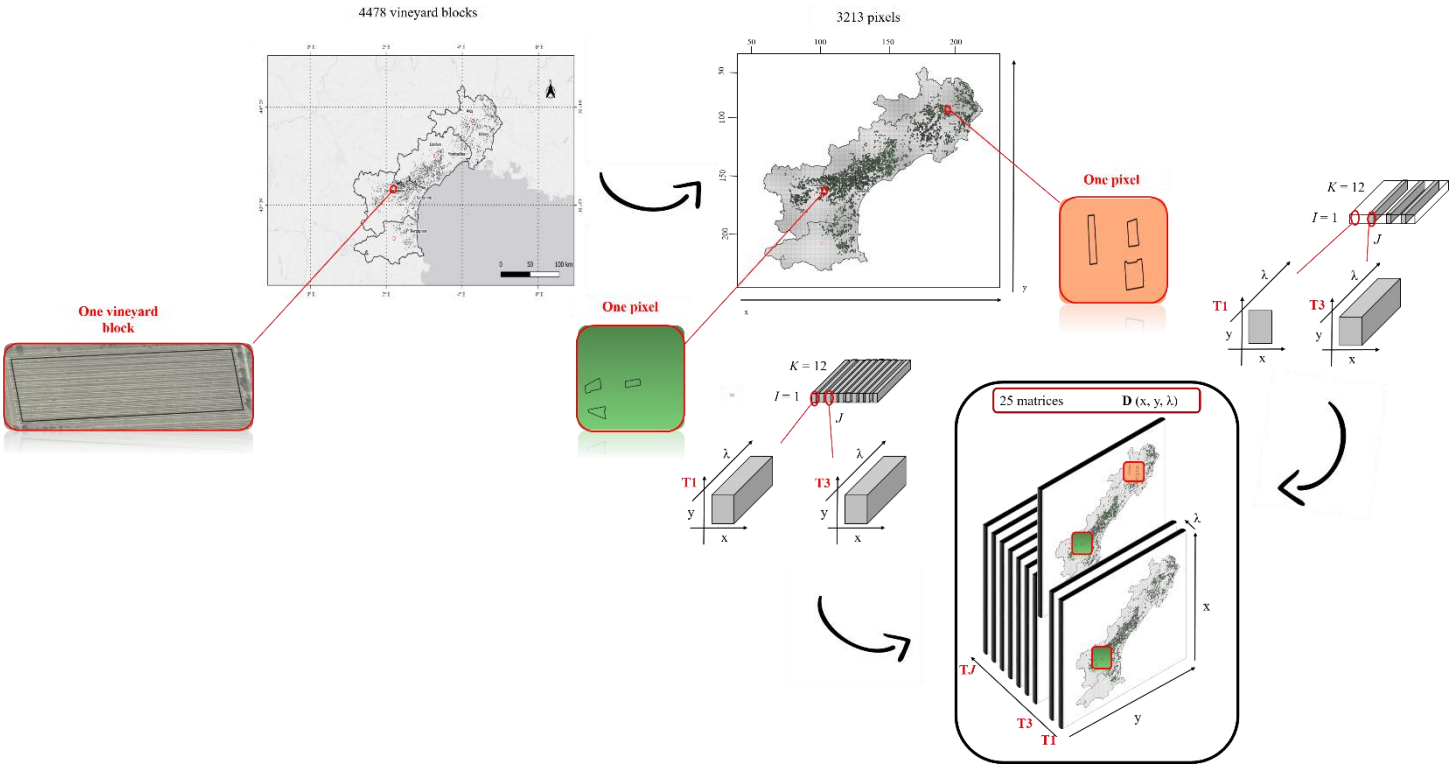


FIGURE 3. 10 - Workflow diagram of data structuring in order to apply the MCR-ALS model to deal with temporal information gaps.

MCR-ALS model

The MCR-ALS method was used to decompose the time series data of Sentinel-2 image into the signatures or pure spectra of the image constituents and into their concentration (relative amounts) on the image (distribution map). The first step is to unfold the three-dimensional data cubes (x × y × λ) as a 2-D data matrix (x × y) × λ, ready for MCR-ALS multiset analysis (Figure 3.11). Then, for an accurate reproduction of the original data (**D**), the **C** and $\mathbf{S^T}$ were estimated and then optimised iteratively in an Alternative Least Square (ALS) algorithm until convergence was reached (Zhang, 2015). The rows of this matrix **D** ((x × y) × λ) contain the original spectral measurements for every pixel and every date. Matrix **C** ((x × y) × *n*) contains the concentrations or relative amounts of the constituents in the pixels and $\mathbf{S^T}$ matrix (*n* × λ) is the pure spectra associated with these constituents at λ spectral bands (Figure 3.11). In order to have physically meaningful spectral profile shapes, non-negativity constraints in **C** and $\mathbf{S^T}$ were applied.

For determining the total number of the component model for adequate resolution of the multiset data, the percentage of variance explained (r^2) as well as the meaningfulness, in the authors' opinion, of the resulting pure spectral profiles were taken into account.

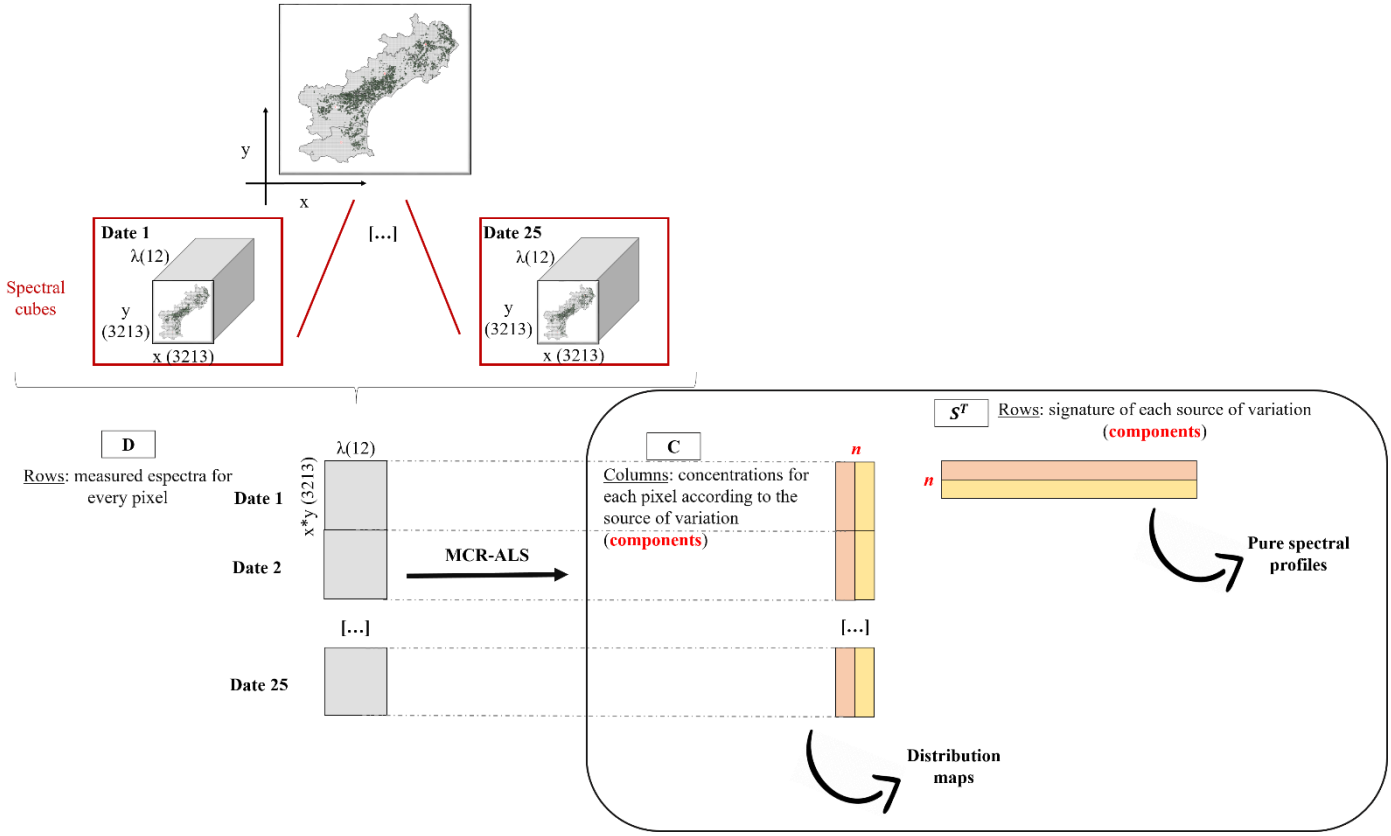


FIGURE 3. 11 - Multivariate Curve Resolution and alternating Least Squares (MCR-ALS) analysis of the multiset data from the Sentinel-2 satellites.

3.2.3.4 Spatial analysis

Semivariograms were performed using GeoFis 1.0 software (Leroux et al., 2018). This software was used to derive semivariogram models and estimate their featured parameters, i.e. C_0 (nugget effect), C_1 (sill) and A_1 (range) and the Cambardella Index (I_c) (Cambardella et al., 1994). The Cambardella index was considered here to demonstrate the interest of quantifying how the 2-D image (distribution map) of the concentrations (relative amounts) of every component were spatially organised in the LR region for some dates as an illustration of the exploratory potential of the MCR-ALS methodology in the context of viticulture at large scales. The hypothesis of this part was the same as in Section 3.1.3.3 of the previous part, where it has been considered that if spatially structured patterns are revealed, it is assumed they are determined by environmental variables (soil, climate, etc.) that are spatially structured.

3.2.3 Results

3.2.3.1 Data signal decomposition

The purpose of this part was to obtain a model that faithfully represented the problem under study, i.e. the assessment of the temporal alteration related to a heatwave episode in the samples (3213 pixels) at a regional scale in 2019. For the resolution of this dataset, the two-component model was considered the most appropriate. Although the percentage of variance explained (r^2) of the three-component model (r^2 = 99.63) was slightly higher than for the two-component model (r^2= 99.53), the profiles of the former one were not interpretable from the author's perspective. Therefore, for two-model components, the profiles of the components were identified as spectra of the soil (Co1) and of the vegetation (Co2) (Figure 3.12).

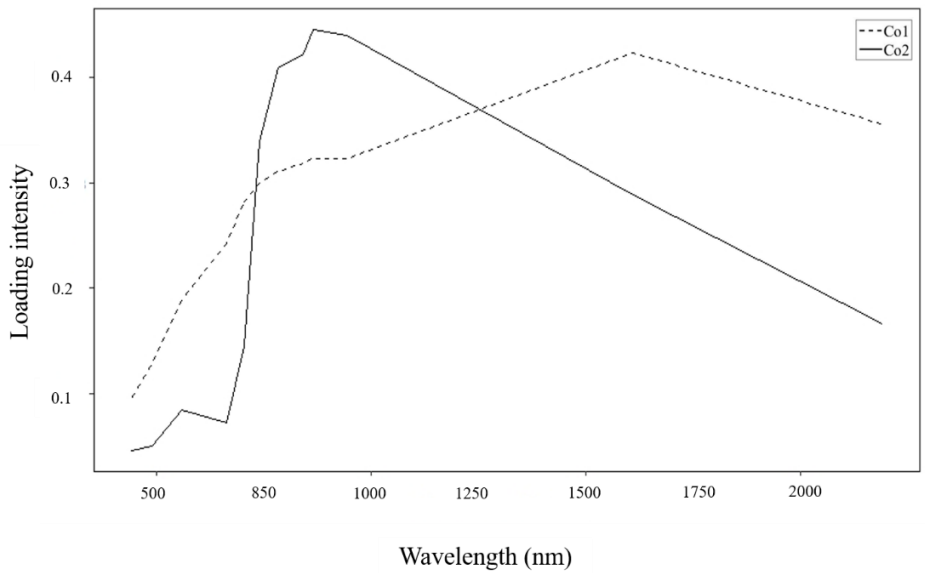


FIGURE 3. 12 - Spectral loading for two-component MCR-ALS model for the year 2019 (Co1 and Co2).

3.2.3.2 Spatial analysis and characterisation

Spatial analysis

The MCR-ALS methodology made it possible to consider each date of the study period collected by the Sentinel-2 satellite. Via 2-D reconstructed images of the concentration values at a pixel level in the *x–y* plane, 25 date distribution maps were obtained. However, the

reconstitution of a nearly complete distribution map containing the information of at least 3/4 of the LR pixels (≈2500 pixels) was possible on 8 dates over the study period (1st May to 31st August).

Figure 3.13 shows the eight chronologically ordered distribution maps of the year 2019 relative to the estimated concentration value of the Co1 component (soil) of each pixel in relation to its spectral profile in each time slice at the level of the LR vine-growing region. Therefore, each distribution map for Co1 highlights the evolution of pixels whose contributions were directly related to the profile identified as the soil in 2019. According to Figure 3.13, the distribution maps from the 7th July to the 22nd July represent the dates with the higher concentration (yellowish colour trend) and from the beginning of August, the dates of lower concentration (bluish colour trend). Therefore, the spatial evolution of the Co1 contribution over the study period can be observed on a regional scale from the concentration values using the distribution maps. The concentration values of the Co1 varied spatially across the eight distribution maps from May to August. A peculiar zone is observed in the pixel's x = [150, 200] y= [50, 110] which seemed to follow a different trend than the rest of the region from early June until July 7th where the region seems to spatially homogenise the concentration values until the end of the study period.

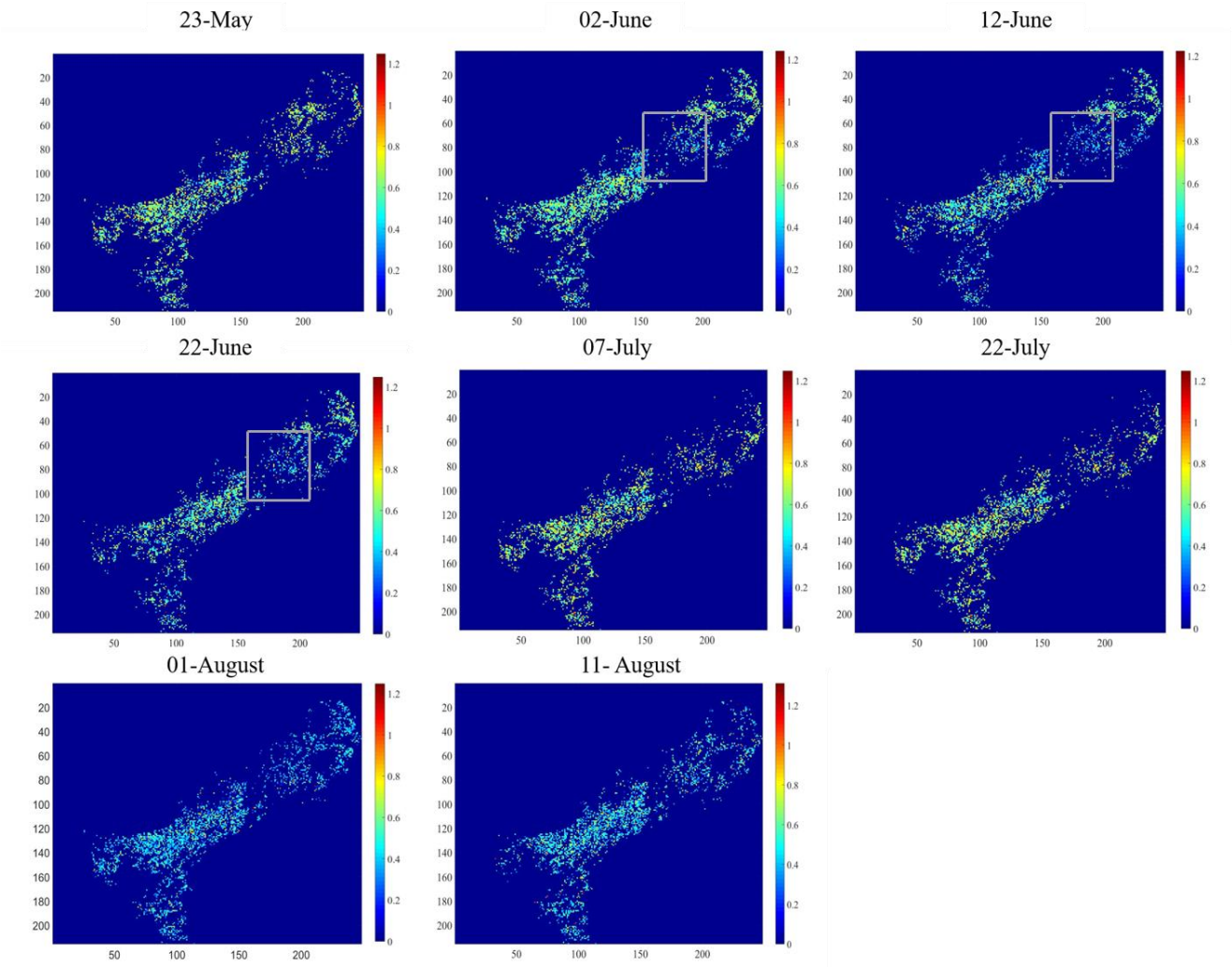


FIGURE 3. 13 - Distribution maps of the matrix concentration values for pixels in the vine-growing region at regional scale for component 1 (Co1) identified as the spectral profile of the soil in the year 2019 during the dates covering the vine growing season. Concentration colour gradient ranges from the lowest values (0) in dark blue to the highest values (1.2) in red. The grey square highlights the peculiar zone observed in the pixel's x = [150, 200] y= [50, 110].

Figure 3.14 shows the same eight chronologically ordered distribution maps of the year 2019 but relative to the estimated concentration value of the Co2 component (vegetation) of each pixel in relation to its spectral profile in each time slice at the level of the LR vine-growing region. Each map of Co2 highlights pixels whose contributions were directly related to the profile identified as that of the vegetation in 2019. The maps from 12 June to 22 July show the highest concentration, i.e. the highest presence of vegetation. However, in the case of this second component, there appears to be a greater spatial disparity between the distribution maps. For example, between the images of June 22 and 7-22 July, the high concentrations (yellowish to reddish colours) appear more clearly in the northern part of the region for the first date,

whereas for the second and third date, they seem to have migrated towards the southern part of the region. This dynamic could be related to the heatwave phenomenon that occurred in the LR region at the end of June 2019, specifically from the 23rd of June until the 8th of July temperatures of up to 45 °C were recorded. However, it is worth placing these results in line with those obtained in the previous part (3.1) of this chapter. In fact, in the previous part it has been shown through agronomic knowledge by experts, that other factors such as different phenological stages of vineyards (e.g. grape variety), leading to different vigour, management practices, etc., are factors that could have an important role in explaining the different spatial patterns (concentration differences) on the distribution map over time.

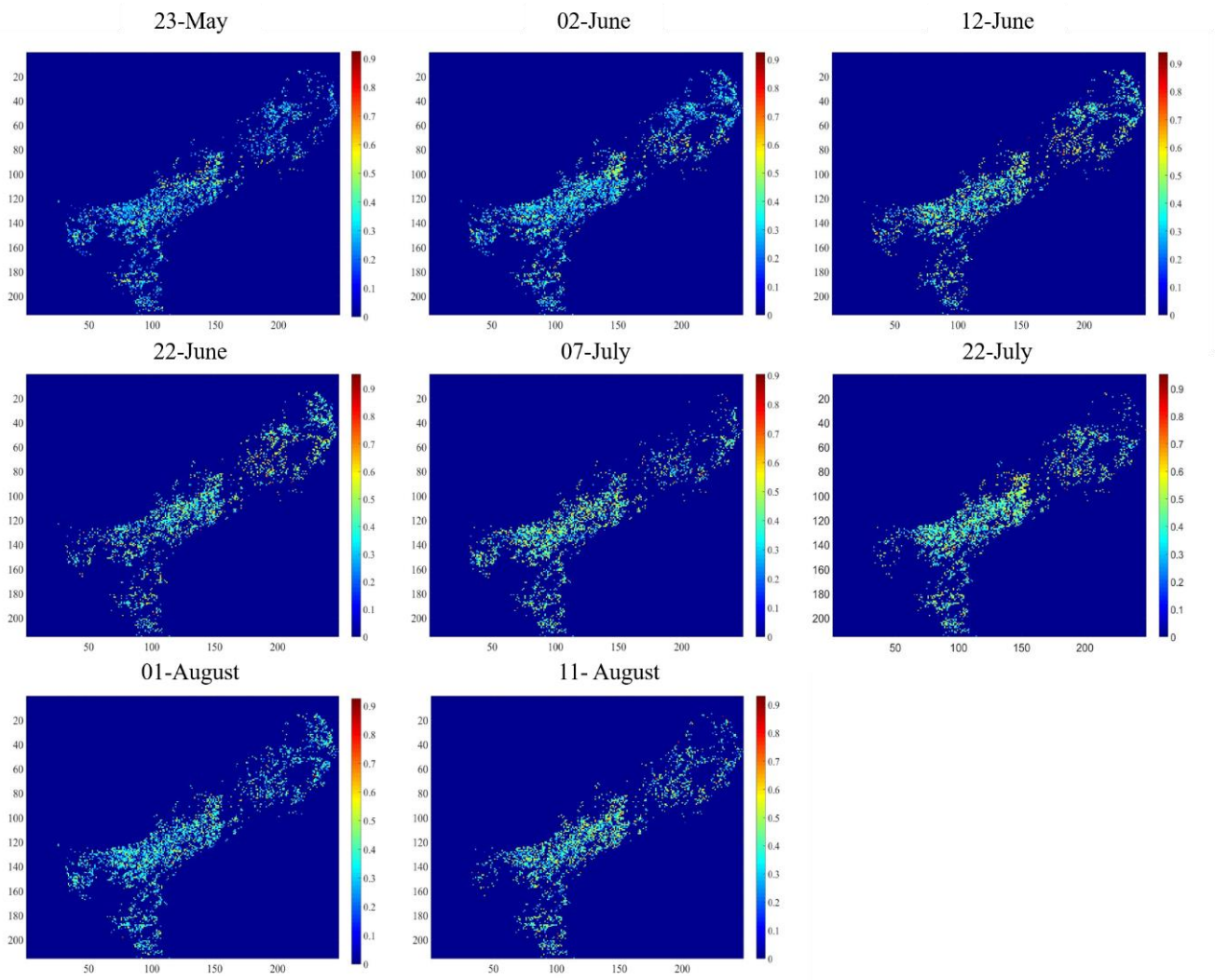


FIGURE 3. 14 - Maps of concentration values for pixels in the vine-growing region at regional scale for component 2 (Co2) identified as the spectral profile of the vegetation in the year 2019 for the 8 dates covering the vine growing season. Concentration colour gradient ranges from the lowest (0) in dark blue to the highest values (0.9) in red.

Geostatistical characterisation

The spatial variability of vine growth at the LR scale was shown visually by eight distribution maps (Figure 3.13 and Figure 3.14). The distribution maps revealed a high spatial and temporal variability at the scale of the study area. Using geostatistical methods, a characterisation of this high spatial and temporal heterogeneity was approached. For illustrative purposes, to demonstrate the potential of the method to characterise temporal dynamics, the spatial variability of Co2 (vegetative component) at regional scale was chosen to be characterised in the first (23-May) and last (11-August) maps with the Cambardella index (I_c) (Table 3.6). The different spatial organisation of the concentration values over time was confirmed by the semivariogram models, which showed that in May, about 50-60% of the variability was explained by a spatial phenomenon (I_c = 45%), while in August the variability explained by a spatial phenomenon decreased significantly (Ic = 88%). This significant spatio-temporal heterogeneity could be due to the great impact on the vegetative development of certain factors at the beginning of the growing season (May), such as (i) the different vineyard management such as the training systems and weeding practices (ii) different vine precocity and (iii) topographical characteristics (exposure, slope) but which become less decisive at the end of the season, i.e. this spatial variability in vine growth becomes more homogeneous.

TABLE 3. 6 - Semivariogram parameters and spatial variability index of vegetation component (Co2) concentration values for late May and early August. A1 (Range), C0 (Nugget), C1 (Sill) and Ic (Cambardella Index).

Date	Components	Semivariogram model	A_1 (km)	C_0	C_1	I_c (%)
23-May	Co2	Exponential	4e+06	0.009	0.011	45
11-August	Co2	Exponential	1e+06	0.013	0.001	88

3.2.4 Discussion

The application of the MCR-ALS algorithm allowed the study of remote sensing time series images for the regional characterisation of vineyard blocks. The MCR-ALS methodology applied in a non-supervised way at regional level requires a high level of agronomic knowledge and expertise able to integrate and interpret (i) local specificities (management, varieties, regulatory constraints, production objectives, etc.), (ii) climatic characteristics of the current year and (iii) knowledge of the plant, the soil and their interactions. Therefore, this part does not go beyond the presentation of an interesting multi-way methodology to take into account a different temporal approach than the one shown so far. However, the MCR-ALS approach is original from an exploratory point of view because it proposes a model that:

- allows to represent the spatial variability of the components (soil-vegetation) for the revisit dates of the Sentinel-2 satellites, i.e. no temporal interpolation needed,
- can be easy to visualise spatially as it deals with two spatial dimensions (x-y pixels), creating 'images' (distribution maps) directly. Nevertheless, the distribution maps shown in the results are difficult to interpret visually due to the spatial resolution at which they have been analysed (regional scale). At these larger scales, perhaps a kriging process (as in the preceding part, 3.1) could be considered to provide more visibility for interpreting the maps obtained.

In general, the interpretation of the distribution maps is enhanced by the fact that the approach gives, at each date, and for each pixel, the concentration of the two components (soil-vegetation). However, the MCR-ALS approach has operational limitations mainly related to the differences in data manipulation between remote sensing and chemometrics domains. The process of characterising the LR region as a single image has raised many questions outside the research scope presented in this part, e.g. how to create a relevant image for further analysis in chemometrics when dealing with large-scale remote sensing images. The proposed study is based on a pixel size of 1000 m * 1000 m containing a variable number of vineyard blocks depending on their geographical location in the LR region. The size of the pixels, as well as the randomness (quantity and qualities) of the type of blocks that can be present in the same pixel is clearly a limiting factor for the correct interpretation of the results at a regional scale. However, there is still no clear definition of a functional scale for working with satellite images at regional scale. Geostatistics could provide interesting methods to optimally define the size of the grid in order to minimise random (erratic) variability while maximising the relevant information (spatially organised) (Tisseyre et al., 2018). Another complementary approach could be to use the same methodology on at different resolution scales, i.e. different sample footprints (Sentinel-2 pixel, plot, domain…) to study the sensitivity of the method and thus obtain more relevant results from an agronomic point of view. The MCR-ALS approach presented here represents a specific application case for the LR wine region in 2019 with 8 maps showing temporal and spatial dynamics of vegetation and soil profiles from late May to early August. In view of the results, this approach should be further investigated, as it is able to characterise the variability of phenomena that evolve over time, such as canopy growth. The introduction of the spectral-temporal-spatial dimensions in the same methodology is a complex issue to address in remote sensing, hence the need to improve the adaptability of this approach to other research areas besides chemometrics.

3.2.5 Conclusion

As a reminder, the objective of this part was to test the MCR-ALS method on our temporal series of images in viticulture to verify if some of the limitations identified by the

PARAFAC method could be overcome. The work was not as comprehensive as with the PARAFAC method in that the validation of the agronomic information highlighted was not studied in the same depth. The work demonstrated the potential of a data resolution methodology with a suitable temporal approach, as presented with the MCR-ALS methodology, for the analysis of remotely sensed images time series. The results obtained from remote sensing data collected in 2019 over 4978 vineyard blocks showed that the MCR-ALS method has a potential to provide information on temporal variability coupled with spatial variability for the characterisation of the LR wine region. In fact, showing the possibility of working with real data without the need to interpolate (temporally) the spectral data present a great advantage in processing time series of multispectral images at the regional scale. However, due to its large-scale geographical area of application, its validation requires expertise and experience that is not currently available at this scale. As it was not intended to go beyond the presentation of the methodology as an alternative temporal approach to PARAFAC method, validation through the spatial index (I_c) showing that there is a change in the spatial structure of vegetation over the year 2019 was considered more than relevant to demonstrate the potential for applications.

Conclusion of Chapter 3

At the regional scale, time series of multispectral images can be very useful because of their high spatial and temporal resolution (Lobo et al., 2004). The objective of this third part was to establish the importance of accounting simultaneously for the temporal and the spectral dimensions to provide information for crop monitoring at a regional scale. To do this, the potential of multidirectional unsupervised methodologies capable of considering time as an additional variable in the analysis was presented. For the application of this temporal approach, a cube-shaped data structure was determined to allow the integration of the time series in its wholeness. Two different three-way resolution approaches were used to integrate the temporal dimension in the characterisation of the variability of vine phenology at a regional scale. Tisseyre (2012) determines that characterisation is the stage that allows the elaboration of agronomic information from one or several data obtained during observation. Thus, unsupervised characterisation using time series of multispectral data is considered a relevant exploratory approach due to the difficulty of obtaining complete ground truth data at the regional scale.

The Part 3.1 aims to highlight the interest of a multi-way approach such as the PARAFAC method in order to simultaneously account for spectral and temporal information from remote sensing data in agriculture assuming trilinearity of the data. The results demonstrated the interest of this type of approach to provide insight and knowledge that can be taken into account in subsequent analyses. The intention was to highlight how enriching, exploratory studies considering the inclusion of the temporal dimension, without neglecting the clear limits they entail, such as the scarce knowledge of large-scale interpretative specificities. Thus, the spatial patterns observed could represent relationships between spectral and temporal information that were relevant to the profession, which made clear the need for external validation by experts to demonstrate the interest of this type of methodology at large scales. Having addressed the possible limitations of the PARAFAC approach when dealing with large-scale image time series without a complete ground truth data set to validate all the potential relevant information provided, another multidirectional approach called MCR-ALS is presented in Part 3.2. The MCR-ALS method allows to consider simultaneously spatial (x-y pixel) and spectral dimensions with a multiset data and a non-continuous temporal approach. The aim of this Part 3.2 did not go beyond determining whether the MCR-ALS method is a potential alternative to fit the temporal characteristics of the dataset. The results of these two parts clearly show the importance of accounting for the temporal dimension along with the spectral dimension in crop monitoring because of its impact on the possible information obtained. Finally, the proposed approaches are sufficiently generic to consider their application (transfer) to other spatial scales and resolutions, noting that there is still work to be done in

adapting data management between the domains of remote sensing and chemometrics to exploit their full potential.

The difficulty of having a complete ground truth data set able to validate the potential information provided by the analysis of time series of multispectral images is a real issue. Indeed, to would require to have all the agronomic information on the crop and its evolution, soil, management practices, biotic and abiotic stresses, etc. This type of information may be available at the vineyard scale for research issues. However, at this scale, the variability of agronomic information is limited compared to the regional scale. This limits the interest to explore the potential of agronomic information that can be extracted from temporal image series. This issue is important because it will repeat with the development of new remote sensing acquisition platforms. In this context, it is necessary to develop new approaches, the use of non-supervised methods validated by experts capable of integrating a systemic vision of the crop at the regional scale, as proposed in this study, seems a relevant possibility.

When some agronomic observations are available at the regional scale, depending on the size of the data set and its spatial resolution another possible approach is to propose supervised methods. The next chapter (Chapter 4) relies on demonstrating the potential of a supervised multiway regression method using multispectral time series data applied to regional crop monitoring.

Chapter 4. Potential for regional crop monitoring with a supervised multiway regression method using multispectral time series data.

Introduction of Chapter 4

The previous chapter (Chapter 3) focused on the exploration of remotely sensed time series data from two different unsupervised multi-way chemometric methods: PARAFAC and MCR-ALS. Both have proven to be a convenient approach to improve the large-scale understanding of different patterns of vineyard cropping behaviour by integrating their temporal dynamics. Moreover, the relevance of integrating temporal dynamics of crop surface reflectance into the analysis was validated at a regional scale from a framework of expert observations.

This chapter moves from an exploratory analysis to an analysis with a well-defined target variable. As shown above, at large study scales, it is often difficult to measure all the data needed to apply a descriptive model. However, when it is possible to measure some variables, the characterisation of an agronomic phenomenon from these measured data can take different forms. In this context, one possible framework is to propose a supervised approach, such as a predictive model, to estimate agronomic information from one or several measured data types. Here, a supervised method is used to consider the question of the possible multivariate response of individual vineyards to an extreme weather (stress) event at the regional scale, based on satellite time series. To assess the impact of stress factors, capturing the anomaly state of vegetation is of great significance (Hua et al., 2019). Therefore, many studies have used VIs to monitor vegetation conditions because of their simplicity and reasonable predictive efficiency (Cogato et al., 2019b). However, as demonstrated so far, there is an interest in integrating the temporal factor to provide more detailed information on vegetation response in order to obtain a more accurate cause-effect response relationship. Since remote sensing data can be meaningfully organised into a multidimensional structure (e.g. a cube), the hypothesis is that monitoring the evolution of certain phenomena, such as vegetative growth, and considering simultaneously the spectral and temporal dimension, could have the potential to reveal a stress situation over time scales, as the stress trigger and its effect will vary considerably over time.

Cube data structures require multi-way chemometrics analyses that are capable of handling the trilinear structure in the data (fold method) to perform the supervised analysis on remotely sensed data. An extension of the ordinary regression model PLS, the N-way Partial Least Squares regression (N-PLS) (Bro, 1996) is described and applied to effectively capture the causal relationships between a crop-specific driver response, i.e. a heatwave (driver) and estimated loss of vineyard yield (response). To illustrate the importance of understanding the temporal dynamics of a crop's spectral response to underlying variations, such as environmental factors, an analysis that: (i) deals with high-dimensional data, taking advantage of any multi-way structure in the data (Smilde, 1997); (ii) uses all temporal data and (iii) avoids multicollinearity issues that are inherent to multiple linear regressions (Inoue et al., 2012) is needed. The overall objective of this chapter is to develop a trilinear model to characterise the

temporal dynamics of the vegetation response to a heatwave at large scales. Therefore, based on a case study of the impact of a heatwave at the end of June 2019 on vineyard yields in the Languedoc-Roussillon region (south of France), the first part of the chapter (4.1) presents the potential of the N-PLS methodology as a useful modelling technique for the analysis of multispectral image time series and the second part of the chapter (4.2) shows the possible application of the model developed in the previous chapter to spectrally, temporally and spatially assess the impact of an extreme weather event on yield, as spatial and temporal dynamics are interdependent with regard to disruptive and extreme climate conditions at large scales (Neethling et al., 2019).

4.1 Potential of Multiway PLS (N-PLS) Regression Method to Analyse Time Series of Multispectral Images: A Case Study in Agriculture

Article 3: Potential of Multiway PLS (N-PLS) Regression Method to Analyse Time-Series of Multispectral Images: A Case Study in Agriculture

Published in *Remote Sensing*, 2022, *14* (1), pp.216. DOI: 10.3390/rs14010216

E. Fornieles-Lopez[1,2], G. Brunel[1], R. Florian[3], B. Gaci[1,2], M. Metz[1,2], N. Devaux[4], J. Taylor[1] B. Tisseyre[1] and J.M. Roger[1,2]

[1]ITAP, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France

[2]Chemhouse Research Group, 34000, Montpellier, France

[3]CNRS, IMS, nº 5218, Groupe Signal et Image, Univ. Bordeaux, F-33405 Talence, France

[4]LISAH, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France

Abstract: Recent literature reflects the substantial progress in combining spatial, temporal and spectral capacities for remote sensing applications. As a result, new issues are arising, such as the need for methodologies that can process simultaneously the different dimensions of satellite information. This paper presents PLS regression extended to three-way data in order to integrate multi-wavelengths as variables measured at several dates (time-series) and locations with Sentinel-2 at a regional scale. Considering that the multi-collinearity problem is present in remote sensing time-series to estimate one response variable and that the dataset is multidimensional, a N-way Partial Least Squares (N-PLS) regression approach may be relevant to relate image information to ground variables of interest. N-PLS is an extension of the ordinary PLS regression algorithm where the bilinear model of predictors is replaced by a multilinear model. This paper presents a case study within the context of agriculture, conducted on a time-series of Sentinel-2 images covering regional scale scenes of southern France impacted by the heatwave episode that occurred on 28th June, 2019. The model has been developed based on available heatwave impact data for 107 vineyard blocks in the Languedoc-Roussillon region and multispectral time-series predictor data for the period from May to August 2019. The results validated the effectiveness of the proposed N-PLS method in estimating yield loss from spectral and temporal attributes. The performance of the model was evaluated by the R^2 obtained on the prediction set (0.661), and the Root Mean Square of Error (RMSE), which was 10.7 %. Limitations of the approach when dealing with time-series of large-scale images which represent a source of challenges are discussed; however, the N–PLS

regression seems to be a suitable choice for analysing complex multispectral imagery data with different spectral domains and with a clear temporal evolution, such as an extreme weather event.

Keywords: fold methods; chemometrics; Sentinel-2; multispectral remote sensing

4.1.1 Introduction

Benefiting from the large number of high-quality spectral-temporal images captured from Earth observation satellites, remote-sensing based on multitemporal and multispectral imagery has a great potential to provide timely and comprehensive agricultural and environmental information. With the availability of more and more remote sensing platforms, massive amounts of remotely sensed data are produced (Albanwan & Qin, 2018); for example, every 5 days, the Sentinel-2 multispectral satellite constellation can provide a global coverage of Earth's land surface. According to Bovolo and Bruzzone, (2015), new policies for the free distribution of satellites data, the distribution of archive data that makes large-scale retrospective analysis possible, and the increasing number of satellites with higher revisit frequency, are the main reasons for the growing importance of a need for methodological research on multitemporal data analysis.

These large amounts of available remote sensing data have largely been applied to hazard assessment, coastal applications, agricultural applications, natural resource management, etc. (Mishra et al., 2021). In the field of agriculture, repetitive information on crop status throughout the season at different scales and for different actors, provides the capacity to assist the adaptive evolution of agricultural practices (Weiss et al., 2020). Chen et al. (2008) summarised six main applications of remote sensing in agriculture management and monitoring. For all of these applications, the information of interest consists of traits or features of the agricultural systems, and especially how these vary in space and time. Since plant growth and plant developmental processes are strongly influenced by fluctuations in environmental conditions, and especially ambient temperature (Venios et al., 2020), capturing the dynamics of crop growth over time, especially at critical growth stages is crucial. Physiological and physical properties of the production system, and consequently the reflectance spectrum, change according to growth conditions, and time of measurement (Filella et al., 1995). Consequently, remotely sensed data, obtained by satellites over the time, can provide a set of detailed data on plant growth and development (Plant et al., 2000).

Regardless of the applications, different approaches to deal with the increasing volumes of data and variables of interest can be found in the remote sensing literature. As Dorigo et al. (2007) state in their review, regression methods constitute a widely used tool for explanatory and prediction purposes. When it comes to producing estimates, a popular method to implement this prediction, is the Partial Least Squares method (PLS) (Dorigo et al., 2007). Numerous papers have previously discussed this method from a geometrical, mathematical and statistical point of view (Dayal & MacGregor, 1997; Phatak & Jong, 1997). PLS methods have been extensively used (Arenas-Garcia & Camps-Valls, 2007) as they are well suited to deal with multivariate data structures with high covariance and redundancy (Abdi & Williams, 2013), such as are found in remote sensing datasets.

However, challenges remain with respect to addressing these increased data volumes and increased data variability (Bishop, 2013). Exploiting multispectral image time-series is a promising but still relatively underexplored research direction due to the complexity of jointly analysing spatial, spectral and temporal information (Mou et al., 2019). The increasing complexity of information leads to a natural generalization of the concept of the 'dataset' from conventional tables to higher-dimensional arrays (Henrion, 1994). As Bro (1996) specified, when including time as an additional dimension, the dataset becomes three-way, which cannot be handled by commonly used conventional prediction models.

Among the alternatives proposed in the literature for modelling multi-way data, deep learning has the potential to outperform traditional classification and regression techniques (Li et al., 2019). However, deep-learning approaches often requires a large dataset to perform the learning and do not provide immediate access to intermediate steps to analyse and understand the agronomic processes that impact production. An interesting alternative to modelling multiway data has been proposed in the chemometrics literature, the N-way Partial Least Squares regression (N-PLS). This is a modelling strategy introduced by Bro (1996) and further advanced by Smilde (1997). N-PLS aims at building a model between a high-order array $\underline{\mathbf{X}}$ of independent variables and a response array $\mathbf{y}$ (Hanafi et al., 2015). In this paper, PLS regression is extended to multiway data (N-PLS), with the main emphasis on three-way data, by including time as a dimension. It is important to note that N-PLS imposes a trilinear structure on the data (Sena & Poppi, 2004) and when introducing time into the analysis, this implies that the emphasis is exclusively on the development of a correlational structure over time (Coppi, 1994), i.e. time will determine both the variability between individuals and the variability between variables (wavelengths). The N-PLS methodology allows a joint evaluation of time and variables, in order to determine the interaction between them, to know if and how the variables modify their 'behaviour' at different occasions. To the authors' knowledge, such an approach has not been proposed in the analysis of remotely sensed satellite imagery.

Therefore, the objectives of this study are (i) to propose a formalism to apply the N-PLS approach to a time-series of images at the regional scale in order to predict a variable of interest with a small ground truth dataset (ii) to show the value of the approach for its potential to generate knowledge in an agricultural case study, and (iii) to identify the possible limitations of the approach when dealing with image time-series on large scales with series that may be incomplete for certain areas.

The work is organised as follows: Section 4.1.2 introduces the proposed N-PLS method and the development of the model as well as the description of the case study that the methodology is applied to. The results are presented in Section 4.1.3, with the discussion in Section 4.1.4.

4.1.2 Materials and Methods

4.1.2.1 Type of problem the N-Way Partial Least Squares aim to address in remote sensing

The type of problem that the proposed methodology of this study addresses is the detection of a rapid extreme climate event (frost, hail, heatwave, etc.) that may affect production over a large area (e.g. a regional scale). It is assumed that the extreme climate event has an effect on production by affecting either the pigment content of the leaves, the biomass itself (growth limitation) or the physical architecture of the plants (Cogato et al., 2019b). Therefore, a time-series of multispectral satellite image should provide relevant information to detect the date of occurrence, the severity and also the spatial footprint of the extreme event under study. However, monitoring only a single variable, or only a few variables, is problematic as many variables are correlated and interrelated and have an effect on one another. Consequently, the use of a multidirectional (spectral and temporal) unfolding method (N-PLS) for a given extreme weather event (heatwave) that simultaneously examines all spectral information at different points in time should be well adapted to the challenge of anomaly and event detection in temporal data.

Considering imagery time-series, a fixed classical object/variable scenario, i.e. an agricultural field, can be observed several times under different conditions, yielding a separate data table for each condition (Henrion, 1994). Another option would be to group all the objects/variable conditions together taking into account the time-series of images, which leads to a consideration of a cubic or a three-dimensional data array as shown in Figure 4.1 Such an array has N images containing the corresponding objects (O) at T_n different dates, with each image being constituted of M_m spectral bands or wavelengths (variables). Given the data dimensionality, multiway arrays have been proven to be a natural and efficient representation of the data, in particular, tensor subspace learning methods have been shown to outperform their corresponding vector subspace methods, especially for small sample size problems (Zhao et al., 2011). Therefore, the N-PLS algorithm should constitute a relevant approach to retrieve the information contained in the spectral bands of satellite imagery to highlight the target phenomenon taking into account its effect on the temporal evolution of the imagery.

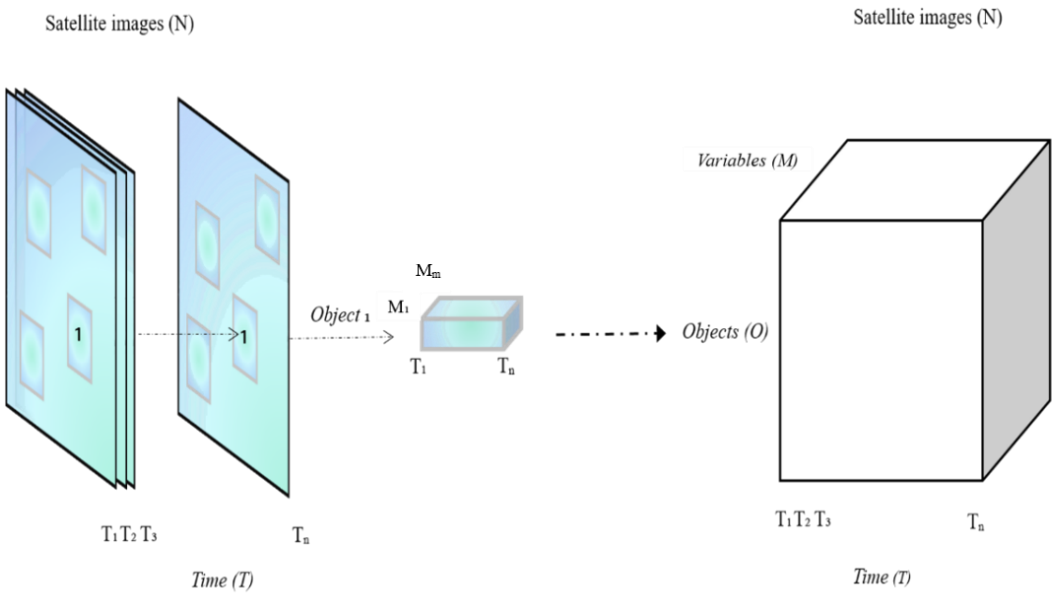


FIGURE 4. 1 - Production of three-dimensional array from satellite imagery time-series data.

4.1.2.2 N-Way Partial Least Squares

In general, the PLS approach is particularly useful when one or a set of dependent variables needs to be predicted by a (very) large set of predictor variables (or time-series) that are strongly cross-correlated (Abdi, 2010). The strength of N-PLS is that it summarises all latent information from a large 3-way dataset of object variables ($\underline{\mathbf{X}}$) and relates it to a dependent variable ($\mathbf{y}$) using a relatively low number of parameters, which makes the prediction more robust (Hansen et al., 2002). This situation leads to an analysis that is able to extract the maximum information possible from samples measured at different times based on cubic structure.

For these remote sensing data, the 3-PLS1 method (Bro, 1996) was considered here to generate a three-way array $\underline{\mathbf{X}} = ||X_{i,j,k}||$. The first of the three dimensions of cube $\underline{\mathbf{X}}$ corresponds to the objects (I), the second to time (J) and the third to spectral bands (K). Thus, the data were organised in a three-way array of independent variables $\underline{\mathbf{X}}$ ($I \times J \times K$) and a response vector $\mathbf{y}$ of size ($I \times 1$) that is defined by the objects dimensions (Figure 4.2).

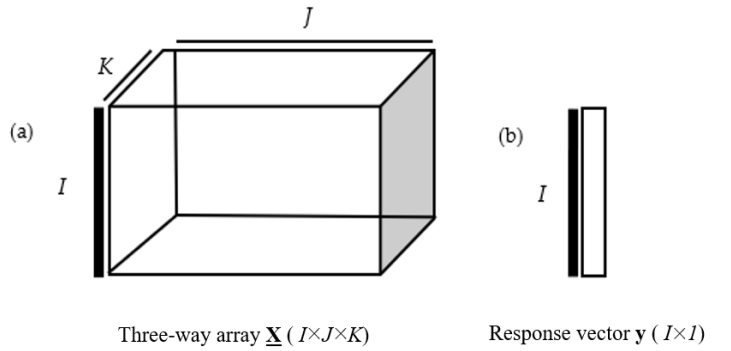


FIGURE 4. 2 - (a) Representation of three-way array (**X**) and (b) response vector **y** to compute a tri-linear PLS1 procedure.

The N–PLS regression was used to relate the three-way array **X** (remote sensing data) to the response vector y (ground truth data). As Abdi & Williams (2013) specified, PLS regression performs a simultaneous decomposition of **X** and **y** by means of a set of latent variables that explain as much as possible of the covariance between **X** and **y**. As Bergant and Kajfež-Bogataj (2005) detailed, if **X** (*I* × *JK*) is a properly unfolded two-way form of a three-way array **X** (*I* × *J* × *K*), the 3-PLS1 method can be written in mathematical form (Equation 4.1), where **X** and **y** are cantered along the first dimension, *I*.

$$\mathbf{S} = \mathbf{XW} \qquad \boldsymbol{y} = \mathbf{Sb} + \mathbf{r}. \tag{EQ. 4.1}$$

Estimating the weight array **W**, allowing for the scores in **S** to be expressed directly in terms of the **X**-columns, is the essential part of the method (Jong, 1998). The regression coefficients **b** (Equation 4.1) can be estimated afterwards by a least square approach (Bergant & Kajfež-Bogataj, 2005). The vector **r** in Equation 4.1 presents the part of **y** not explained by the model. For the estimation of **W**, the algorithm for three-way **X** and a single response y proposed by Jong (1998) was considered. If the initial values of $\boldsymbol{y_a}$ (*a* is the latent variable counter) are set to the original values $\boldsymbol{y}$ (a = 0 and $\boldsymbol{y_0} = \boldsymbol{y}$), the algorithm can be summarised as follows (Bergant & Kajfež-Bogataj, 2005):

1. Compute the reshaped covariance matrix $\check{\mathbf{Z}} = y_a^T\ \mathbf{X}(1 \times JK)$.
2. Define the first singular weight vectors $\boldsymbol{w}_a^J$ and $\boldsymbol{w}_a^K$ from $\boldsymbol{Z}$: $[\boldsymbol{w}_a^J, \boldsymbol{w}_a^K] = \text{svd}(\boldsymbol{Z},1)$. From hence, store them as additional columns in separate weight arrays $\mathrm{W}^J = [\boldsymbol{w}_1^J \ldots \boldsymbol{w}_a^J\ ]$ and $\mathrm{W}^K\ [\boldsymbol{w}_1^K \ldots \boldsymbol{w}_a^K\ ]$.
3. Compute the weight array by a Khatri-Rao product (Bro et al., 2001; Bergant & Kajfež-Bogataj, 2005) so that $\mathbf{W} = \mathbf{W}^K \otimes \mathbf{W}^J$. The Khatri-Rao product defined as a column-wise Kronecker product of the two matrices (Bergant & Kajfež-

Bogataj, 2005) was used to overcome constraints related to multiway arrays decomposition.

4. Calculate $\mathbf{S} = \mathbf{XW}$.
5. Calculate the regression coefficients regressing $\boldsymbol{y}$ on $\mathbf{S}$ as $b = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\boldsymbol{y}$.
6. Calculate the residual $r = \boldsymbol{y} - Sb$.
7. Increase a to a + 1 and continue from step 1 to the appropriate description of $\boldsymbol{y}$. The inclusion of an additional latent variable (a + 1) in the model is terminated when the joint analysis of RMSEC (Root Mean Square Error of Calibration) and RMSECV (Root Mean Square Error of Cross-Validation) (Goodarzi & Freitas, 2009) indicates overfitting due to sampling variability.

This brief description of the method highlights its interest for the case described in this article. It allows to a simultaneous consideration of the temporal (J) and spectral (K) components while keeping the information carried by these two components in the analysis.

4.1.2.3 Model description

Structuration of time-series data

In order to build a cubic data structure (objects × time × spectral bands), all spectral bands were re-interpolated on a date grid, regularly spaced over the period of interest, i.e. the period that the extreme climatic event may have impacted crops. The objective was to ensure that the cubic $\underline{\mathbf{X}}$, had the same number of steps (N days of measurements) to prevent temporal data gaps due to clouds and inconsistent numbers of available satellite images. The interpolation at a date t was performed band-by-band, by a convolution of the chronology measured with a Gaussian filter (Alam et al., 2008) centred on t and with full width at half maximum (P).

Temporally sparse satellite observations, especially for areas that can be affected by cloudy conditions, may not be sufficient for monitoring rapid changes such as some phenological crop stages (Wang et al., 2018). Moreover, Hird & Mcdermid (2009) have demonstrated that there is an impact of varying atmospheric conditions and sun-sensor-surface viewing geometries on satellite-derived time-series. A preliminary sensitivity analysis with a noise reduction approach and given time steps provided information on the sensitivity of the variation of the P and N parameters to interpolation. The parameters P and N involved in the interpolation setting were optimised by cross validation of 2 blocks repeated 5 times of a N-PLS between cube $\underline{\mathbf{X}}$ and vector $\mathbf{y}$ using the calibration dataset (see following section). For this analysis, values of P and N ranged from 10 to 50 and 5 to 30, respectively.

Calibration and validation of the model

Calibration and validation subsets were created for model assessment. Considering the size and the potential non-regularity of the distribution of the samples from the dependent variable, a calibration set (3/4) and a validation set (1/4) were defined by the distribution of **y**, as follows:

1. The vector y was sorted in ascending order.
2. After sorting, every fourth individual was placed in the validation set, the others retained in the calibration set.

The optimal number of latent variables was selected using a cross-validation of 2 blocks repeated 10 times of an N-PLS between the $\underline{\mathbf{X}}$ cube and the **y** from calibration set. As proposed by Bergant and Kajfež-Bogataj (2005), the final calibration fit can be written as in Equation 4.2.

$$\hat{\mathbf{y}} = \mathbf{Sb} = \mathbf{XWb} = \mathbf{XB_{NPLS}} + \mathbf{b_0} \quad \text{(EQ. 4.2)}$$

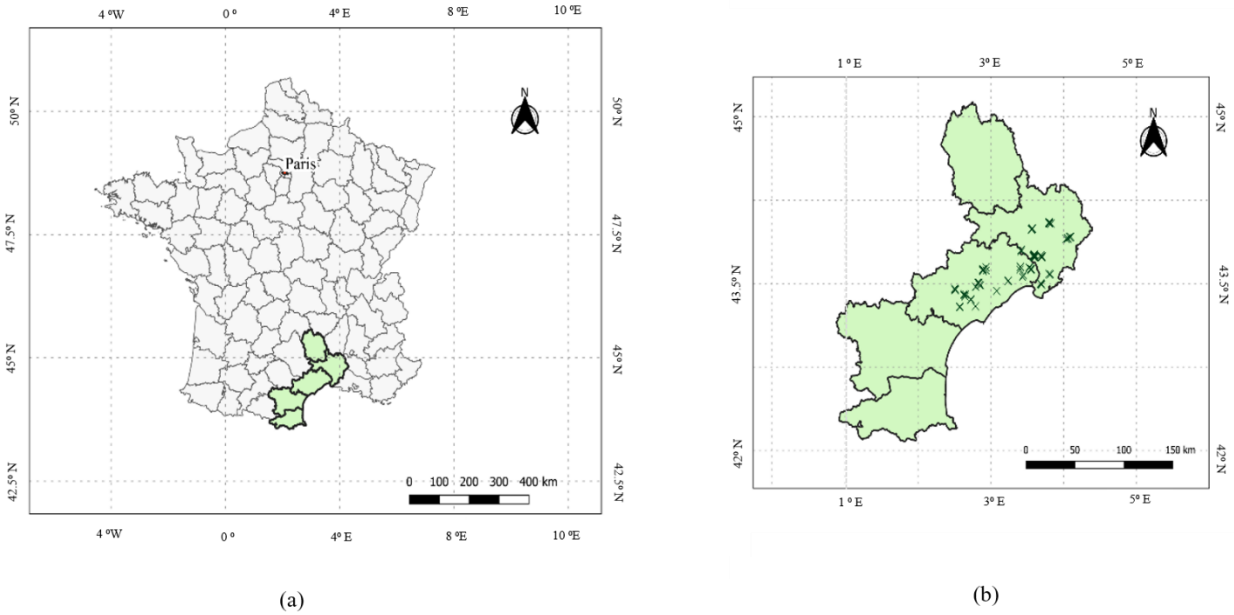
where $\hat{\mathbf{y}}$ is the vector of estimated responses and $\mathbf{b_0}$ the intercept of the linear regression model. The matrix of regression coefficients $\mathbf{B_{NPLS}} = \mathbf{Wb}$ (Equation 4.2) can be used on new data for the estimation of unknown response values. The prediction performance of the empirical model was quantified by the standard coefficient of determination (R^2), the bias and the standard error parameters (Ahmad Fadzillah et al., 2013; Malegori et al., 2017)

4.1.2.4 Case-Study

According to Cogato et al. (2019a), the increasing frequency of heatwave events represents a severe threat to viticulture because periods of extreme heat may affect grape yields and quality. Thus, the detection of the impact of heatwaves on perennial crops from dynamic spectral-temporal features using medium-resolution data, could provide valuable information to identify the possible effects of heatwaves on production and the spatial footprint of the phenomena at the regional/national scale. A preliminary study of eight heatwaves during the 2016–2017 and 2017–2018 growing seasons, showed that medium-resolution spectral data from Sentinel-2 time-series can provide valuable information on the possible effects of heatwaves on grapevines (Cogato et al., 2019b). The N-PLS algorithm should therefore constitute a relevant approach to retrieve the information contained in the spectral bands of the time-series that relates the target phenomenon (heatwave in this case).

Study area

The study area corresponded to a large wine-growing region, the Languedoc-Roussillon, located in the south of France (Figure 4.3). It extends over approximately 27,400 km2, from the Spanish border to the delta of the Rhône.



(a) (b)

FIGURE 4. 3 - Location of the study area in Southern France (a) and the location of the vineyards that contained the 107 blocks of interest for the study (b).

On Friday 28th June, 2019, a part of this wine territory experienced an extreme climatic event with a heatwave that reached 45 °C in some places (Figure 4.4). This heatwave occurred mid-season, that is to say at a critical stage in terms of vine growth.

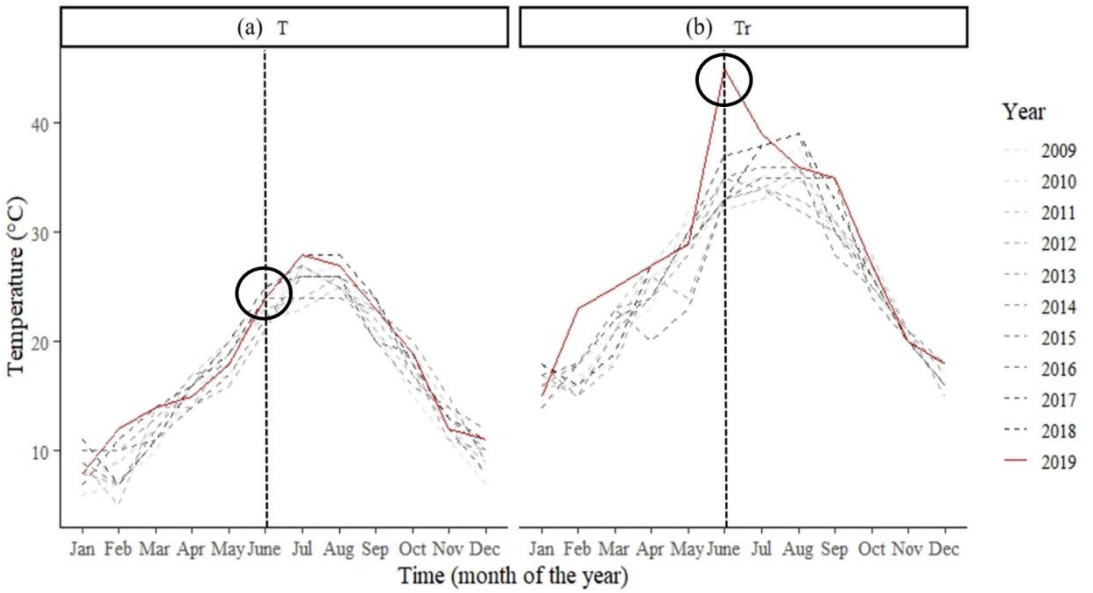
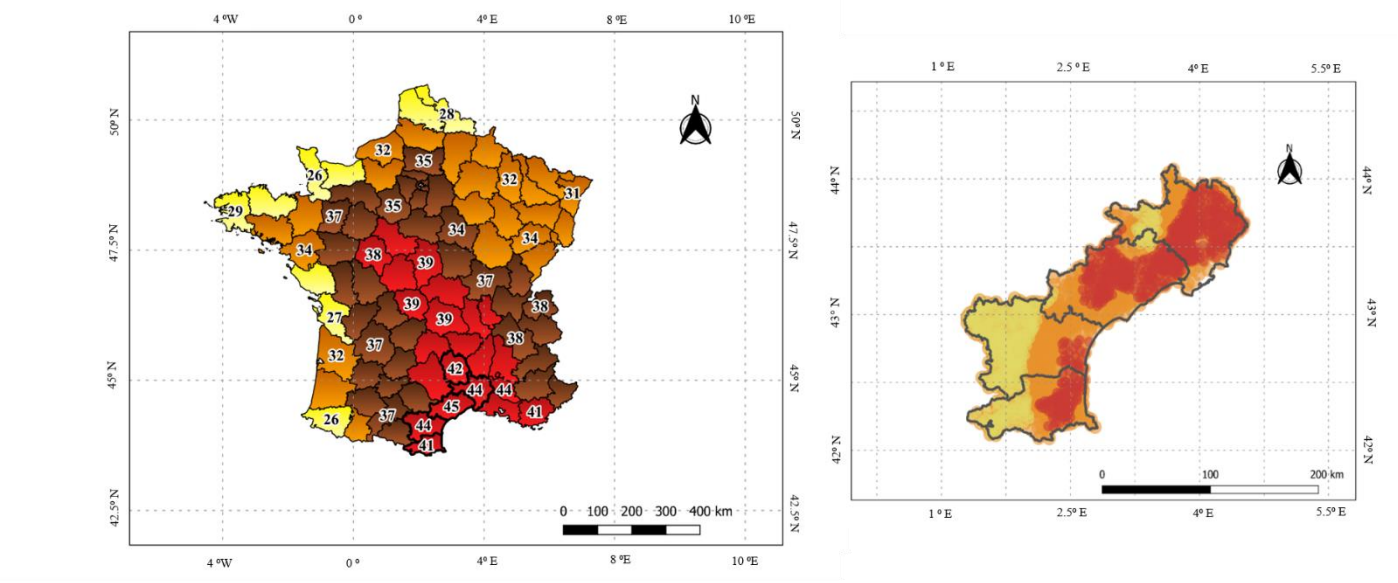


FIGURE 4. 4 - Mean monthly temperature (a) and maximum monthly temperature records (b) from 2009 to 2019 across the Languedoc-Roussillon region, France, highlighting a peak in the maximum monthly temperature corresponding to the extreme weather event that occurred in June 2019, while the mean monthly temperature for this month was not extreme. The vertical black dashed line highlights the month of the heatwave. Source: Historique de Météo-France.

It should also be noted that this extreme heatwave did not affect the entire region but only a part of it, as shown in Figure 4.5. In addition, the incidence of the heatwave may present a significant spatial variability according to the local elevation, aspect and slope of the individual blocks. The presence of natural features, such as forests or hedges, in the vicinity of the vineyards was also likely to mitigate the effects of the heatwave. This contributed to an observed local variability in the incidence of this climatic event on the vineyard blocks.



(a) (b)

FIGURE 4. 5 - (a) Maximum temperature record map 28th June 2019 France. (b) Maximum temperature record map 28th June 2019 Languedoc-Roussillon. Temperatures ≥ 40 °C are shown in red, temperatures < 40 °C and ≥ 30 °C are shown in orange and temperatures < 30 °C are shown in yellow. Source: Météo-France.

Remote sensing data

Data acquisition and pre-processing

The Sentinel-2 Multispectral Imager is a European Space Agency (ESA) satellite for Earth observation that provides imagery with 13 spectral bands from the near-infrared (NIR) to short-wave infrared (SWIR). The radiometric images are provided with a range of spatial resolutions of 10 m, 20 m and 60 m (Table 4.1). Each Sentinel-2A or -2B satellite revisits the same area every 10 days (5 days with the twin satellites together).

Images containing the study vineyard blocks (see Section 4.1.2.4) were selected and processed via the Google Earth Engine (GEE) platform that enables large-scale processing of Sentinel-2 L2A (Sentinel-2A and Sentinel-2B) products. The time period considered for the study was from 13th May to 20th August 2019, which is the most relevant period to monitor vine growth vegetation in this region (Devaux et al., 2019). Following Hollstein et al. (2016) decision trees and Bayesian models for cloud detection were applied to the Sentinel-2 images via GEE Javascript API to leave them in ready-to-use formats. Considering the revisit period of the Sentinel-2 (A/B) satellites, 25 images should have been potentially available over each block for the chosen time period. However, after cloud detection, the number of available images for each vineyard was 11 on average, with a range from 7 to 16 images depending on

the location of the various vineyards. By using the GEE cloud-based platform to select Sentinel-2 images and pre-process them, the output was a file containing the spectral band values for each date (available images) and for each vineyard block.

Spectral bands

There were 12 spectral bands (among the 13 available from Sentinel-2 satellites) used in the model (Table 4.1). Spectral band 10 at 1380 nm was not used as it corresponds to a high atmospheric absorption band dedicated to visible and sub-visible cirrus cloud detection (Hollstein et al., 2016). To avoid mixed pixels, a 10 m inner buffer was imposed at the border of each block to ensure only pure vine pixels were considered. Pure pixels within each block were then averaged for each spectral band at each individual date, i.e. regardless of vineyard block size, each block had one value for each spectral band at a given date.

TABLE 4. 1 - Spectral bands for the Sentinel-2 satellite considered by the analysis.

Sentinel-2 Band	Central Wavelength (nm)	Bandwidth (nm)	Spatial Resolution (m)
Band 1–Aerosol	442.7	21	60
Band 2–Blue	492.4	66	10
Band 3–Green	559.8	36	10
Band 4–Red	664.6	31	10
Band 5–Vegetation Red Edge	704.1	15	20
Band 6–Vegetation Red Edge	740.5	15	20
Band 7–Vegetation Red Edge	782.8	20	20
Band 8–NIR	832.8	106	10
Band 8A–Vegetation Red Edge	864.1	21	20
Band 9 –VNIR	945.1	20	60
Band 11–SWIR	1613.1	91	20
Band 12–SWIR	2202.4	175	20

Ground-truth data

The ground-truth data were obtained from 107 vineyard blocks in the Languedoc-Roussillon region (Figure 4.3). These were all blocks with characteristics representative of the Languedoc-Roussillon wine region in terms of plantation density, management practices and diversity of varieties. All blocks were non-irrigated. These blocks were selected because they all showed some effect related to the heatwave, such as stalled development, leaf burn and leaf drop.

The severity of this effect on each of the 107 vineyard blocks was assessed locally by the winegrowers and advisors. Severity was evaluated several weeks after by an estimation of percentage yield loss (Figure 4.6). There was an acknowledgement that it was sometimes difficult to attribute the losses to the heat alone. Figure 4.6 summarises the distribution of the 107 blocks in relation to estimated yield loss.

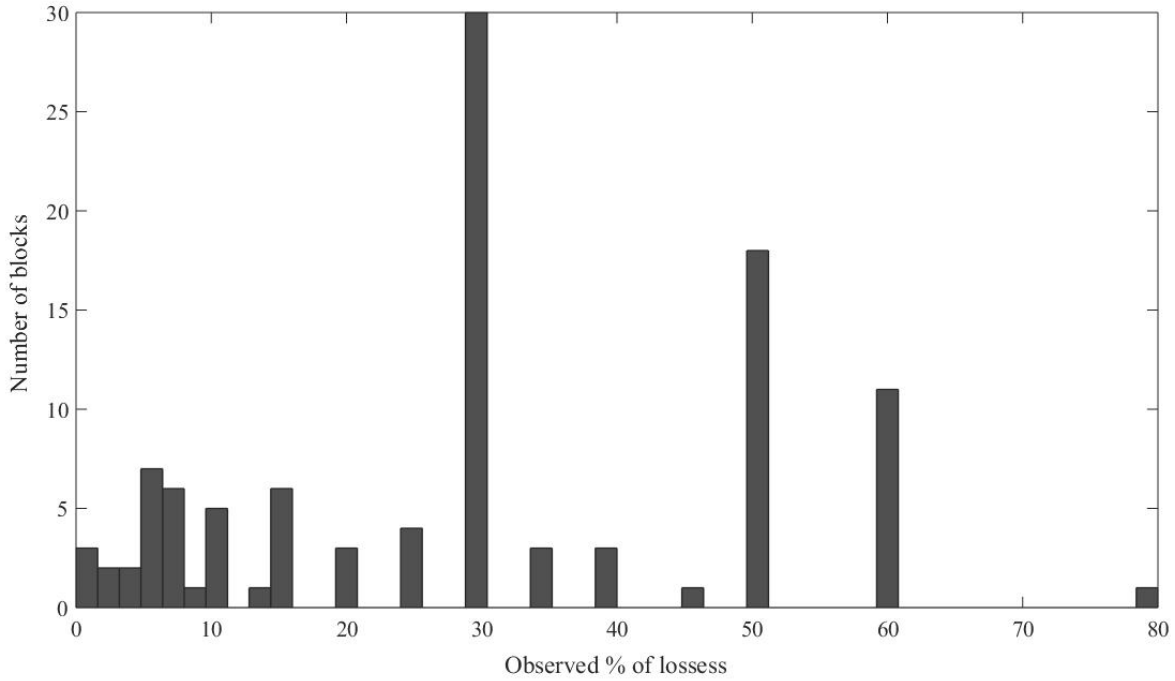


FIGURE 4. 6 - Percentage of yield losses observed by winegrowers on 107 blocks in southern France.

Modelling

Model construction

The final data set was characterised by the repeated observation of 12 spectral bands (variables) averaged for each of the 107 different vine blocks (objects) at 25 potentially available dates over four months (13th May to 20th August), in 2019. This data set was meaningfully arranged in a three-way array $\underline{\mathbf{X}}$ ($I \times J \times K$).

As discussed previously, the number of images per block varied according to each block's geographical location and possible cloud (or other) effects. Thus, $\underline{\mathbf{X}}$ was incomplete and there was a need for interpolation to obtain a continuous data cube. *J*, defined as the number of dates selected to represent the time-series, can be determined by optimising the parameters *N* and *P*. For this model, *J* was determined by using ranges of *N* and P described in Section 4.1.2.3. Once *J* was determined, interpolation was performed (again following the method in Section 4.1.2.3) to have a consistent time step. At the end of the image-processing step, the data were presented as:

- a cube $\underline{\mathbf{X}}$ (107, *J*, 12) where the first dimension corresponds to the vineyard blocks (*I*), the second dimension corresponds to time (*J*), which is optimised during modelling, and the third dimension of the three-way array $\underline{\mathbf{X}}$ corresponds to spectral bands (*K*) averaged for each field,
- a vector **y** (107), corresponding to the estimated percentage yield loss by the winegrowers from the 107 blocks.

Model validation

The calibration and validation subsets were generated using the method outlined in Section 4.1.2.3. The joint analysis of RMSEC (calibration set) and RMSECV (cross-validation) as proposed by Goodarzi & Freitas (2009) was used to determine the optimal number of latent variables in the regression model.

Model evaluation

The N-PLS was applied as an approach to detect and characterise the vineyard response (estimated by the percentage of losses) to extreme heat stress by using 12 available reflectance spectral bands from a time-series of Sentinel-2 imagery. As defined in Section 4.1.2.3, the prediction performance of the N-PLS model was quantified by the standard determination coefficient R^2, the bias and the standard error parameters. The regression b-coefficients resulting from the application of the methodology were plotted (i) against time as a function of wavelengths and (ii) against spectral bands as a function of time identify the time and the wavelengths that best highlighted this phenomenon. To further illustrate this, a 3D view of the b-coefficients vs. time vs. spectral bands was also generated. Figure 4.7 summarises the implementation and the workflow scheme of the N-PLS adapted to the case study.

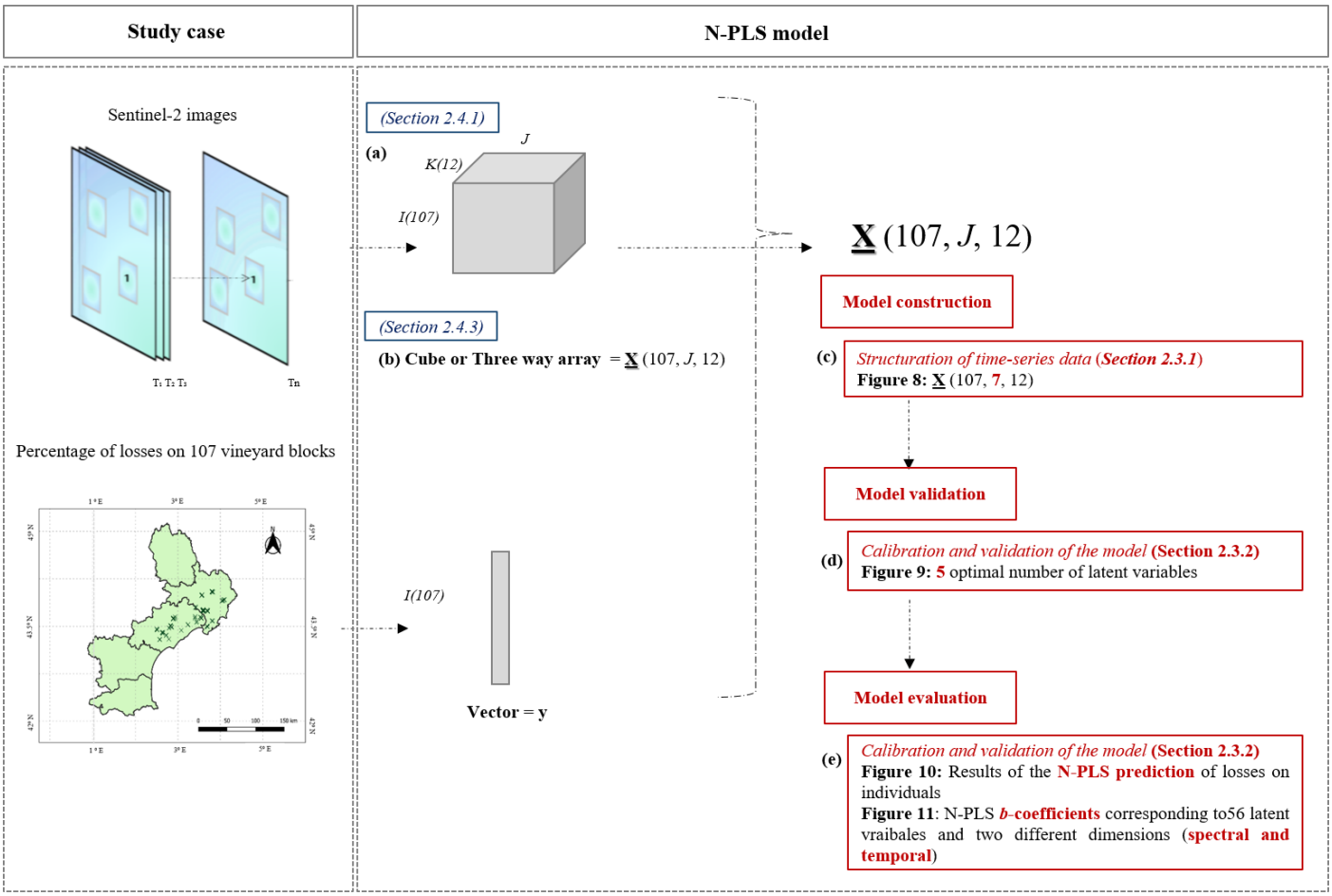


FIGURE 4. 7 - Workflow scheme of N-PLS application to the case study.

4.1.3 Results

4.1.3.1 Optimisation of model parameters over the study site

Figure 4.8 shows the evolution of the RMSECV for a cross-validation of 2 blocks repeated 5 times with a N-PLS performed with the **X** cube and the **y** losses from the calibration set when varying the Gaussian filter width (*P*) and interval between dates (*N*). This procedure led to an observation of the lowest error of cross-validation (RMSECV) at $P = 30$ and $N = 15$. $N = 15$ equates to $J = 7$, i.e. 7 dates over the study period, and resulted in a cube **X** of dimensions (107, 7, 12). Figure 4.8 highlights three situations in which the cross-validation error (RMSECV) increased when applied to these data, as follows: (i) when the Gaussian filter is too weak ($P > 25$); (ii) when the time step is short ($N < 10$ days); (iii) when the time step is long ($N > 20$ days).

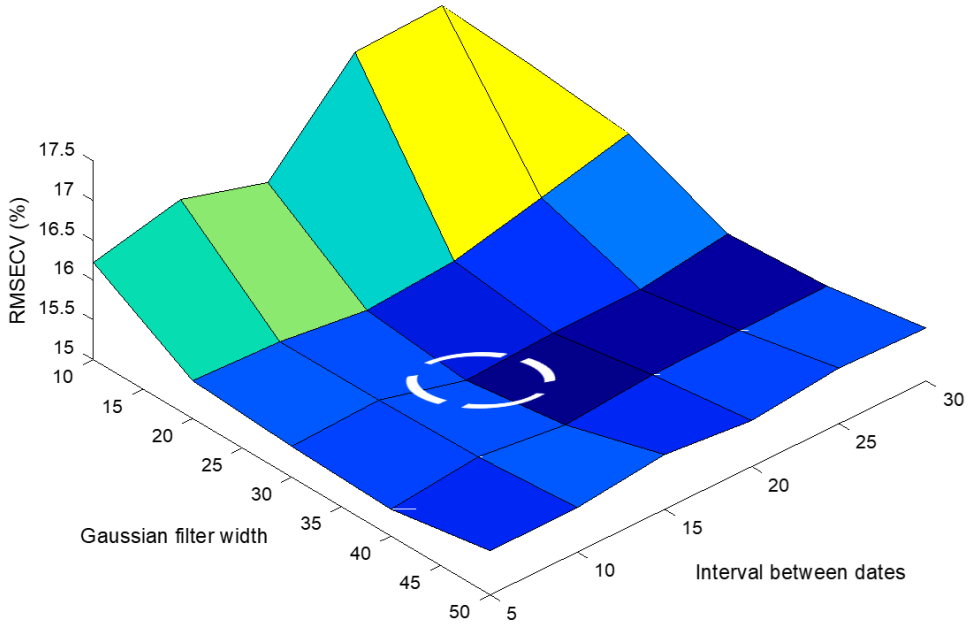


FIGURE 4. 8 - Evolution of RMSECV as a function of varying intervals between dates (*N*) and Gaussian filter width (*P*) obtained for 2 blocks repeated 5 times with a N-PLS performed with a $\underline{\mathbf{X}}$ cube and the **y** yield losses. Parameter optimisation was achieved at, $P = 30$ and $N = 15$ as indicated by the white dash circle.

Using the optimised *N* and *P* values, Figure 4.9 presents the evolution of the cross-validation RMSEC and RMSECV against the number of latent variables used in the model. This figure highlighted the following classical phenomenon: a phase of a decrease in the RMSEC, which corresponded to an improvement of the explanatory value of the latent variables, then a phase of increase of the RMSECV (while the RMSEC continued to decrease), which corresponded to the overlearning phase. On the basis of this joint analysis, 5 latent variables were retained for the rest of the study.

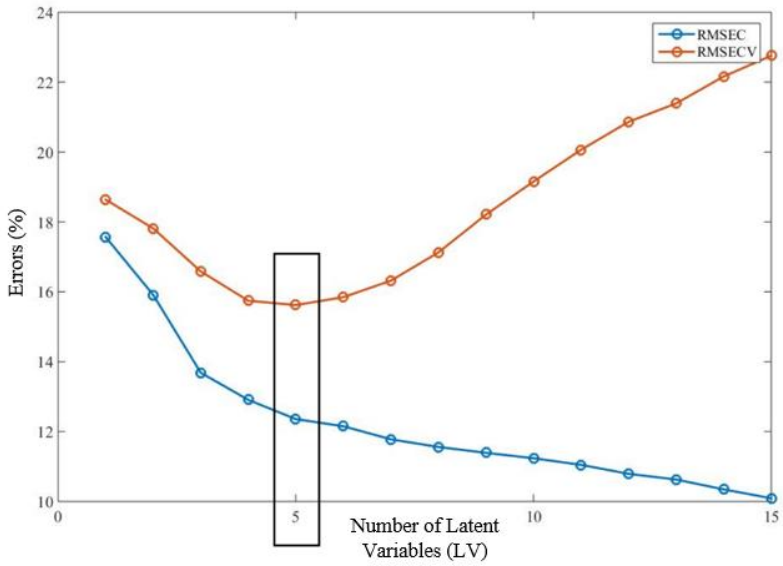


FIGURE 4. 9 - Evolution of the RMSEC and the RMSECV for a cross-validation of 2 blocks repeated 10 times of a N-PLS between the **X** cube and the **y** losses. The black frame indicates the optimal number of latent variables (5 LV).

4.1.3.2 Quality of the N-PLS model

The quality and performance of the N-PLS model with 5 latent variables are presented in terms of R^2, the bias and the standard error of prediction of yield losses in the calibration (Figure 4.10.a) and validation (Figure 4.10.b) analyses. The N-PLS model showed a performance accuracy (R^2) of 0.56 in the calibration set and 0.66 in the validation set, with a standard error of cross-validation in the calibration set of 12.4 % and a standard error of prediction of losses in validation set of 10.7 %.

A standard error of prediction over the prediction set of 10.7 % (Figure 4.10.b) was consistent with the initial variability of the ground truth data and also with the information needed by the growers to characterise the level of yield loss in a block that was caused by the heatwave. These results demonstrated the relevance of the approach to detect incidences of heatwaves in viticulture and in particular to estimate the yield loss based on spectral and temporal attributes of the images. The results were further validated by the calibration data.

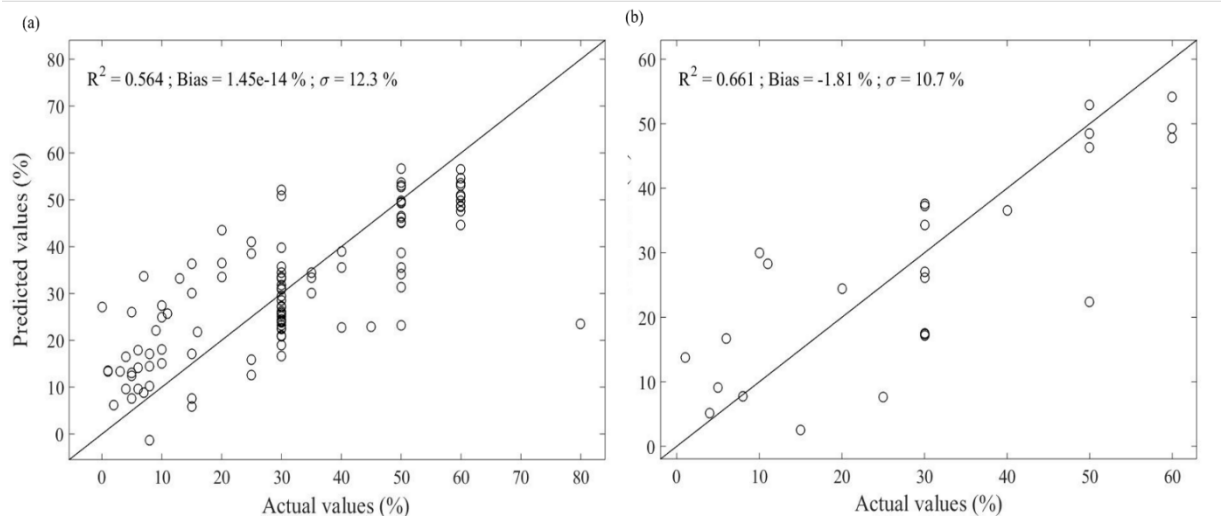


FIGURE 4. 10 - Results of the N-PLS prediction of losses on individuals in the calibration set (a), with 80 vineyard blocks and in the validation test (b), with 27 vineyard blocks. The standard error of cross-validation in the calibration set was 12.4 % and the standard error of prediction of yield losses in validation set was 10.7 %.

Figures 4.11 and 4.12 show plots of the coefficients $\mathbf{B}_{\mathbf{NPLS}}$ corresponding to the 5 latent variables selected for the model. The N-PLS model estimated yield loss for each block by making the inner product of *b*-coefficients with a spectral-temporal profile (7 × 12) contained in the cube $\underline{\mathbf{X}}$ (107,7,12). Thus, the b-coefficients had the same dimension as the spectral-temporal profile and could be analysed to identify how specific parts or periods of the spectral-temporal profile were more or less related to yield losses.

Figure 4.11 shows the spectral profiles for each spectral band (Figure 4.11.a) and the temporal profiles by date (Figure 4.11.b) of the b-coefficients. Regarding the spectral profile (Figure 4.11.a), the spectral bands were clearly organised into the following three groups: (i) bands B1 and B2 (442.7 and 492.4 nm), which presented no variations throughout the study period (ii) bands B3, B9 and B12 (559.8, 945.1 and 2202.4 nm), which showed a slight decrease by the end of June and (iii) the remaining bands B4, B5, B6, B7, B8, B8A and B11 (663.5, 704.1, 740.5, 782.8, 832.8, 864.1 and 1613.1 nm), which presented a very pronounced V-shaped profile with a minimum b-coefficient observed on the 20th of June, identifying these spectral bands as significant in order to identify and characterise the heat stress.

Regarding the temporal profiles (Figure 4.11.b), they were very similar in shape and mainly differed in translation. The higher coefficients were observed for the dates at the beginning and at the end of the study period while the lowest b-coefficients were observed for 20th June, followed by the intermediate dates (5th June and 5th July). The spectral profiles highlighted the end of June as the period that affected on yield loss. Given the interpolation

performed, the date of 20th June was the closest to the date of the onset of the heatwave and this date should be considered as the best indicator.

When looking at the *b*-coefficients, independently from the time dimension, it can be seen that the heat sensitivity was related to a specific spectral shape, with two absorption zones at 700 and 1600 nm (Figure 4.11.b). This spectral shape was close to the typical spectrum of vegetation. It has already been demonstrated that the red edge band has a good potential for estimating plant stress (Cogato et al., 2019b). These results suggested that spectral bands can provide useful information about heat stress in vineyards. By integrating the temporal dimension, it was noticed that the vineyards that suffered most yield losses from the heatwave not only presented a specific spectral shape, but that this spectrum deepened close to the moment of the heatwave (Figures 4.11.a and 4.12). Figure 4.12 combines Figure 4.11 to shows the 3D shape of the *b*-coefficients accounting for the spectral and temporal dimension simultaneously.

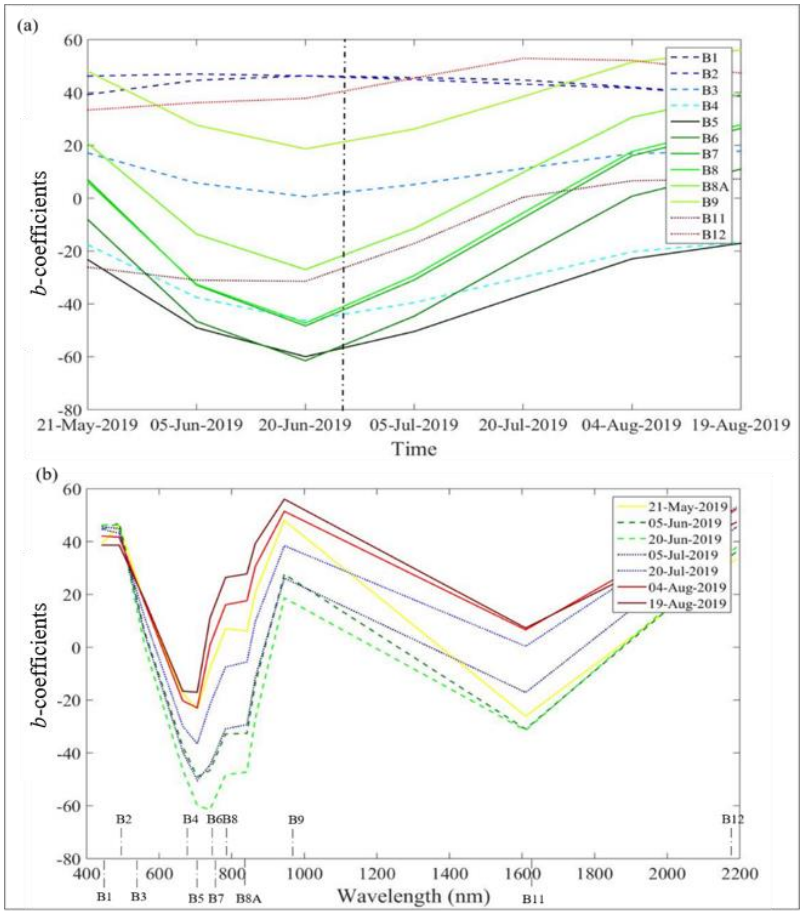


FIGURE 4. 11 - N-PLS *b*-coefficients corresponding to 5 latent variables and the following two different dimensions: (a) plotted according to the spectral dimension and (b) plotted according to the temporal dimension. Black dot-dash line highlights the most relevant date of the heatwave.

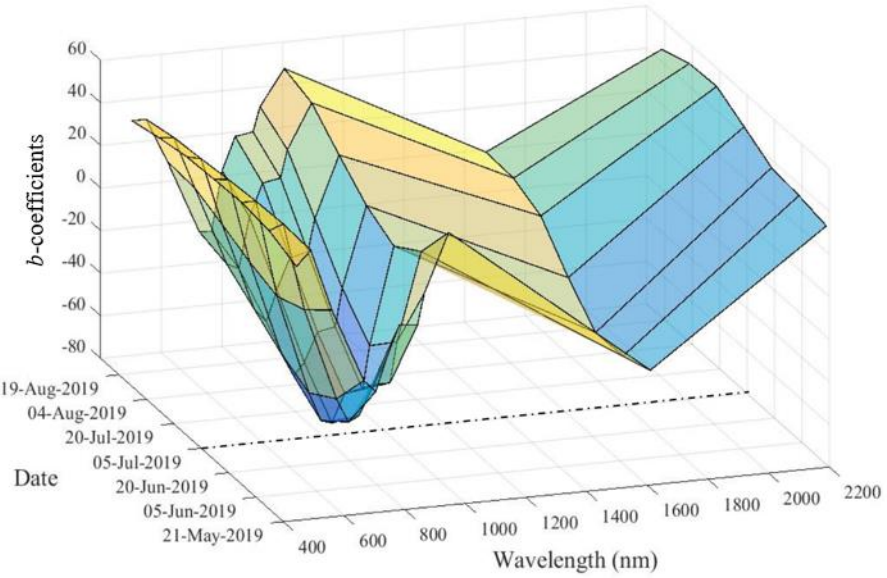


FIGURE 4. 12 - Combined representation of the temporal and spectral profiles as a 3D view of N-PLS *b*-coefficients corresponding to 5 latent variables. Black dot-dash line to highlight the most relevant date of the heatwave.

4.1.4 Discussion

A generic example of the application of N-PLS algorithm to time-series of multispectral images was provided in the form of an agricultural study. This paper showed the potential for methods originally developed in the spectrometry/chemometrics domains to be applied to a time-series of Sentinel-2 data. The application showed the value of considering simultaneously the temporal and spectral dimensions with a sensor whose band repartition was partly optimised for vegetation monitoring.

The application of N-PLS to characterise and estimate the impact of an extreme event, such as a heatwave on grapevines, showed that it was possible to predict yield losses with a performance (R^2) of 0.66 and a RMSE of approximately 10 %. The theory for the use of remote sensing instruments to monitor electromagnetic radiation reflectance changes in crops is well demonstrated in the scientific literature (Knipling, 1970). In situations where crops interact with any given aspect of their environment, such as seasonal climatic variations or meteorological extreme events, the interactions between plants and light reflectance translates into changes in plant signal patterns that can be interpreted using satellite data (Segarra et al., 2020). Besides the model of yield loss estimation, the N-PLS analysis showed the interest of adopting a systemic analysis which accounts simultaneously for the spectral and temporal characteristics of the considered data. The approach allowed the identification of the spectral bands that

responded most strongly to the phenomenon of interest while keeping the information of the period of influence.

Previous studies have reached similar accuracies to assess the effects of heat stress on grapevines (Cogato et al., 2019b). However, notably, the linear nature of the dimensionality reduction in the method presented here allowed for a simple interpretation using the computed *b*-coefficients, which are directly related to the importance of the explanatory variables, i.e. to the best spectral signature related to the event under study. Since b-coefficients have the same dimension as the spectral-temporal profile, spectral bands around Red Edge (700 nm) and the SWIR region (1600 nm) in late June and early July, were considered relevant to quantify the effect of a heatwave that occurred in late June. It has already been demonstrated that the Red Edge band has a good potential for estimating plant stress, either due to water loss or to leaf pigment modifications (Seelig et al., 2009; Cogato et al., 2019b; Laroche-Pinel et al., 2021a). In particular, this band has been useful for studying how experimental water deficits change the characteristics of plant physiology (e.g. chlorophyll a/b ratio), as plants experiencing a water deficit change their foliar chlorophyll composition, resulting in a shift of red-edge reflectance towards shorter wavelengths (Easterday et al., 2019). Regarding the shortwave infrared region (SWIR), from approximately 1300 nm to 2500 nm, the absorption of radiation is largely dominated by water (Segarra et al., 2020). These results generally supported the prediction that wavelengths, where the water absorption coefficient is weak, will penetrate further into canopies and thus will be best for estimations of water content, as is the case in the 1600 nm region (Sims & Gamon, 2003). This implied that the SWIR region of the spectrum may potentially be efficient for drought stress detection and resulting changes in canopy tissues (Cogato et al., 2019b). The results of the current analysis confirmed, with a data-driven approach, that a combination of Red Edge and SWIR regions was a valuable indicator in identifying grapevine heat stress on a regional scale.

It is essential to place the results presented in this paper within the reality of multitemporal satellite data. Satellite observations are heavily conditioned by cloud coverage over the acquisition period. Despite this, this method was able to manage a time-series dataset with different sets of missing values by means of an interpolation on a date grid that was regularly spaced and optimised over the period of interest. The N-PLS methodology does not handle a high percentage of missing values within its cubic data structure and, it is expected that this approach would be affected in cases where cloud conditions during the target period significantly limit the number of available images. Another constraint that the model was able to deal with was the low number of ground truth samples (in this case, yield loss observations for grape fields). Note that ground truthing remains cumbersome and difficult to manage for many environmental and agricultural studies. Such limitations can influence the values of model outputs, resulting in a lower reliability of the supervised approach. However, this work showed that N-PLS gave relevant prediction results for a complex phenomenon with a very limited number of data (n = 80) in the calibration data set.

These results indicated that the algorithm may be suited to other similar satellite systems, such as the Landsat products. However, it is important to keep in mind that while the spectral band specifications are similar between the Sentinel and Landsat systems (cross-calibration was actually performed between the two sensors (Claverie et al., 2018), the Landsat constellation provides imagery with a lower spatial resolution (30 m for most of the spectral bands) which may be a problem for some applications (e.g. intra-field variability in vineyards) but is not crucial for others. The combination of both Landsat and Sentinel-2 products remains a research question (Quintano et al., 2018). However, such an image combination may be interesting to explore with the N-PLS method in order to foster some limit when working at the regional scale, i.e. missing values due to weather conditions that prevent a full time-series of the objects being studied to be obtained.

While this paper has focused on the monitoring of grapevine heat stress, the identification of the spectral bands that best explain the impact on plant canopy and the development of a local model, which can be applied to predict the spatial extent of the phenomenon, can be transferred to other agricultural applications related to climatic phenomena (e.g. hail or frost). Because of the type of approach, the model should only be valid for the year and the region considered. The resulting model remains specific to the learning base used for the calibration and its generalisation to other vintages and/or other agricultural regions is somewhat limited, especially when dealing with extreme climatic events. However, the results obtained can help facilitate the development of empirical models to be applied to other situations and other vintages.

The type of multi-spectral and multi-temporal method presented in this work is expected to be relevant for any application that relies on different spectral domains with a clear temporal evolution. However, its application to more gradual phenomena, such as progressive changes in plant water status over the summer period, would require further testing, as it may be less relevant because changes in plant characteristics may be less obvious.

In general, the methodology presented here can be effective for any environmental and agricultural application of Sentinel-2 that involves the monitoring of a sharp temporal evolution and where all spectral information may be useful.

4.1.5 Conclusions

This study demonstrated how, with a dimensionality reduction algorithm, it is possible to describe the underlying phenomena by highlighting spectral bands and relevant time periods within a time-series of imagery.

This was demonstrated by analysing the footprint of an extreme climatic event and predicting yield losses from a heatwave in 2019 in vineyards in the Languedoc-Roussillon region. In this case, spectral bands around the Red Edge (700 nm) and the SWIR region (1600

nm) were found relevant, as their spectral profile showed a deepening of the time profile during the period of the heat stress, which permitted the band information to be related to reported yield losses.

The proposed method cannot handle temporal data gaps due to clouds and inconsistent availability of satellite images during the acquisition period. Interpolation of the data to achieve a three-way data structure was mandatory for the development of the methodology. By doing this, the method was able to deal with a time-series data set with different sets of missing values and to identify a noise reduction approach with a 15-day time step, which led to a consideration of only seven dates over the study period.

Further investigation is needed to (i) complete the results of this study, expanding the analysis to the spatial distribution of the phenomenon in order to determine its regional dynamics and thus the reasons for its main effects, and to (ii) confirm the applicability of the N-PLS method to different time-evolving phenomena.

4.2 Is it Possible to Assess Heatwave Impact on Grapevines at the Regional Level with Time Series of Satellite Images?

Article 4: Is It Possible to Assess Heatwave Impact on Grapevines at the Regional Level with Time Series of Satellite Images?

Published in *Agronomy*, 2022, *12* (3), pp.563. DOI: 10.3390/agronomy12030563

E. Fornieles-Lopez[1,2], G. Brunel[1], N. Devaux[3], J.M. Roger[1,2] and B. Tisseyre[1]
[1]ITAP, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France
[2]Chemhouse Research Group, 34000, Montpellier, France
[3]LISAH, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France

Abstract: Unexpected climatic conditions or extreme climatic events in vineyards are a worldwide problem that requires accurate spatial and temporal monitoring. Satellite-based remote sensing is an important source of data to assess this challenge in a climate-change context. This paper provides a first insight into the capacity of a multiway analysis method applied to Sentinel-2 time series to assess heatwave impacts in vineyards at a regional scale. N-way Partial Least Squares (N-PLS) regression was used as a supervised technique to predict the intensity of damage caused to vineyards by the heatwave phenomenon that impacted the vineyards in the south of France in 2019. The model was developed based on available ground truth data of yield losses for 107 vineyard blocks in the Languedoc-Roussillon region and multispectral time-series predictor data for the period from May to August 2019. The model showed a performance accuracy (R^2) of 0.56 in the calibration set and of 0.66 in the validation set, with a standard error of cross-validation in the calibration set of 12.4 % and a standard error of the prediction of yield losses in the validation set of 10.7 %. The model was applied at a regional scale on 4978 vineyard blocks to predict yield losses using spectral and temporal attributes. The prediction of the yield loss due to heat stress at a regional scale was related to the spatial pattern of maximum temperatures recorded during the extreme weather event. This relation was confirmed by a chi-square test ($p < 5$ %). The introduction of N-PLS insights into the analysis enables the characterisation of heat stress responses in vineyards and the identification of spectro-temporal profiles relevant for understanding the effects of heatwaves on vine blocks at a regional scale.

Keywords: unfold methods; N-PLS; heat stress-2; water relations; remote sensing

4.2.1 Introduction

Grapevine (Vitis vinifera L.) is widely recognised as one of the most important crops in Europe (Droulia & Charalampopoulos, 2021). Growing evidence of the significant impact of climate change on viticulture is driving new and underexplored research aiming at monitoring and understanding its incidence on vine cultivation. The two factors most frequently addressed in reflections on the possible effects of climate change (CC) on viticulture are thermal and hydrological conditions (Droulia & Charalampopoulos, 2021). Both of these effects have an impact on vine development and fruit composition, determining yields and the quality of grapes, and thus of the wine produced (Pinel et al., 2021). The most measurable effect of CC is that the steady increase in temperature leads to a rise in radiation and in the frequency and severity of more extreme weather events, such as heatwaves (Droulia & Charalampopoulos, 2021).

According to Pinel et al. (2021), the climate of the Mediterranean region until the end of the 1990s was defined by wet winters and warm summers, with balanced rainfall. However, since 2000, low rainfall and increased evapotranspiration have defined the growing period of the vineyards (April–June), intensified by the increasing occurrence of heatwaves at key phenological stages of the vine. Prolonged periods of unusually high temperatures are likely to affect the yield and quality of the vines (Cogato et al., 2019b). In June 2019, an exceptional heatwave episode hit the south of France, causing severe and irreversible damage in vineyards. Persistent temperatures above 35 °C during the growing season drastically affect the plant response and heat acclimatisation mechanisms are activated (Venios et al., 2020). By affecting the photosynthesis rate (Carvalho et al., 2016; Cogato et al., 2019b) and intensifying drought stress (Nicholas and Durham, 2012), heat stress has a considerable influence on the physiology and yield of grapevines (Fraga et al., 2020). Since fluctuations in environmental conditions, and especially ambient temperature (Venios et al., 2020), strongly influence plant growth and plant developmental processes, it is crucial to capture the dynamics of vine growth over time, especially at critical growth stages. Therefore, time series of multispectral images may constitute a relevant tool to assess the incidence and the spatial footprint of a heatwave at a large scale (e.g. the scale of a production basin, region, etc.).

Remote sensing techniques are largely used in agriculture, focusing on traits or features of the agricultural systems that vary in space and time. Based on high-quality multi-temporal and multi-spectral images captured by Earth observation satellites, satellite remote sensing has a great potential to address the challenges of CC due to its ability to provide timely and comprehensive information at different scales and for different actors (Weiss et al., 2020). Due to the increasing availability of remotely sensed data, e.g. the multispectral Sentinel-2 satellites provide revisits every 5 days, global coverage of Earth's land surface makes large-scale analysis possible (Bovolo & Bruzzone, 2015). Benefiting from convenient spatial resolutions at different scales (plot, production basin, appellation, region, etc.), satellite data allows the development of tools and methods that account both for the continuous spatio-temporal reality of a

phenomenon, such as a heatwave, and the spectro-temporal dynamics of the development of a specific crop, such as vines. In the field of viticulture, information from 13 spectral bands (from visible to shortwave infrared), on vine conditions obtained by Sentinel-2 over time provides a detailed time series of data on the physiological and physical properties of the vine (Plant et al., 2000). According to Filella et al. (1995), the reflectance spectrum changes depending on growing conditions and the time of measurement relative to the stage of crop development. Therefore, remotely sensed multispectral images have a particularly important potential to quantify the effects of extreme events in the context of global climate change, at different spatial scales (e.g. plot and region), on grapevine yield.

Different approaches to assess the impact of extreme weather events on major crops can be found in the remote sensing literature (Cogato et al., 2019a). However, the publications dealing with heatwaves in viticulture with remote sensing are sparse (Cogato et al., 2019b). Stress conditions due to fluctuations in ambient temperature certainly affect the physiological behaviour of the grapevine, and thus the spectral response of the canopy at various wavelengths (Cogato et al., 2021). Although the current knowledge on the physiological dynamics regulating the responses of grapevines to heatwaves appears to be well established (Webb et al., 2010; Fraga et al., 2020; Venios et al., 2020), large research gaps still exist in the assessment of the effects of heat stress using spectral features and environmental parameters. Cogato et al. (2019b) proposed a relevant approach based on Sentinel-2 time-series data that highlights the most suitable spectral regions and VIs for heat stress detection. However, such an approach does not take into account the spatio-temporal extent of the phenomenon or the spectral-temporal extent of the cultivation, as is the case with multivariate methods. Moreover, such an approach presents the risk of creating uncertainty about the possible effects of the heatwave timing and limiting knowledge about the phenomenon in question. Furthermore, challenges remain for modelling the effects of extreme weather events in agriculture without overlearning. Regarding the assessment of CC at different scales, quality field data are scarce and difficult to homogenise, so the number of samples available to study the phenomenon is often limited. This is a paradoxical phenomenon, considering that observational data from remote sensing have never been so numerous (increased data volume and high data variability) (Bishop, 2013), while ground reference data, particularly in operational contexts, remain sparse.

The research presented here aimed to evaluate the incidence of a climate event, such as a heatwave, on vine cultivation at the plot level and its extent at the regional scale. The specific objectives of this work were to: (i) propose an adaptable systemic multidimensional methodology able to consider the spectral and temporal dimensions intrinsic to cultivation; (ii) design, calibrate and validate a model to predict yield losses based on spectral-temporal information derived from time series of remote sensing images; and (iii) study and assess the quality and uncertainty of this predictive model when used at the regional scale.

4.2.2 Materials and Methods

4.2.2.1 Study area

The study area corresponded to a large wine-growing region, the Languedoc-Roussillon (LR), extending over almost 27,400 km2 in the south of France (Figure 4.13). The LR vineyards, united under the same administrative label but with their own characteristics, cover four French administrative sectors: Gard (A), Hérault (B), Aude (C) and Pyrénées-Orientales (D) (Figure 4.13). It encompasses a large diversity of varieties, training systems, etc. (Fernández-Mena et al., 2021).

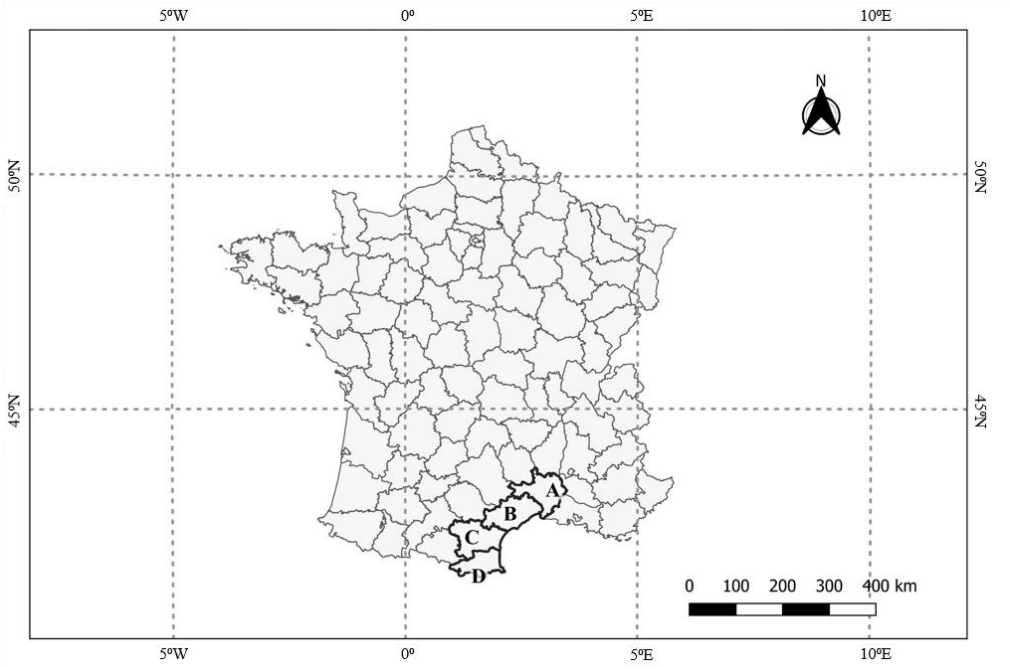


FIGURE 4. 13 - Location of the study area in Southern France for the four administrative sectors: Gard (A), Hérault (B), Aude (C) and Pyrénées-Orientales (D).

The climate, typically Mediterranean, is characterised by hot and dry summers, with sparse summer rainfall and mild winters. The regional level presents a large variability in pedo-climatic conditions (Fernández-Mena et al., 2021); however, soils typically share common characteristics: low fertility, high stoniness, good drainage, absence of a limiting horizon to ensure deep rooting and limited water holding capacity.

Heatwave stress characteristics

In June 2019, the LR wine-growing area experienced a heatwave characterised by a hot wind, blowing from the north-east to the south-west, with temperatures reaching 45 °C (Figure 4.14). The extreme weather episode occurred between 25th June and 8th July 2019, of which 28th June was the most critical day. Given that vine growth in Mediterranean conditions is still occurring, although slowing, in the middle of the season, extreme environmental fluctuations that occur during this period at very rapid time scales will limit the evaporative cooling of the leaves and induce symptoms that can even lead to wilting of the leaves (Schymanski et al., 2013). Extremely high-temperature regimes, characteristic of a heatwave, affect the biochemical and physiological processes necessary for the optimal development of the vine, especially for early ripening varieties (Droulia & Charalampopoulos, 2021).

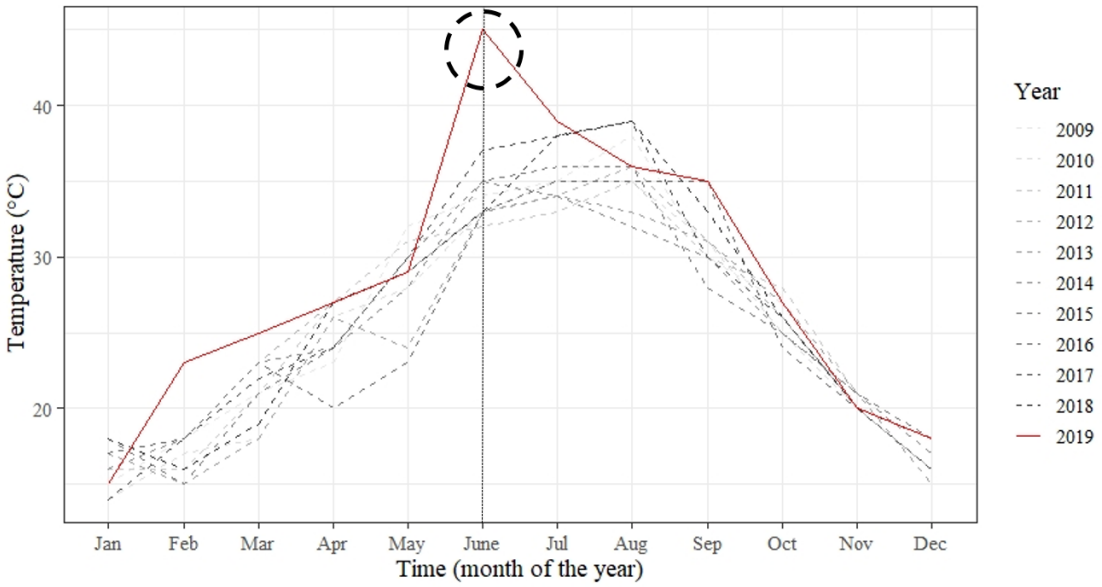


FIGURE 4. 14 - Maximum monthly temperatures recorded from 2009 to 2019 over the Languedoc-Roussillon region, France, highlighting a peak corresponding to the extreme weather event that occurred in June 2019. The vertical black dashed line highlights the month of the heatwave. Source: Historique de Météo-France.

It should be noted that the heatwave did not affect the whole region equally, as shown in Figure 4.15. The northern part of the region (sectors A and B) was the most strongly affected. Indeed, sector B presented the temperature (>44 °C), with a strong spatial variability, since these high temperatures did not affect the eastern part. Sector D was only severely impacted on the western part, and sector C was almost unaffected by high temperatures. Note that Figure 4.15 only presents the main trend of the heatwave. It may hide some local (short range) phenomena due to factors that may locally mitigate or amplify the temperature effect

experienced by vineyard blocks, such as elevation, the presence of forest and the aspect of the topography.

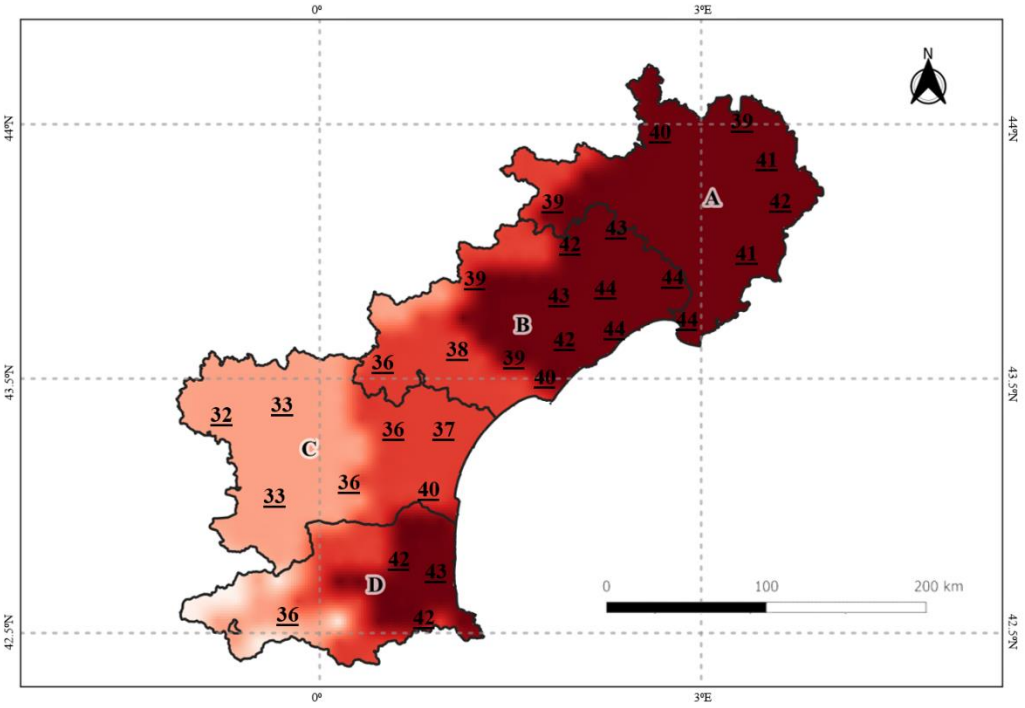


FIGURE 4. 15 - Map of maximum temperature recorded for 28th June 2019 in the Languedoc-Roussillon region. Source: SAFRAN grid, Météo-France.

Ground truth data

The ground truth data were selected from 107 non-irrigated vineyard blocks in the northern part of the LR region (Figure 4.16.a). They all showed some effects related to the heatwave, such as stalled development, leaf burn and leaf drop (Lopez-Fornieles et al., 2022). The severity of this effect was assessed by winegrowers and advisors on each of the 107 vineyard blocks several weeks after the event through an estimation of the percentage of yield loss. Figure 4.16.b summarises the distribution of the 107 blocks in relation to yield loss. Note that the ground truth data concerned only sectors A and B, which were the most impacted by high temperatures during the heatwave. No vineyard blocks were sampled over sectors C and D, despite the potential impact of the heatwave on both of these sectors. However, this unbalanced spatial distribution of the blocks leads to a representative dataset with a large diversity of observed yield loss values (Figure 4.16.b).

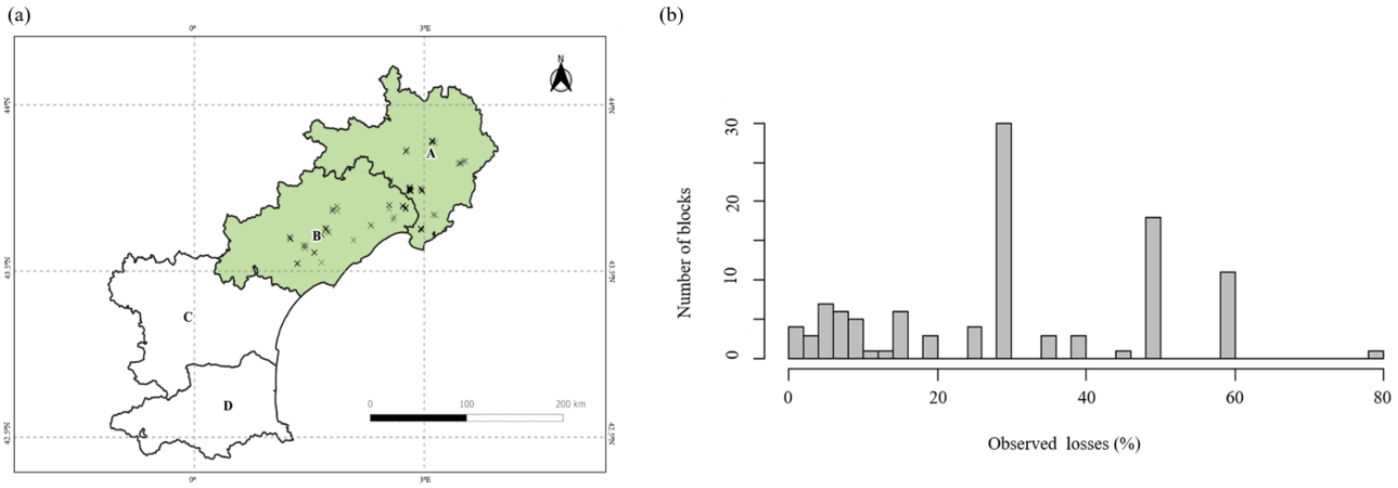


FIGURE 4. 16 - (a) Map of the 107-ground truth blocks with a known estimated percentage of yield loss after the heatwave; and (b) percentage of yield losses observed by winegrowers and advisors on 107 vine blocks in southern France (Lopez-Fornieles et al., 2022).

4.2.2.2 Remote sensing data

Data Acquisition and processing

Satellite images were selected via the Google Earth Engine (GEE) platform, which provides Sentinel-2 L2A (Sentinel-2A and Sentinel-2B) products. Sentinel-2 satellites, with a revisit frequency of 10 days (5 days with the twin satellites (A/B) together), provide 13 spectral bands from visible (Vis) and near-infrared (NIR) to shortwave infrared (SWIR) regions of the spectrum, with a spatial resolution of 10, 20 and 60 m (Table 4.2) (Lopez-Fornieles et al., 2022). Twelve spectral bands (among the 13 available from Sentinel-2 satellites) were used in this study (Table 4.2). Spectral band 10 at 1380 nm was not used as it was designed for the detection of visible and sub-visible cirrus clouds and corresponds to a band of high atmospheric absorption (Hollstein et al., 2016).

TABLE 4. 2 - Spectral bands for the Sentinel-2 satellite considered by the analysis.

Sentinel-2 Band	Central Wavelength (nm)	Bandwidth (nm)	Spatial Resolution (m)
Band 1–Aerosol	442.7	21	60
Band 2–Blue	492.4	66	10
Band 3–Green	559.8	36	10
Band 4–Red	664.6	31	10
Band 5–Vegetation Red Edge	704.1	15	20
Band 6–Vegetation Red Edge	740.5	15	20
Band 7–Vegetation Red Edge	782.8	20	20
Band 8–NIR	832.8	106	10
Band 8A–Vegetation Red Edge	864.1	21	20
Band 9 –VNIR	945.1	20	60
Band 11–SWIR	1613.1	91	20
Band 12–SWIR	2202.4	175	20

The time period considered for the study was from May to August 2019, which was the most relevant period to monitor vine growth vegetation in this region (Devaux et al., 2019). Only images containing the study vineyards (Section 4.1.2.4) from 13th May to 20 August 2019 were selected and processed via Google Earth Engine (GEE) (Lopez-Fornieles et al., 2022). Blocks' boundaries were extracted from the graphical parcel register of France (RPG). According to the highest spatial resolution of Sentinel-2, to avoid mixed pixels, a 10 m inner buffer was imposed over the boundary of each block before average pixel values were computed within the inner boundary for each block, each date and each wave band (Lopez-Fornieles et al., 2022). For the chosen period, 25 images should have been potentially available over each block, but following the cloud detection algorithm for Sentinel-2 imagery proposed by Hollstein et al. (2016), the number of available images was 11 on average, with a range from 7 to 16 images depending on the location of the different blocks (Lopez-Fornieles et al., 2022).

4.2.2.3 Modelling

N-way Partial Least Squares

In this work, data corresponded to a three-way array $\underline{\mathbf{X}} = ||X_{i,j,k}||$ whose dimensions involved the individuals (*I*), i.e. the vine blocks, as a first dimension, the second dimension corresponded to time (*J*) and the third dimension corresponded to mean reflectance at each wavelength (*K*). Therefore, as shown in Figure 4.17, data were organised in a three-way array of independent variables $\underline{\mathbf{X}}$ ($I \times J \times K$) derived from the remotely sensed data (Section 4.1.2.4) and a response vector y of size ($I \times 1$) corresponding to the ground truth data (Figure 4.16.a).

N-PLS regression, as an extension of the classical Partial Least Squares (PLS) method, was chosen to analyse the data. As proposed by Hansen et al. (2002), it is an interesting method to relate a N-way array to a dependent variable. It identifies all latent information from (**X**), which maximises the covariance between **X** and **y** while keeping information provided by each dimension of **X** (Abdi, 2010). In this study, the 3-PLS1 regression method (Bro, 1996) was used to best relate **X** with **y** while keeping information provided by the spectral and the temporal dimensions.

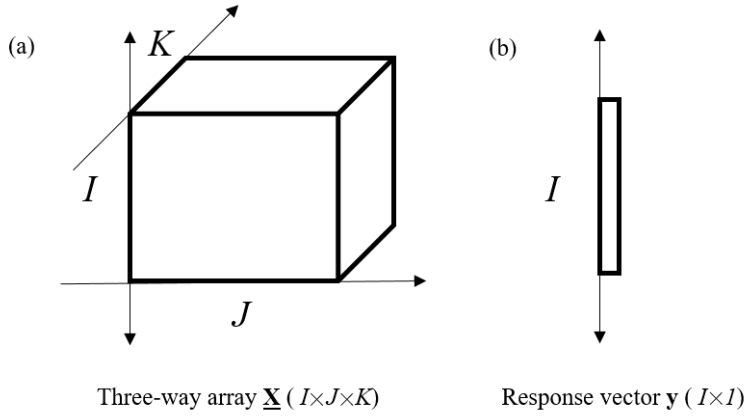


FIGURE 4. 17 - (a) Representation of the three-way array (**X**); and (b) response vector **y** (Lopez-Fornieles et al., 2022).

Data array construction

The dataset was characterised by the observations of the 12 spectral bands averaged over each of the 107 blocks for all available satellite images, between 13th May and 20th August, in 2019 (Lopez-Fornieles et al., 2022). However, the number of images per block varied according to the local atmospheric conditions over each block for each acquisition date. As a result, all potential Sentinel-2 images were not necessarily available at each date. To overcome this issue, an interpolation was performed to obtain a continuous data cube. The interpolation at a date t was done wavelength by wavelength, by a convolution of the chronology measured with a Gaussian filter (Alam et al., 2008) in order to have a consistent time step dimension (*J*) (Lopez-Fornieles et al., 2022). The parameters involved in the interpolation setting were optimised by cross-validation of 2 blocks repeated 5 times of a N-PLS between cube **X** and vector **y**. Parameter optimisation was achieved with a Gaussian filter having width (*P*) = 30 and date interval (*N*) = 15 (Lopez-Fornieles et al., 2022).

At the end of the interpolation step, the dataset was meaningfully arranged in a three-way array **X** of dimensionality 107 (samples, *I*) × 7 (times, *J*) × 12 (wavelengths, *K*) and a vector **y** (107), corresponding to the yield loss rates from 107 blocks estimated by the winegrowers and advisors.

Model calibration and prediction

Calibration and validation subsets were created to build and evaluate the model. Considering the distribution of the samples from the dependent variable (Figure 4.16.b), a calibration set (3/4) and a validation set (1/4) were defined as follows to ensure that the two sets had the same final distribution (Lopez-Fornieles et al., 2022):

1. The vector y was sorted in ascending order.
2. After sorting, every fourth individual was placed in the validation set and the others were kept in the calibration set.

A cross-validation of 2 blocks repeated 10 times of a N-PLS between the $\underline{\mathbf{X}}$ cube and the **y** vector from the calibration set was performed. The joint analysis of the Root Mean Square Error of Calibration (RMSEC) and the Root Mean Square Error of Cross-Validation (RMSECV), as proposed by Goodarzi et al. (2009), was used to determine the optimal number of latent variables (LVs) in the regression model. On the basis of this joint analysis, 5 meaningful latent variables were kept for the model (Lopez-Fornieles et al., 2022).

The prediction performance of the model was quantified on the validation subset (data not used for the model calibration) by the standard determination coefficient R^2, the bias and the standard errors.

Model application at regional scale

In order to assess the potential of the approach in identifying the spatial footprint of the heatwave phenomenon at the regional scale, the calibrated N-PLS model was applied to 4978 vineyard blocks spread over the whole LR region (Figure 4.18). These vineyard blocks corresponded to all data available from the RPG.

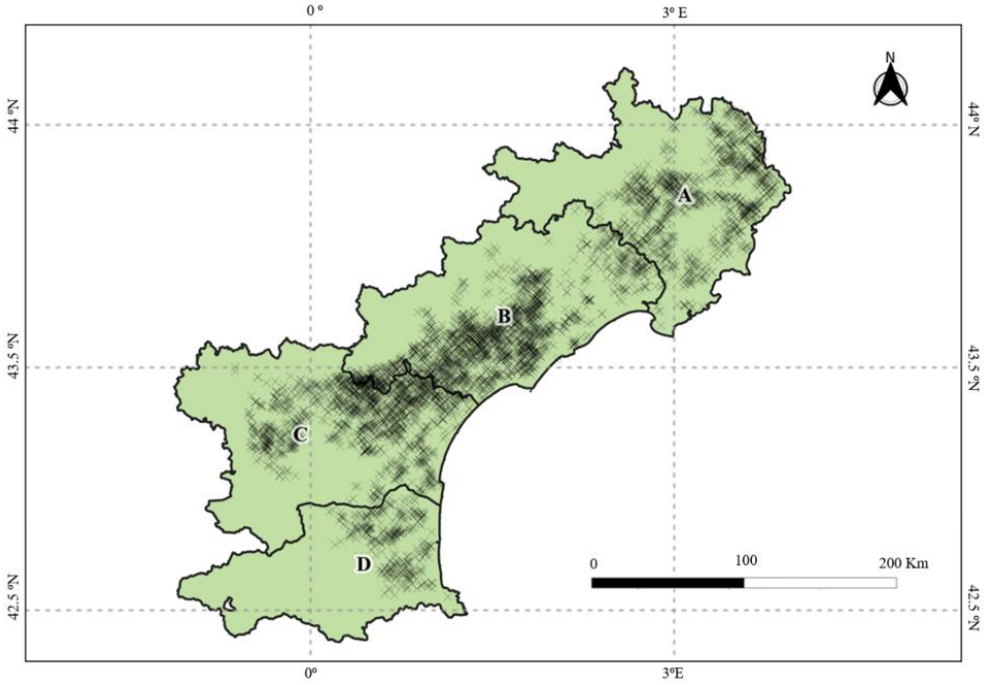


FIGURE 4. 18 - Location of vineyard blocks (4978) of interest in Southern France.

For the implementation of the N-PLS model, the same steps were followed as for $\underline{\mathbf{X}}$ data array construction. The model with the 5 latent variables was therefore applied to a significant three-way array $\underline{\mathbf{X}}_2$ of dimensionality 4978 (samples, *I*) × 7 (times, *J*) × 12 (wavelengths, *K*). The model application provided an estimation of the yield loss over the 4978 vineyard blocks. Figure 4.19 summarises the implementation and the workflow scheme of the N-PLS model calibration, in addition to its validation and its application at the regional scale.

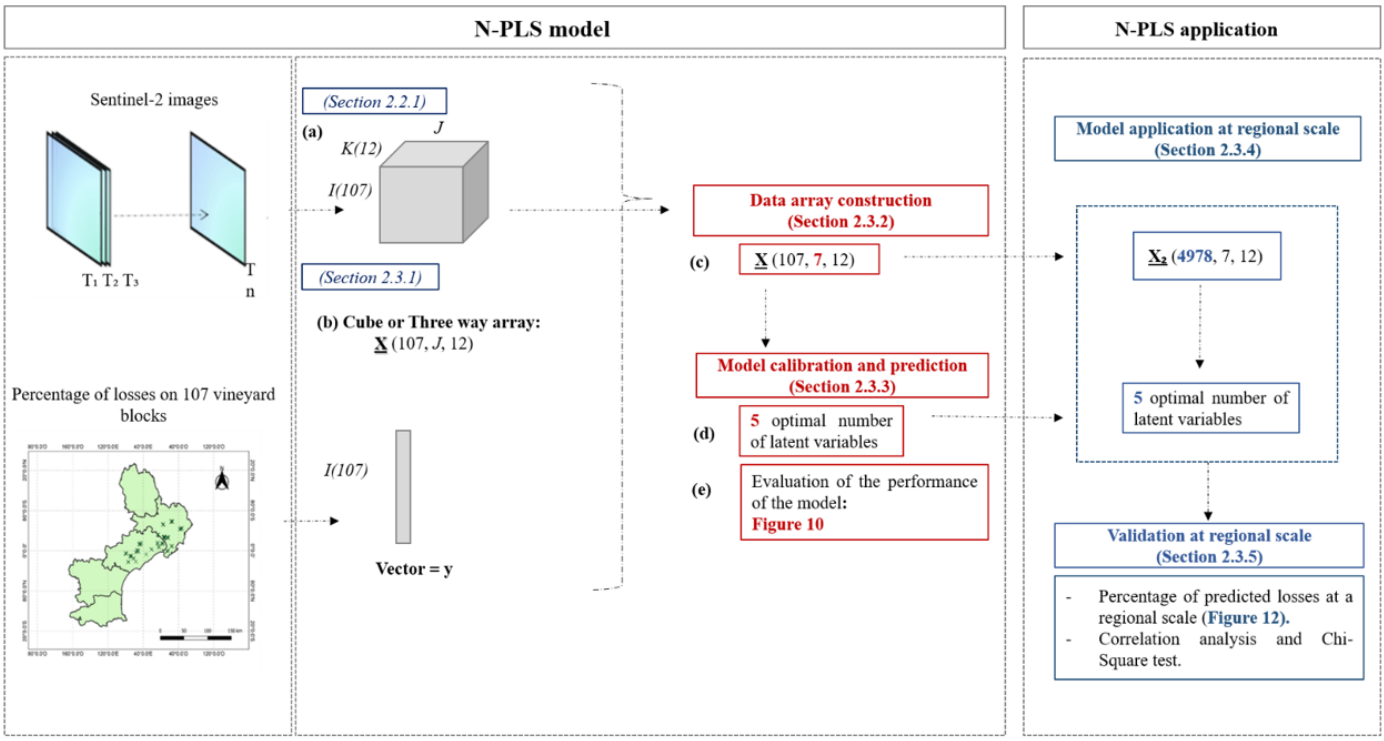


FIGURE 4. 19 - Workflow scheme of the calibration of the N-PLS model, the prediction evaluation and its application to 4978 vineyard blocks at the regional scale.

Validation at regional scale

Yield loss estimations at the regional scale were interpolated following a classical kriging process (see Section 4.2.2.4). Estimations were analysed both qualitatively from visual comparison with the maximum temperature record map of the 28th of June 2019 (Figure 4.15), and quantitatively, based on the occurrence of predicted yield loss values with respect to the maximum temperature recorded over the block. As validation, a chi-square test of independence (Greenwood and Nikulin, 1996) was used to verify the dependency between both information sources at the regional scale. The H_0 hypothesis was that yield loss estimation was independent of maximal temperature recorded during the heatwave event. H_0 was rejected for *p*-values of 0.05 ($p < 0.05$).

Both estimated yield loss and temperature values are continuous variables. In order to account for inaccuracy and short-range variability, both variables were converted into classes to carry out the test. For yield loss estimations, the width of the classes was defined from the data distribution at the regional scale, i.e. it was based on the distribution of predicted values (Scott, 2015). For the temperature values, a theoretical criterion was applied. Fraga et al. (2020) stated that during critical periods of vine development, e.g. in the growth period, if the air

temperature reaches a threshold above 35 °C, a limitation of photosynthesis is to be expected, leading to a decrease in productivity. Therefore, it was decided to set a threshold temperature at 35 °C and to establish two classes above and two classes below this threshold. The widths of the 4 classes were determined by (1) the temperature threshold and (2) the minimum and maximum temperature recorded on the 28th of June 2019.

Model interpretation

As for the classical Partial Least Squares (PLS) method, the three-way PLS model aims at finding new components called latent variables (LVs), which best relates data $\underline{\mathbf{X}}$ (samples × time × wavelengths) to **y** (ground truth data) (Abdi, 2010). Compared to classical PLS, the three-way PLS allows the information supported either by the time dimension or the spectral dimension to be kept and analysed properly. The weight vectors of each LV correspond to a spectral and temporal profile, providing evidence on the spectral bands and their dynamics over time that may best explain the yield losses. The most relevant LVs (i.e. those that best explained yield loss) were then selected and analysed to identify known information supporting the model performance, in addition to new complementary knowledge provided by the time series of images to characterise the heatwave. The standard error of prediction (SEP) (Keller et al., 1994) was used to determine the LVs to be analysed.

In addition to the LV weight vectors, the score of each of the 4978 blocks was calculated for each LV. The score value of a block shows its relation to the spectro-temporal profile defined by the LV. Then, 3 cases were considered:

- if the temporal-spectral profile of a block followed the same signature as the one created from the weight vectors (temporal and spectral bands) of a LV, the score value was positive;
- if the temporal-spectral profile of the sample (vineyard block) followed the inverse signature to the one created from the weight vectors (temporal and spectral bands) of the LV, the score value was negative;
- if the temporal-spectral profile of the sample (vineyard block) followed a different signature to the one created from the weight vectors (temporal and spectral bands) of the LV, the score value was close to zero.

4.2.2.4 Mapping and spatial analysis

Maps were obtained using point kriging interpolation. Kriging was performed with the GeoFis 1.0 software (Leroux et al., 2018), which was used for: (1) the modelling of semivariograms and calculations of their featured parameters, C_0 (nugget effect), C_1 (sill) and

r (range), and (2) the kriging interpolation. The latter was performed on a grid of regularly spaced points 1000 m apart within the geographical boundary of the LR region.

Semivariogram features were also used to compute the Cambardella Index (I_c) (Cambardella et al., 1994) (Equation 4.2):

$$\mathbf{I_c} = \frac{\mathbf{C_0}}{\mathbf{C_0}+\mathbf{C_1}} \quad \text{(EQ. 4.2)}$$

where C_0 is the nugget effect and C_1 is the sill of the semivariogram model. The Cambardella Index was considered here to quantify how the data were organised spatially over the LR region. The common following thresholds were then used: (1) Ic less than or equal to 25 %, the distribution is considered strongly spatially organised; (2) for Ic between 25 and 75 %, the distribution is considered moderately spatially organised and; (3) if Ic is higher than 75 %, the distribution is considered weakly spatially organised (Cambardella et al., 1994; Martínez & Gomez-Miguel, 2017).

4.2.3 Results

4.2.3.1 Quality of the N-PLS model

The performance and quality of the N-PLS model with five latent variables are presented for the calibration set (Figure 4.20.a) and validation set (Figure 4.20.b) in terms of R^2, bias and standard error of prediction of yield losses. The N-PLS model showed a performance accuracy (R^2) of 0.56 in the calibration set and of 0.66 in the validation set, with a standard error of cross-validation in the calibration set of 12.3 % and a standard error of prediction of losses in the validation set of 10.7 % (Lopez-Fornieles et al., 2022). A standard error of 10.7 % over the prediction set (Figure 4.20.b) was consistent with the accuracy level of the ground truth data, which gave the yield loss in 25 % classes (Section 4.2.2.1).

These results prove the relevance of multispectral satellite time series to assess the incidence of a heatwave on grape vine loss when combined with the N-PLS. However, this model was only validated on a small number of vineyard blocks spread over a small representative part of the region. The next sections aim to verify whether the model derived from this small dataset could be applied to the whole regional level in order to verify its ability to highlight heatwave footprints at this scale.

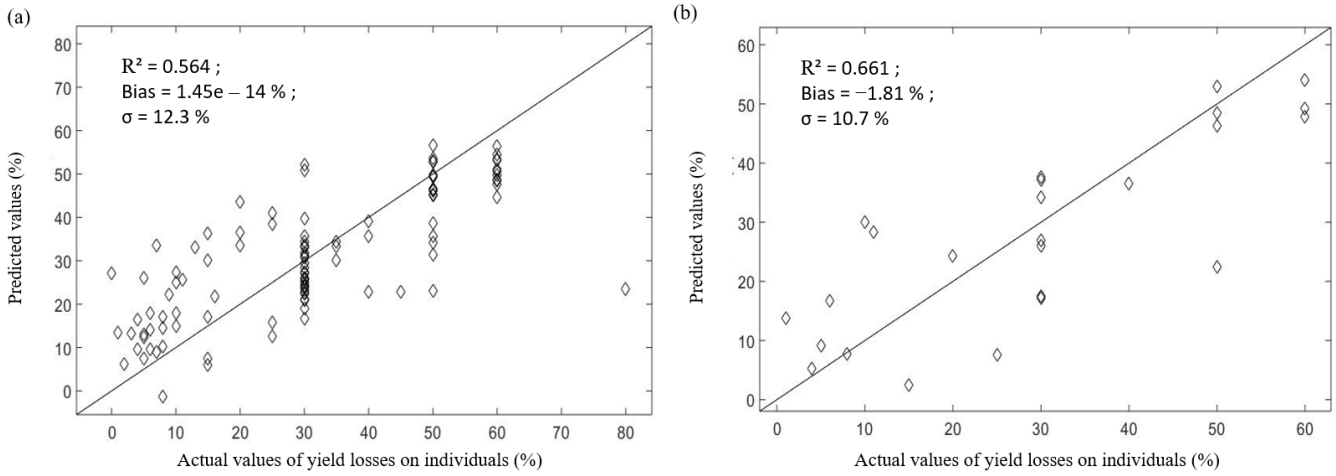


FIGURE 4. 20 - Results of the N-PLS prediction yield losses on individuals in the calibration set (a), with 80 vineyard blocks and in the validation test (b), with 27 vineyard blocks (Lopez-Fornieles et al., 2022).

4.2.3.2 Yield loss prediction at the regional level

Figure 4.21 shows the distribution of the yield loss prediction when the N-PLS model calibrated with the 107 vineyard blocks is applied to 4978 vineyard blocks of the region. From the mean (μ) and the standard deviation (σ) values of the distribution, four classes of yield loss were defined as follows: predictions between 0 and 15 % yield loss represented a low impact (*a*), predictions between 15 and 27 % of losses represented a moderate impact (*b*), predictions between 27 and 40 % of losses represented a high impact (*c*), and predictions > 40 % represented a severe impact of heat stress (*d*).

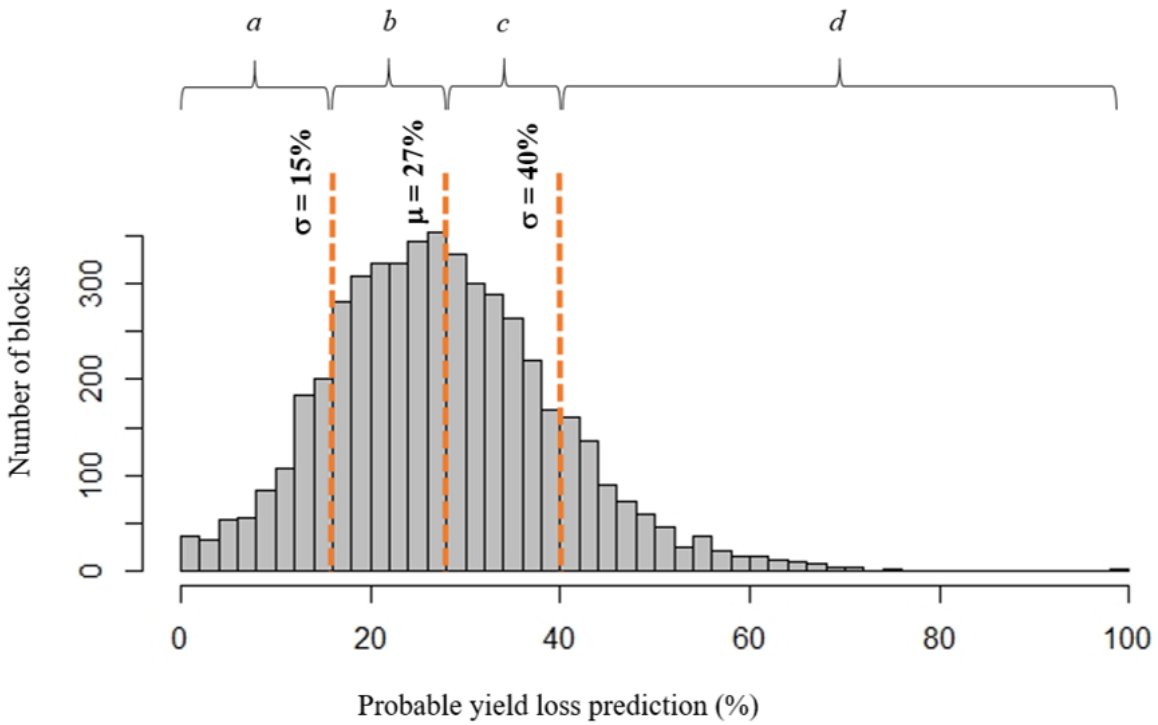


FIGURE 4. 21 - Histogram of the prediction of yield loss for the 4978 vineyard blocks. The vertical orange dashed lines highlight the mean (μ) and the thresholds corresponding to μ +/− σ (σ standing for the standard deviation). The 4 classes from a to d defined on these thresholds were linked to the degree of the impact of the heatwave, with a being the class with the lowest impact and d the class with the highest impact.

Figure 4.22 shows the kriged map of the yield loss prediction at the regional scale. The kriging was performed with a Gaussian semivariogram model, with a nugget effect (C_0) of 122, a sill (C_1) of 175 and a range of 140 km. The Cambardella index (I_c) resulting from this semivariogram was 41 %, which highlighted a (moderate) spatial organisation of yield loss predictions, i.e. 41 % of the variability exhibited a spatial structure while the remaining variability (59 %) was stochastic in nature.

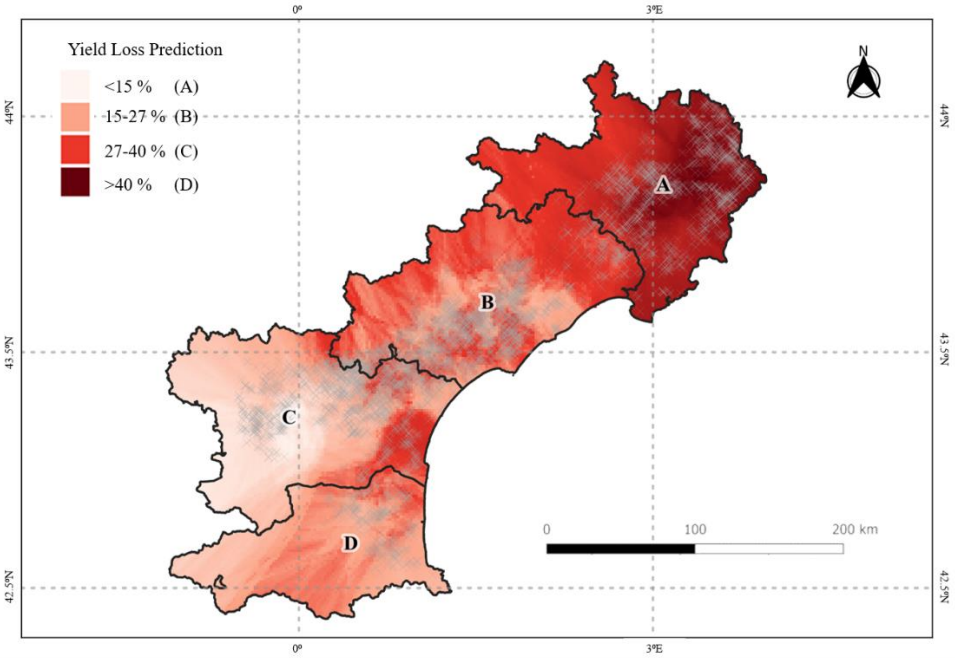


FIGURE 4. 22 - Kriged map of yield loss predictions at a regional scale derived from the N-PLS model.

The kriged map (Figure 4.22) shows large spatial patterns that are more or less related to patterns observed on the maximal temperature map (Figure 4.15). Indeed, it highlights two main zones, one in the northern part (A and B) corresponding to a high yield loss in the majority of the area, and the second in the southern part (C and D), with a lower yield loss. This main trend was clearly in relation with the maximal temperature map (Figure 4.15), which exhibited the same patterns. The kriged map also exhibited local patterns showing some local variations in yield loss. This was clear for sectors B and C, where smaller zones of yield loss were clearly highlighted.

In order to validate the qualitative observations made from the kriged map, a quantitative analysis of the results was proposed. Given the difficulty of obtaining reliable ground truth data at this scale, the proposed validation was based on the analysis of a contingency table between the yield loss classes with the maximum temperature observed on the day of the heatwave (Table 4.3). The same classes of yield losses were considered as for the kriged map.

TABLE 4. 3 - Contingency table cross-referencing the yield loss prediction classes with the maximum temperature classes recorded on 28th June 2019.

	(30–32.5 °C)	**(32.5–35 °C)**	**(35–40 °C)**	**(40–45 °C)**
Low impact: (0–15 %)	333	0	0	0
Moderate impact: (15–27 %)	417	552	0	0
High impact: (27–40 %)	0	1289	704	0
Severe Impact: (40–80 %)	0	0	749	934

Table 4.3 shows a clear relationship between the variables, and this relationship was clearly positive, i.e. the number of blocks impacted by the heatwave increased with the maximal temperature at the regional scale. The H_0 hypothesis (independence of data distribution within the contingency table) tested with a classical chi-square test was rejected (p-value < 0.01). Therefore, classes of estimated yield losses were significantly related to the heatwave at the regional scale.

Note that the results obtained are very specific. The model did not estimate a significant loss of yield when the recorded maximum temperatures were relatively low (<32.5 °C). On the contrary, when the recorded maximum temperatures were very high (>40 °C) the model estimated very high yield losses. Table 4.3 also highlights some inaccuracy between yield loss classes; for example, the 32.5–35 °C temperature class can lead to moderate and large predicted yield losses. This observation should be considered in the light of the definition of the classes, which were only roughly defined on the basis of the overall distribution of the data and which would certainly merit from adjustment according to the zone of the region, or according to the grape varieties, training systems, management practices, etc. It should certainly be considered in the light of the important nugget effect, which may be explained by a high inter-block variability (grape variety, block aspect, micro-topography, etc.) when working at this scale.

These results demonstrated the relevance of the model (derived from N-PLS) when applied at the regional level to predict yield losses associated with the heatwave of 28th June 2019. Based on these insights, the kriged map presented in Figure 4.22 may represent a relevant spatial footprint of the heatwave impact of grape yield losses at the regional level.

4.2.3.3 Insights of time series and spectral analysis of the N-PLS model

Table 4.4 shows the standard error of prediction (SEP) of yield loss for each LV derived from the N-PLS (Davies and Fearn, 2006). LV3 and LV4 were the LVs that best predicted the yield loss (lowest SEPs). Therefore, the next section focuses on studying, first, the LV4 that presented the lowest SEP (−0.04), and, in a second step, the LV3 that showed the second-lowest

SEP (−1.06). The LV4 and LV3 were analysed thereafter regarding their relevance towards the yield loss estimation.

TABLE 4. 4 - Standard error of prediction (SEP) for each of the five latent variables derived from the N-PLS model on the calibration set.

LV1	LV2	LV3	LV4	LV5
1.52	1.20	−1.06	−0.04	−1.81

Figure 4.23.a presents the weight vectors of LV4 as a spectro-temporal profile, i.e. for each date. The weight vectors can be viewed as the bands that most impact the vine response in relation to the heatwave. It showed several interesting patterns: (1) for reflectance between 800 and 1000 nm, high weights were observed at the beginning of the season (May), whereas low (negative) weights were observed right after the heatwave (5th of July), and the weights decreased again until the end of July and then increased until the end of August; (2) for the reflectance between 1600 and 2200 nm, the opposite trend was observed, with low weights (negative) at the beginning of the season and high weights after the heatwave (5th of July), and the highest weights were observed in a few weeks (20th July) after the heatwave. Reflectances between 750 and 1350 nm are known to be strongly related to leaf structure (Laroche-Pinel et al., 2021a), whereas reflectances between 1350 and 2200 nm are strongly linked to water absorption (Chen et al., 2005). Both of these ranges of reflectance were strongly impacted from immediately after until three weeks after, the heatwave. Therefore, the spectro-temporal profile of LV4 summarised the dynamic of the incidence of the heatwave on the vineyards' canopy. The LV4 weights showed that this may result in a drastic change in canopy structure partly due to a change in water content a few weeks after the event. Note that LV4 weights also highlighted the slow recovery of the canopy after 20th July for these wavebands.

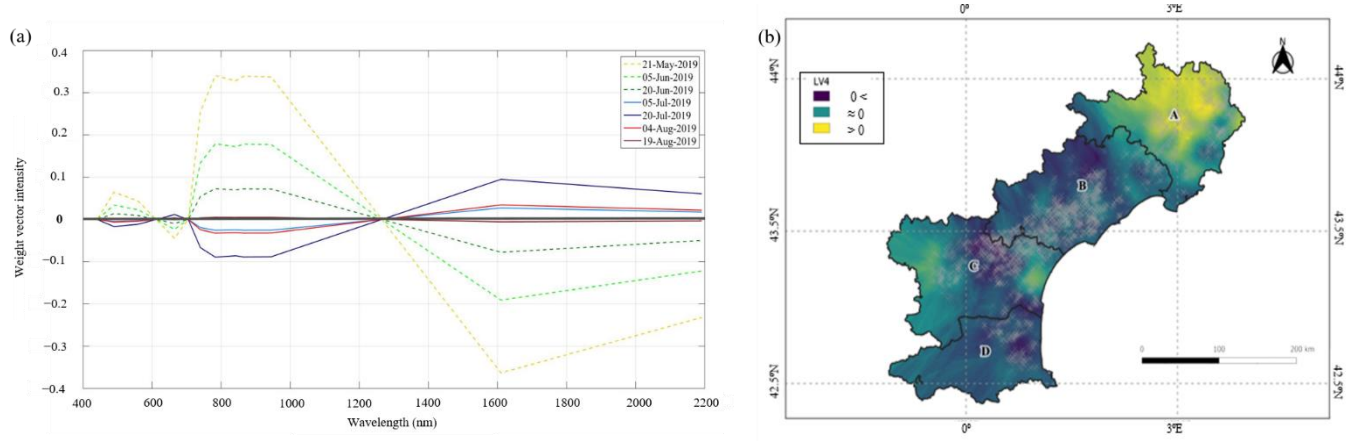


FIGURE 4. 23 - (a) LV4 weight vectors from N-PLS applied to the calibration set (75 % of the 107 vineyard blocks) for each date; (b) kriged map of the LV4 score values for blocks at the regional scale. The most negative score values of the blocks on LV4 (b) are shown in blue, and the most positive score values in yellow.

The score values (Figure 4.23.b), represented the agreement between the LV4 weights and the spectro-temporal profiles of each block. The score map obtained after kriging describes the spatial organisation of the blocks that present: (i) similar spectro-temporal profile (score > 0); (ii) opposite spectro-temporal profile (score < 0); and (iii) no related spectral-profile (score ~0). The score map for LV4 was supposed to highlight blocks that were directly impacted (or not) by the heatwave. The spatial organisation of the scores was confirmed by the semivariogram model, which showed (Table 4.5) that around 50 % of the variability was explained by a spatial phenomenon (I_c = 48 %). The spatial patterns were strongly related to the maximal temperature recorded on 28th June, showing that the blocks of the northern part of the region were in agreement with the LV4 spectro-temporal profile. Note also that, for other sectors, patterns were less related to the main trend of the heatwave. Blocks with a score > 0 underwent a drastic change in canopy structure, partly due to change in water content a few weeks after the heatwave. Patterns of strongly impacted blocks can also be seen in sectors B and C. Note, however, that for sectors other than sector A, the patterns were less related to the main trend of the heatwave.

TABLE 4. 5 - Semivariogram parameter descriptors and spatial variability index for score values A_1 (Range), C_0 (Nugget), C_1 (Sill) and I_c: Cambardella Index.

Latent Variables	Semivariogram Model	Range (km)	C_0	C_1	I_c (%)
Scores LV3	Spherical	17	0.011	0.019	29
Scores LV4	Gaussian	27	0.003	0.004	48

LV3 showed an unusual trend with regard to the common evolution of the vine canopy over the season (Figure 4.24.a). It had the highest weights (positive or negative) at the end of the season, from the July, and especially in August, when growth has usually stopped and canopy reflectance should be relatively stable (until senescence onset). These high weights corresponded to (1) the 664 nm reflectance (negative weights), (2) reflectances between 800 and 1000 nm (positive weights), and (3) to a lesser extent, reflectances between 1600 and 2200 nm (negative weights). They corresponded respectively to (i) the visible spectrum (400–700 nm), which is affected mostly by photosynthetic pigments content (chlorophyll and carotenoids) (ii) leaf and canopy structure (750–1350 nm) and (iii) water content (1350–2200 nm) (Gates et al., 1965). The highest ranges of reflectance (positive or negative) reached their maximum long after the heatwave episode (between 20th July and 19th August). Thus, LV3 weights may represent vines that were strongly affected by the extreme weather episode and that recovered their photosynthetic capacity of part of their leaf canopy later in the season (Teskey et al., 2015).

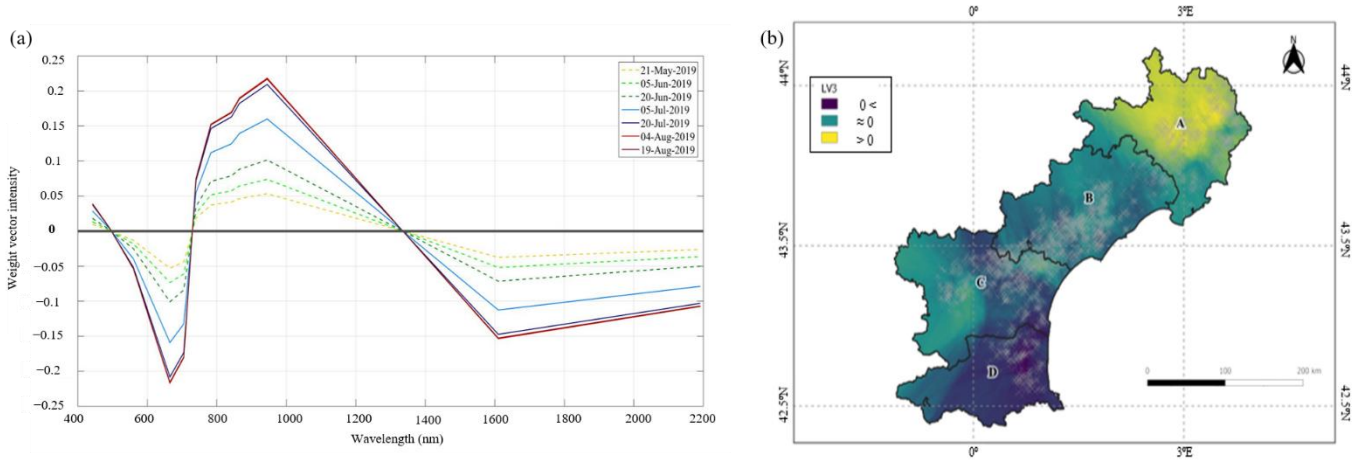


FIGURE 4. 24 - (a) Representation of LV3 weight vectors from N-PLS applied to the calibration set (75 % of the 107 vineyard blocks) for each date; (b) kriged map of the LV3 score values for blocks at the regional scale. The most negative score values of the blocks on LV3 (b) are shown in blue, and the most positive score values in yellow.

The score map for LV3 (Figure 4.24.b) was expected to highlight, on the one hand, blocks showing biomass growth after the heatwave (score > 0) and, on the other hand, blocks showing the opposite phenomenon (score < 0), i.e. an increase in photosynthetically active biomass in spring (before the heatwave) and a stabilisation or even a decrease in August (which should correspond to the expected behaviour of vines in normal conditions). The spatial organisation of the scores was confirmed by the semivariogram model used to krige the scores (Table 4.5), with around 70 % of the variability explained by a spatial phenomenon (I_c = 29 %). The spatial patterns showed a distinct north–south difference in the LR region, which again

corresponded to the maximum temperature patterns recorded on 28th June 2019, and thus to the vineyard blocks most affected by the extreme event.

4.2.4 Discussion

This study demonstrated the potential of temporal series of multispectral remote sensing images to discriminate and characterise the impact of a heatwave on vineyards at a regional scale. It also demonstrated that the high dimensionality (temporal and spectral) of the data required adopting a systemic methodology that accounted simultaneously for the spectral and temporal characteristics of the considered data. The N-PLS (3-PLS1 in our case) may be considered as a relevant approach to handle such problems. It allowed the spectral and temporal dimension of the data to be considered simultaneously in order: (i) to calibrate a model of prediction and (ii) to keep information captured by the spectral and the temporal dimensions through latent variables, which provided insights into the changes undergone by vine canopy.

The approach, although calibrated on a few fields, was successfully applied at the regional level, showing the robustness of the methodology and its ability to map the spatial footprint of the heatwave that affected the south of France in 2019.

It is essential to place the results presented in this paper within the reality for many environmental and agricultural studies where the ground truthing remains weak and hard to manage. The N-PLS model shows the interest in successfully dealing with a low number of ground truth samples (in this case, 107 yield loss observations for grape fields). Nevertheless, it should be noted that the N-PLS model here was still specific to the learning base used for the calibration. Specifically, the model accounted for the timing of the heatwave that the dataset, used for the calibration, had undergone. The direct application of the model to other vintages or to other regions should then not be considered. Despite this limitation, the approach allowed the identification of spectral changes in canopy reflectance that may be the signature of an early summer heatwave. This signature can be applied, with experts set up, to other case studies in order to identify potential heatwave effects when no ground truth data are available. Another limitation is that the model integrated the whole season dynamic (from May to August). As a result, it is not suitable to be used as a monitoring model to identify and spatialise the heatwave effects in real time or even a few days after the event.

Although this study focused on the assessment of grapevine heat stress at the regional level, the generation of potential knowledge from a multispectral time series with the intention of understanding the phenomenon in question (i.e. a heatwave), was achieved through the N-PLS approach. Previous studies have shown significant abilities to assess the effects of heatwaves (Webb et al., 2010; Cogato et al., 2019b) by evaluating the physiological and spectral responses of grapevines. Notwithstanding this, the adoption of a PLS-multiway analysis

presents the advantage of accounting for changes in the spectral responses over time. Indeed, the two latent variables most representative of yield loss in the model (SEP values of −1.06 for LV3 and −0.04 for LV4) provided knowledge regarding heat stress in vineyards by means of weight and score values. LV4 defined the spectral response of the vineyards to thermal stress, marking a clear temporal evolution where all the spectral information was affected, thus reversing the 'theoretical' vegetation profile. LV3 reported on the period after the heatwave. In this case, the spectro-temporal profile of LV3 generated insights related to the unusual vegetation growth observed in mid-August. A possible explanation for the late recovery of greenness could be the experience of the vineyards after the heatwave, as the north was the area most affected by the extreme climatic episode. However, the interpretation of the other LVs in relation to the heatwave remains more challenging. Indeed, the N-PLS aimed to extract LVs that best explained the output variable. For that purpose, it may generate LVs that first model general trends to better extract specific phenomena. As a result, the first LVs may be more related to the evolution of a 'standard' vegetation canopy signature in viticulture. The analysis of LVs as proposed in this study remains difficult and requires further understanding of both the N-PLS approach and the phenomena under study.

The validation of the model at the regional scale was performed based on the maximal temperature observed on 28th of June 2019. The use of only maximum temperature is a relatively simple approach for a very complex phenomenon. There is still some debate about how the impact of a heatwave on vineyards should be assessed and which factors need to be taken into account, such as the onset dates, the duration of the heatwaves (Fraga et al., 2020), and humidity and the resulting vapour pressure deficits (De Boeck et al., 2010). Similarly, summarising the effect of a heatwave to a yield loss, as was done in this study, is certainly reductive. Other response variables (vigour, chlorophyll content, etc.) would certainly have been interesting to consider and more directly related to remote sensing variables. However, from an operational point of view, yield loss was an integrative response of the vine plant and, moreover, important for the wine industry. Nonetheless, the statistical analysis of the relationship between the maximum temperature recorded on 28th June 2019 and yield loss predictions (p-value ≤ 0.001) showed that the model may have captured the main trend of the heatwave impact on yield at the regional scale.

At this regional level of analysis, it is unclear whether the small spatial patterns observed in Figure 4.22, on sectors B and C, are indeed local variations of the impact of the heatwave on the vine. Given the strong spatial structure observed, the results may support this hypothesis. Indeed, spatially structured environmental factors, such as soil type, elevation, aspect, etc. can explain local variations in heatwave characteristics (duration, maximum temperature, etc.), which would explain the local variations observed in predicted yield losses. It would be interesting to validate this hypothesis because, if it proves to be correct, the use of image time

series associated with a fine mapping of meteorological conditions may be a powerful tool to better characterise the effect of meteorological conditions during a heatwave.

At the regional scale, a high nugget effect was observed, indicating that a significant proportion of the variance remained independent of spatially organised factors or was explained by very short-range phenomena. Other factors undoubtedly affected the change in spectral response of the vine canopy, such as the variety, the training systems and the management practices. For the record, the LR region presents a wide variety of rootstocks, cultivars and clones, in addition to different soil characteristics and training systems. Since the model was applied without considering these factors, they may explain the strong variability that was observed at very short ranges from one field to another. It is difficult to know if these factors affected the spectral response of the canopy or if they locally mitigated/amplified the effect of the heatwave as it was observed from remote sensing or both. This interrogation promotes interesting questions to study the potential of the temporal series of multispectral images to better study how field characteristics may drive the response to heatwaves (Schaffer & Andersen, 2018; Venios et al., 2020).

The applied multidirectional approach (N-PLS) presented here represents a specific case of viticulture for the heatwave of 28th June 2019 in LR. In general, though, it is a type of approach that can be effective in characterising and assessing the impact of extreme weather events that suddenly affect the spectral response of the vine canopy, e.g. hail and frost, with a distinct, disrupted temporal evolution. However, its application to more gradual phenomena, such as progressive changes in water status or nutritional problems, may be less adapted as changes in plant characteristics may be less obvious.

4.2.5 Conclusions

This study demonstrated how, with a proper dimensionality reduction algorithm such as the N-PLS, a time series of multispectral images can provide an estimation of the impact of a heatwave on vineyard blocks at a regional level. The methodology proved to be relevant to provide the spatial footprint of the heatwave through its evolution over time by means of the observed response of spectral information.

The relationship between the percentage yield losses and the maximal temperature recorded on 28th of June 2019 at the regional scale was shown. Insights into the phenomena explaining canopy responses were provided from the spectro-temporal signatures of the latent variables, showing the potential of the approach to provide knowledge on canopy changes during such an event. The main limitation of the proposed methodology was its necessary calibration on the spectral temporal signature of the event under study. This prevented any application of the calibrated model to other heatwaves whose timing would be different.

The proposed methodology is potentially transferable to other phenomena that evolve over time and, in particular, to any sudden climatic event that may affect the growth dynamics and leaf composition of the vine canopy (frost, hail, plagues to some extent, etc.). However, the approach seems less suitable for more gradual phenomena, such as plant water status, as the temporal evolution is less evident.

Further research is needed to identify and characterise the effects of factors affecting the specificity of the spectral temporal response of vine canopy towards a heatwave. This should provide a new methodology to better analyse incidences of heatwaves on canopy responses at a large scale, and the potential mitigating or amplifying effects, such as microclimate, topography, training systems and variety.

Conclusion of Chapter 4

Multispectral time series images are multidimensional massive datasets that have shown a great potential for different kinds of applications, including vegetation and crop monitoring (Lee et al., 2020). However, due to its high-dimensional nature and the high variability of the spectral information over time, the challenge remains to extract meaningful information from these data. The objective of this Chapter 4 was to test whether the N-PLS method and the subsequent model developed can be suitable for the handling of multispectral time series images. The N-PLS method was derived from a classical statistical method (PLS) adapted to multi-way data that explicitly considered the geometric structure of these data via a dimensionality reduction process.

This approach has two main assumptions (multicollinearity and linear relationships between the independent variables) that should be mentioned since in recent years, machine learning techniques have proposed alternative methods to avoid being affected by these two assumptions. However, a main problem with machine learning models, such as deep learning, is limited model interpretability and the necessity to have large data base to perform properly the learning step (Mishra et al., 2021).

The first part of the chapter (4.1) aimed to show the potential of the N-PLS multi-way approach for the analysis of Sentinel-2 time series for the detection, characterisation and prediction of the effect of extreme weather events on an agricultural attribute of interest, in this case the impact of a heatwave on vineyard yield at the regional scale. A latent variable model, such as the one obtained by N-PLS, provided: (i) a low-dimensional model for high-dimensional data to better understand an evolving phenomenon over time and (ii) an interpretable and causal model through the *b*-coefficients. Through the *b*-coefficient, it was possible to identify that the spectral bands around the red edge (700 nm) and the SWIR region (1600 nm) were relevant for characterising the heat stress period in relation to reported yield losses. The case study presented, using 107 selected vineyard blocks, showed the importance of taking into account both spectral and temporal information collected by remote sensing data, while demonstrating the validity of the N-PLS model because of its ability to provide knowledge beyond just its predictive accuracy.

As the multi-way N-PLS approach has proven to be of interest when dealing with phenomena with a clear temporal evolution, such as the dynamics of regional vegetation growth, phenological identification of crops or the impact of crop stress, the next step shown in the second part of the chapter (4.2) was its application to characterise spectrally, temporally and spatially the impact of the heatwave at the regional scale (4978 vineyard blocks). This is reasoned by the fact that the multidirectional N-PLS approach has also proven to be able to

generate an interpretable and applicable model at a large scale (regional scale) with a very limited learning base, which makes it a particularly effective approach for this type of application. First, vine heat stress was assessed at the regional level, with a validation of the model at the regional scale based on the maximum temperature observed on the 28 June 2019. Although the validation was rather limited in view of the limited ground truth data at this particular spatial scale, its value and interest for this type of application was demonstrated. Secondly, the scores (spatially mapped with geostatistics) and loadings (representing spectro-temporal signatures) generated as model outputs for LV3 and LV4 (lowest SEPs) were used as key parts of the chemometric modelling, as they facilitated the interpretation of the N-PLS method by providing information on the spectral-spatial relationships of the vine canopy over time.

Some of the concluding perspectives of both articles revealed the interest of further investigating the specificity of the temporal spectral response of the grapevine canopy to a heatwave. The next chapter of the manuscript continues this line of research on the specificity of the spectral-temporal response based on a multi-way variable selection approach.

Chapter 5. Determination from multispectral time series data of the most discriminant spectral and temporal domains for grapevine vegetative growth's characterisation in relation to an extreme weather event

Introduction of Chapter 5

The previous chapter (Chapter 4) has shown the interest of the application of a supervised chemometric method that considered the spectral and temporal dimensions of remotely sensed imagery simultaneously. By integrating the temporal dimension into the multi-way N-PLS model, it was possible to explain the relationship between the different spectral domains (Red-Edge and SWIR) and the impact of the heatwave on the canopy, as well as to determine certain spectral-temporal signatures that explain heat stress and its spatial footprint. Thus, it was possible to determine that an event occurred and also when it happened. However, it was not possible to determine the possibility of identifying other factors over time that could have enhanced (or not) the impact of the heatwave on yield loss. At a more operational level, it is important to specify in more detail which spectral and temporal domains were affected by changes in vegetation growth during this period of stress. In a broader sense, it is necessary to know the impact of an extreme weather event on spectral signatures and their temporal response, in order to identify key risk factors for future monitoring. Although the direct impact of a heatwave on vineyard vegetation is already well studied in the literature (Fraga et al., 2020), the reason for its variable repercussions on a large scale, i.e. the reason for the different % yield loss depending on the vineyard block, is still a field to be explored. To this end, in this chapter it is proposed to deepen the spectral and temporal domains using a multi-way variable selection method, the N-way Covariance Selection (N-CovSel) (Biancolillo et al., 2022). The study uses this supervised chemometric approach to implement multispectral satellite feature extraction in order to characterise the spectral bands or spectral domains, as well as the determining dates or periods to understand and identify differences in yield loss within 107 vineyard blocks affected by the heatwave that occurred between the 23rd of June and the 8th of July 2019 in the LR region. It is assumed that the spectral, temporal or spectro-temporal variables selected by the N-CovSel methodology may characterise the disparity in yield losses on a regional scale that can result from an extreme weather event in early summer.

A variable selection method to reduce the size of the dataset is an interesting approach as in certain cases, such as remote sensing time series, data can quickly reach a large size due to their high spectral and temporal resolution. With this kind of approach, it is possible to determine and set aside variables that lack relevant information in order to process the database more efficiently. Furthermore, the proposed N-CovSel methodology is able to manage variable selection, taking into account the inherent relationship between reflectance variations and the time windows in which they occur.

Article 5: Potential of N-CovSel for variable selection: a case study on time series of multispectral images

Published in *Frontiers in Analytical Science*, 2022, *2*. DOI: 10.3389/frans.2022.872646

E. Fornieles-Lopez[1,2], B. Tisseyre[1], A. Cheraiet[1], B. Gaci[1] and J.M. Roger[1,2]
[1]ITAP, Univ. Montpellier, INRAE, Institut Agro, Montpellier, France
[2]ChemHouse Research Group, 34000 Montpellier

Abstract: Multispectral image time-series have been promising for some years; yet, the substantial advance of the technology involved, with unprecedented combinations of spatial, temporal, and spectral capabilities for remote sensing applications, raises new challenges, in particular, the need for methodologies that can process the different dimensions of satellite information. Considering that the multi-collinearity problem is present in remote sensing time-series, regression models are widespread tools to model multi-way data. This paper presents the results of the analysis of a high order data of Sentinel-2-time series, conducted in the framework of extreme weather events. A feature extraction method for multi-way data, N-CovSel was used to identify the most relevant features explaining the loss of yield in Mediterranean vineyards during the 2019 heatwave. Different regression models (uni-way and multi-way) from features extracted from the N-CovSel algorithm were calibrated based on available heatwave impact data for 107 vineyard blocks in the Languedoc-Roussillon region and multispectral time-series predictor data for the period from May to August. The performance of the models was evaluated by the R^2 and the Standard Error of Prediction (SEP) as follows: for the temporal N-PLS model (R^2 = 0.62 - SEP= 11.4 %), for the spectral N-PLS model (R^2 = 0.61 - SEP = 13 %) and the temporal-spectral PLS model (R^2 = 0.63 - SEP= 11.7 %). The results validated the effectiveness of the proposed N-CovSel algorithm in order to reduce the number of total variables and restricting it to the most significant ones. The N-CovSel algorithm seems to be a suitable choice to interpret complex multispectral imagery by temporally discriminating the most appropriate spectral information.

Keywords: Multi-way, Feature extraction, Covariance Selection, Remote Sensing, Time-series, Grapevines.

5.1 Introduction

From the point of view of data visualisation and interpretation, multi-way analysis allows simplification of the results, providing more adequate and robust models using relatively few parameters (Salvatore et al., 2013). According to Henrion (1994), as information becomes more complex i.e. extremely diverse in terms of information, size and behaviour, the concept of a 'dataset' naturally expands from traditional tables, such as matrices, to higher-dimensional arrays. In fact, the use of multi-way analysis allows connected pieces of information to reflect variation spread across components, events or sources that are represented differently and, yet, complement each other in the simultaneously analysed data (de Juan & Tauler, 2019). Multispectral imaging (MSI) is a well-known imaging technique that has its origins in remote sensing. In practice, regardless of the applications, different approaches to deal with the increasing data volumes and variability of the data from satellite-based time-series imaging, such as Sentinel-2 (A/B), can be found in the remote sensing literature (Picoli et al., 2020). In recent years, spatial-spectral feature extraction has been a developing field of research managing high-dimensional data (Hong et al., 2020). However, in addition to spatial information, the Sentinel-2 satellites contain spectral information with 5-day revisit time, which provides a detailed overview of land and vegetation.

Multispectral imaging techniques applied to temporal series represent an important research tool to assess the impacts of Climate Change (CC) on agricultural systems as it allows spatially and temporally continuous phenomena to be monitored. The main abiotic factors in the life cycle of crops, especially during the growing period, are weather conditions, which determine the quantity and quality of agricultural production (Raza et al., 2019). One of the most measurable effects of CC is the gradual rise in temperature, which leads to an increase in the frequency and severity of extreme weather events (Droulia & Charalampopoulos, 2021). According to Venios et al. (2020) fluctuations in environmental conditions, particularly ambient temperature, strongly influence plant growth and development processes. As a result, remote sensing has the potential ability to assess the impact of an extreme weather effect, e.g. a heatwave, as the reflectance spectrum changes depending on growth circumstances and the time of measurement relative to the stage of crop development (Filella et al., 1995; Cogato et al., 2019b). However due to the complexity of combining spatial, spectral, and temporal information derived from remote sensing, there are still challenges in dealing with increased data volumes and variability of these data (Bishop, 2013). Making the most of multispectral image time-series is a promising but still relatively underexplored research direction in the context of life sciences.

The use of multi-way analysis in remote sensing, such as N-way Partial Least Squares (N-PLS) regression, shares all the advantages of latent-based regression and discrimination methods, i.e. data visualization and interpretation (Favilla et al., 2013; Lopez-Fornieles et al.,

2021). In addition, it allows the representation of data patterns, feature correlation and covariance structure characteristic over time-series images (Coppi, 1994; Favilla et al., 2013). When a two-dimensional signal characterises each sample, as generated by MSI, such as the wavelength/time information, it is often needed to define which are the most relevant features to predict the studied dependent properties. When it comes to deal with complex datasets, a generalised option is the selection of variables (or feature extraction) (Trevino & Falciani, 2006) as these methods allow to: i) select a relatively small number of total variables and restrict it to the most significant ones, i.e. for subsequent applications in regression/classification models and ii) to understand which variable contributes the most to the investigated system, i.e. interpretative purposes (Biancolillo et al., 2021). Several variable selection methodologies have been proposed in the literature (Mehmood et al., 2012), and yet most of variable selection methods refer to contexts in which data is collected in a matrix rather than a in higher-order structure, thus losing the multi-way analysis advantage (Favilla et al., 2013). However, Biancolillo et al. (2022) proposed an alternative variable selection approach for multi-way data, N-way Covariance Selection (N-CovSel). The N-CovSel algorithm is based on the same main principle as the covariance selection algorithm (CovSel) introduced by Roger et al. (2011) for data collected in data matrices. The latter approach is designed to select variables in regression and discrimination contexts, and to assess the relevance of variables based on their covariance with the response(s). Iteratively, the predictor with the highest covariance is selected and the data matrix ($\underline{\mathbf{X}}$) and the variable of interest to predict ($\mathbf{y}$) are orthogonalised with respect to this variable (Biancolillo et al., 2022). By providing filter selection based on model parameters and integrating them into the model construction, the N-CovSel algorithm opens the possibility to select information in a complex dataset as a multi-directional structure.

Regarding agricultural systems and CC impact with time-series of multispectral images, such as a variable selection approach for high order data arrays, could bring a better understanding of how crop growth dynamic is affected by the occurrence of an extreme weather event. Therefore, the objectives of this study are to:

i. propose a formalism to apply the N-CovSel approach to a time-series of images at the regional scale in order to predict a small variable of interest,
ii. show the value of methods originally developed in the analytical chemistry domain to be applied to larger scales and life sciences domains and,
iii. to identify the possible limitations of the approach when dealing with time series of satellite images.

The work is organised as follows: Section 5.2 introduces the proposed N-CovSel algorithm and the development of the model as well as the description of the case study that the methodology is applied to. The results are shown in Section 5.3, with the discussion in Section 5.4.

5.2 Materials and Methods

5.2.1 Notations

Upper case bold and underlined characters will be used for N-way arrays, e.g. $\underline{\mathbf{X}}$ (I,J,K) indicates a 3-way array with I samples described by J times at K wavelengths. Upper case bold characters will be used for matrices; e.g. $\mathbf{X}$ and lower case bold characters will be used for column vectors, e.g. $\mathbf{y}$. Non-bold italics will be used for scalars. Upper case characters for fixed values, e.g. the number of samples I and lower-case characters will be used for running indexes, e.g. a slice k from the third mode of $\underline{\mathbf{X}}$. A column of $\underline{\mathbf{X}}$ will be noted $\mathbf{x}_{.jk}$ and a slice of $\underline{\mathbf{X}}$ will be noted $\mathbf{X}_{j.}$ or $\mathbf{X}_{.k.}$.

The N-CovSel method allows the selection of the best set of predictors (features) in an N-way array ($\underline{\mathbf{X}}$) on the basis of its covariance with a response vector ($\mathbf{y}$) or a response matrix ($\mathbf{Y}$) (Biancolillo et al., 2022).

5.2.1.1 Definition of features

When selecting features in a N-way array, different solutions are possible. In fact, it is possible to define different features depending on the number of way arrays of the input data structure. As determined by Biancolillo et al. (2022), for a 3-way data, i.e. a cube, two distinct options are possible: i) a 2-D feature (Figure 5.1.a and Figure 5.1.b), i.e. a variable in one mode without discarding any variable in the other (e.g. a slice $\mathbf{X}_{.j.}$ or a slice $\mathbf{X}_{.k.}$) and ii) a 1-D feature (Figure 5.1.c), i.e. a single variable in each mode (e.g. the column $\mathbf{x}_{.jk}$).

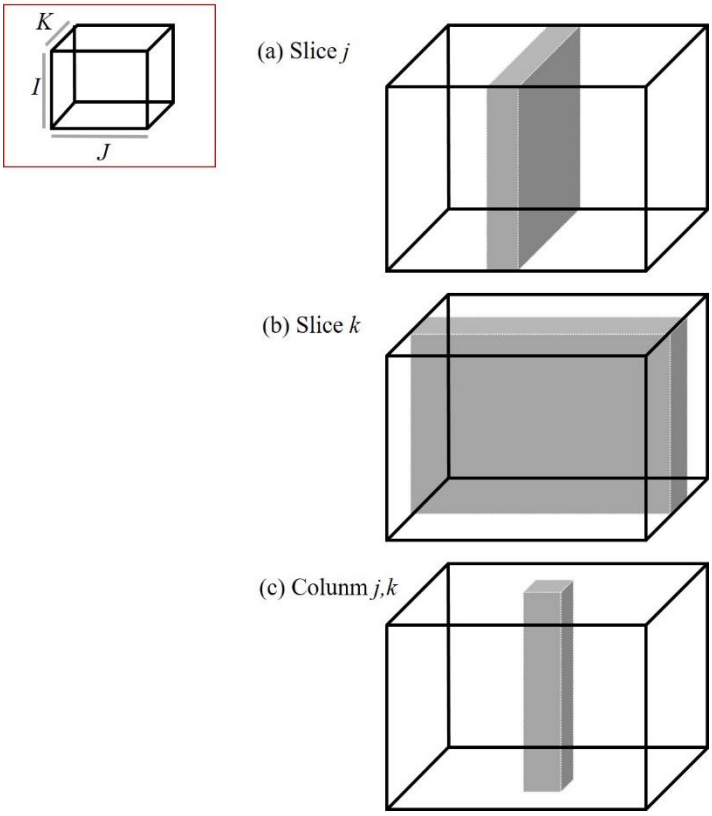


FIGURE 5. 1 - Features in a 3-way array represented in (a) J-axis slice, (b) K-axis slice and (c) J,K-column (Biancolillo et al., 2022).

5.2.1.2 Algorithm

N-CovSel algorithm is an extension of the above mentioned CovSel feature selection approach by Roger et al. (2011) to high-order data. To assess the relevance of features in a 3-way array (**X**) context to predict a response vector **y** relying on covariance, Biancolillo et al. (2022) defined the N-CovSel algorithm as follows:

1) Determine the structure of the features to be selected, i.e. columns or slices.
2) Define the number of features to be selected.
3) Select the feature of **X** with the highest squared covariance with **y**.
4) Deflate **X** of the information present in the selected feature.
5) Continue from Step 3 until the value defined in step 2 is reached.

5.2.2 Case-study

The Languedoc-Roussillon (LR) wine-growing area experienced a heatwave from the 23rd of June to the 8th July of 2019, with temperatures reaching 45°C on 28th June, 2019. Extreme weather events, such as a heatwave, occurring on very rapid time scales during crucial

periods of vine plant development (e.g. growing stage) will induce symptoms that may lead to stalled development, leaf burn and leaf drop (Schymanski et al., 2013; Lopez-Fornieles et al., 2022). According to Cogato et al. (2019a), remote sensing data could provide valuable information from spectral-temporal dimensions to characterise the impact of heatwaves on perennial crops by providing a detailed time series of data on the physiological and physical properties changes of the cultivation (Plant et al., 2000).

The N-CovSel algorithm should therefore constitute a relevant approach to create a model on a reduced set of information highlighting the extreme weather phenomenon taking into account its spectral-temporal evolution.

5.2.2.1 Ground truth data

Ground truth data were selected from 107 non-irrigated vineyard blocks in the northern part of the LR region that all showed some effects related to the heatwave (Figure 5.2.a). The severity of this effect was assessed by winegrowers and advisors on each of the 107 vineyard blocks by estimating the percentage of yield loss several weeks after 28th June 2019 corresponding to the peak of the extreme weather event. Severity was assessed several weeks later by estimating the percentage yield loss based on heatwave-related effects such as stalled development, scorching and leaf drop. It was acknowledged that it was sometimes difficult to attribute losses exclusively to the heatwave. Figure 5.2.b summarises the distribution of the 107 blocks according to yield loss.

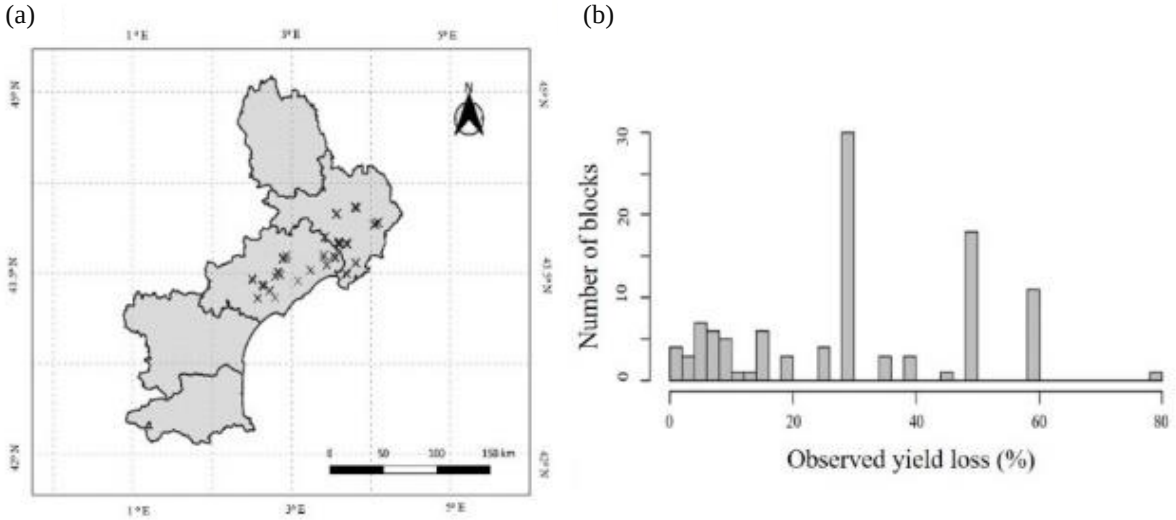


FIGURE 5. 2 - (a) Map of the 107-ground truthed blocks with known estimated percentage of yield loss after the heatwave and (b), percentage of yield losses observed by winegrowers and advisors on 107 vine blocks in southern France (Lopez-Fornieles et al., 2022).

5.2.2.2 Remote sensing data

Satellite images were selected via the Google Earth Engine (GEE) platform that provides Sentinel-2 L2A products. Sentinel-2 (A/B) satellites, with a revisit frequency of 10 days (5 days with the twin satellites together), provide 13 spectral bands from visible (Vis) to shortwave infrared (SWIR) with a spatial resolution of 10 m, 20 m and 60 m depending on the spectral band (Table 5.1) (Lopez-Fornieles et al., 2022). Spectral band 10 at 1380 nm was not used in this study as it is designed for the detection of visible and sub-visible cirrus clouds (Hollstein et al., 2016).

TABLE 5. 1 - Spectral bands for the Sentinel-2 satellite considered by the analysis.

Sentinel-2 Band	Central Wavelength (nm)	Bandwidth (nm)	Spatial Resolution (m)
Band 1–Aerosol	442.7	21	60
Band 2–Blue	492.4	66	10
Band 3–Green	559.8	36	10
Band 4–Red	664.6	31	10
Band 5–Vegetation Red Edge	704.1	15	20
Band 6–Vegetation Red Edge	740.5	15	20
Band 7–Vegetation Red Edge	782.8	20	20
Band 8–NIR	832.8	106	10
Band 8A–Vegetation Red Edge	864.1	21	20
Band 9 –VNIR	945.1	20	60
Band 11–SWIR	1613.1	91	20
Band 12–SWIR	2202.4	175	20

Images containing the study vineyards were selected and processed via Google Earth Engine (GEE) (Lopez-Fornieles et al., 2022). Images were selected over a period encompassing the heatwave event; from 13th May to the 20th August 2019. Before calculating the average pixel values for each block, each date and each waveband, in order to avoid mixed pixels: i) blocks boundary were extracted from the graphical parcel register of France (RPG) and ii) a 10 m inner-buffer was imposed over the boundary of each block (Lopez-Fornieles et al., 2022).

For the time period considered for the study (from May to August), defined as the most relevant period for monitoring vine growth vegetation in LR region (Devaux et al., 2019), 25 images should have been potentially available on each block. However, the number of images per block varied according to the local atmospheric conditions over each block for each acquisition date (Lopez-Fornieles et al., 2022). The number of available images for each block was 11 on average, being 8 the standard deviation of the set of values.

5.2.2.3 Modelling

Data array construction

To overcome the challenge of heterogeneity in the number of images per block, an interpolation was performed to obtain a continuous data cube $\underline{\mathbf{X}}$ ($I \times J \times K$). The interpolation at a date t was done wavelength by wavelength, by a convolution of the chronology measured with a Gaussian filter (Alam et al., 2008) in order to have a consistent time step dimension (J) between the 13th May and 20th August, 2019. The parameters involved in the interpolation setting were fixed to the Gaussian filter width (P) = 30 and date interval (N) = 5.

At the end of the interpolation step, the data set was meaningfully arranged in a three-way array $\underline{\mathbf{X}}$ of dimensionality 107 (samples, I) × 19 (times, J) × 12 (wavelengths, K) and a vector **y** (107), corresponding to the yield loss rates of the107 blocks.

Model calibration and validation

A calibration and validation subset were created to build and evaluate the model. Considering the samples from the variable to be predicted, a calibration set (3/4) and a test set (1/4) have been defined by its distribution (Figure 5.2.b), as follows Lopez-Fornieles et al. (2022):

1) The vector **y** was sorted in ascending order.
2) After sorting, every fourth individual was placed in the validation set and the others were kept in the calibration set.

At the end of this step, the data were therefore: i) a calibration set $\underline{\mathbf{X}}_c$ (I=80, J=19, K=12) and ii) a test set $\underline{\mathbf{X}}_t$ (I=27, J=19, K=12).

Regression model application

As explained in Section 5.2.1.1, when selecting features in a 3-way array, different outcomes of N-CovSel were obtained. Thus, the structure of the initial data was reduced in either time or wavelength slices (2-D features) or in columns (1-D features), i.e. date-wavelength coupling.

For the structure features (F) in 2-D, the number of best features in the calibration set was defined as follows:

- For the temporal slices (Figure 5.1.a), the number of features defined was F =15. Thus the F=15 dates were sorted in decreasing order of interest, providing a list of indices $\{j_1, j_2, \ldots, j_F\}$.

- For the spectral slices (Figure 5.1.b), as the total number of Sentinel-2 satellites wavelengths is 12, the number of features defined was F =12. Thus, the F=12 wavelengths were sorted in decreasing order of interest, providing a list of indices $\{k_1, k_2, \ldots, k_F\}$.
- For the structure features in 1-D (Figure 5.1.c), the number of features defined was F= 15. Thus, the F=15 date-wavelength coupling was sorted in decreasing order of interest, providing a list of pairs of indices$\{(j_1,k_1), (j_2,k_2), \ldots, (j_F,k_F)\}$.

Once the N-CovSel algorithm had selected the variables' relevancy on the basis of their covariance with the response(s) (Biancolillo et al., 2022), a regression model adapted to the reduced data set was applied. Depending on the structure of the selected features, different data analysis strategies can be applied. In the case of 2-D, as the feature selection is of higher order, features have been analysed using multi-way approach, whereas in the case of 1-D the most intuitive option was to combine them into a matrix, and then applying a traditional chemometric approach (Biancolillo et al., 2022), as follows:

- For the temporal slices, F= 15 N-way Partial Least Squares (N-PLS) models (Bro, 1996) were then calculated on the calibration set, using the slices $\{j_1\}$, $\{j_1, j_2\}$, …, $\{j_1, j_2, \ldots, j_F\}$.
- For the spectral slices, F= 12 N-PLS models (Bro, 1996) were then calculated on the calibration set, using the slices $\{k_1\}$, $\{k_1, k_2\}$, …, $\{k_1, k_2, \ldots, k_F\}$.
- For the columns (date-wavelength), F= 15 Partial Least Squares (PLS) models (Wold et al., 2001) were then calculated on the calibration set, using the columns $\{(j_1,k_1)\}$, $\{(j_1,k_1), (j_2,k_2)\}$, …, $\{(j_1,k_1), (j_2,k_2),\ldots, (j_F,k_F)\}$.

Figure 5.3 summarises the workflow of the N-CovSel model calibration, and its implementation for a regression model according to the structure of its outcomes (slice or column).

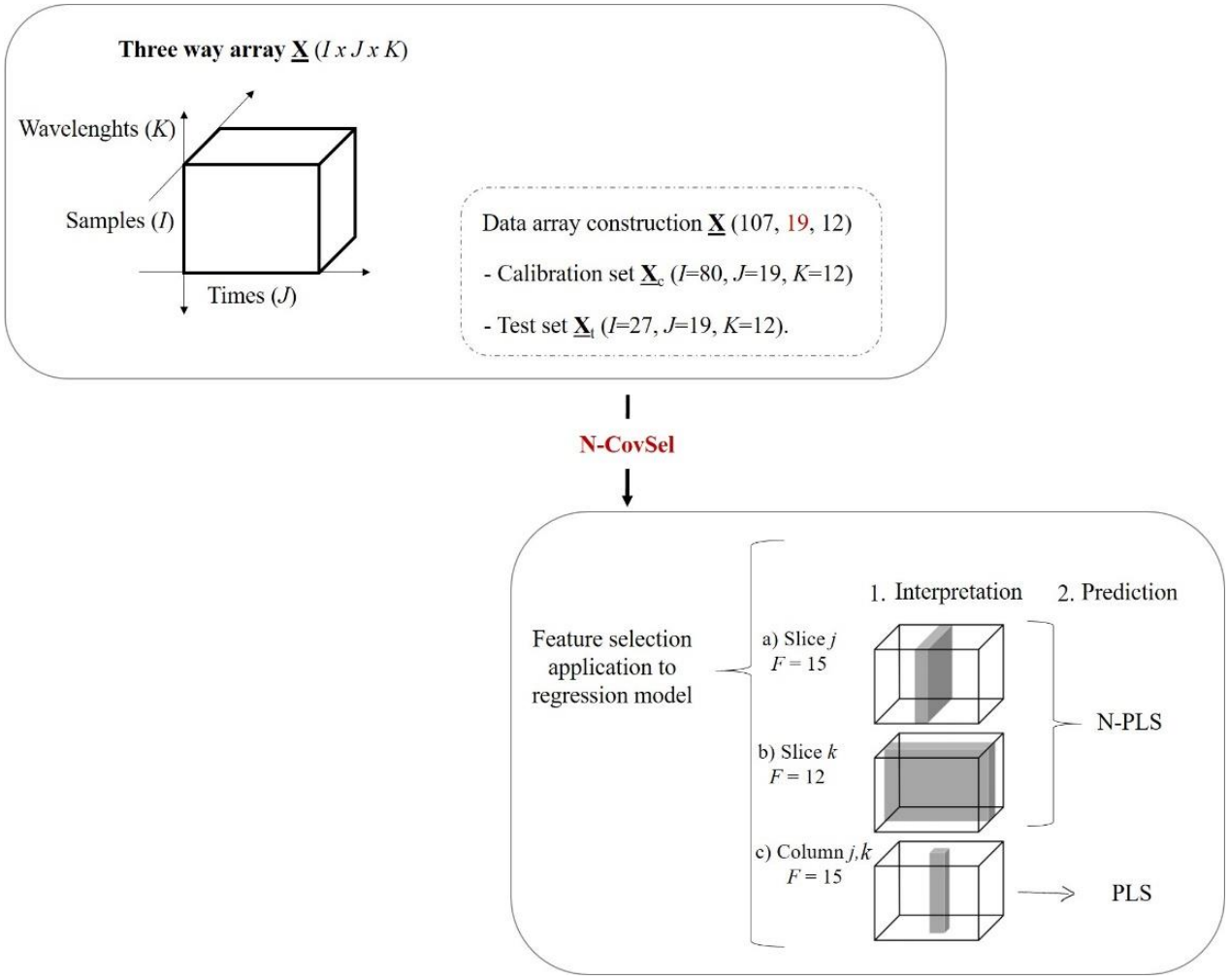


FIGURE 5. 3 - Workflow diagram of the N-CovSel model calibration and the suitable choice of the regression method according to the structure of the features selected by the algorithm.

Model evaluation

For each regression model calculated (N-PLS and PLS), a Standard Error of Calibration (SEC) was calculated, using the maximum number of latent variables (LV). In addition, a cross-validation of 8 random blocks repeated 20 times, provided a Standard Error of Cross-Validation (SECV), using the same number of LVs. The joint analysis of SEC and SECV, according to the specific features (*F*) of the models, either according to the number of slices used (2-D) or the number of date-wavelength couplings used (1-D) was considered for the selection of optimal N-PLS and PLS models, respectively.

These three different PLS models (two multi-way and one uni-way) were then applied to the test set. Bias and Standard Error of Prediction (SEP) were calculated on this prediction. Thus, the predictive performance of the regression models was quantified by the standard

coefficient of determination R^2, the bias and the standard error parameters in the calibration and test subsets.

5.3 Results

5.3.1 Three-way data array over the study case

The remote sensing data were organised in a three-way array ($\underline{\mathbf{X}}$) without temporal data gaps due to clouds and inconsistent number of available satellite images. Figure 5.4 shows the interpolated spectra on the J =19 dates, for the I = 107 plots.

Figure 5.4 shows typical properties of a time series that should not be neglected in satellite-based studies and applications. A high correlation between wavelengths was observed for nearby dates, which can be explained by the following factors: i) remote sensing datasets themselves tend to be data structures with high covariance and redundancy (Lopez-Fornieles et al., 2022) and, ii) by interpolating missing data, the correlation within the multivariate data structure was increased. Moreover, other potential sources of uncertainties such as multiplicative and additive effects may have affected the reflectance measurement values for the interpolated spectra (Richter et al., 2012). According to Liu et al. (2006), the combination of factors such as varying atmospheric conditions, varying sun-target-satellite geometry and sensor degradation could influence the final measurement value on time-series images by causing the above-mentioned effects.

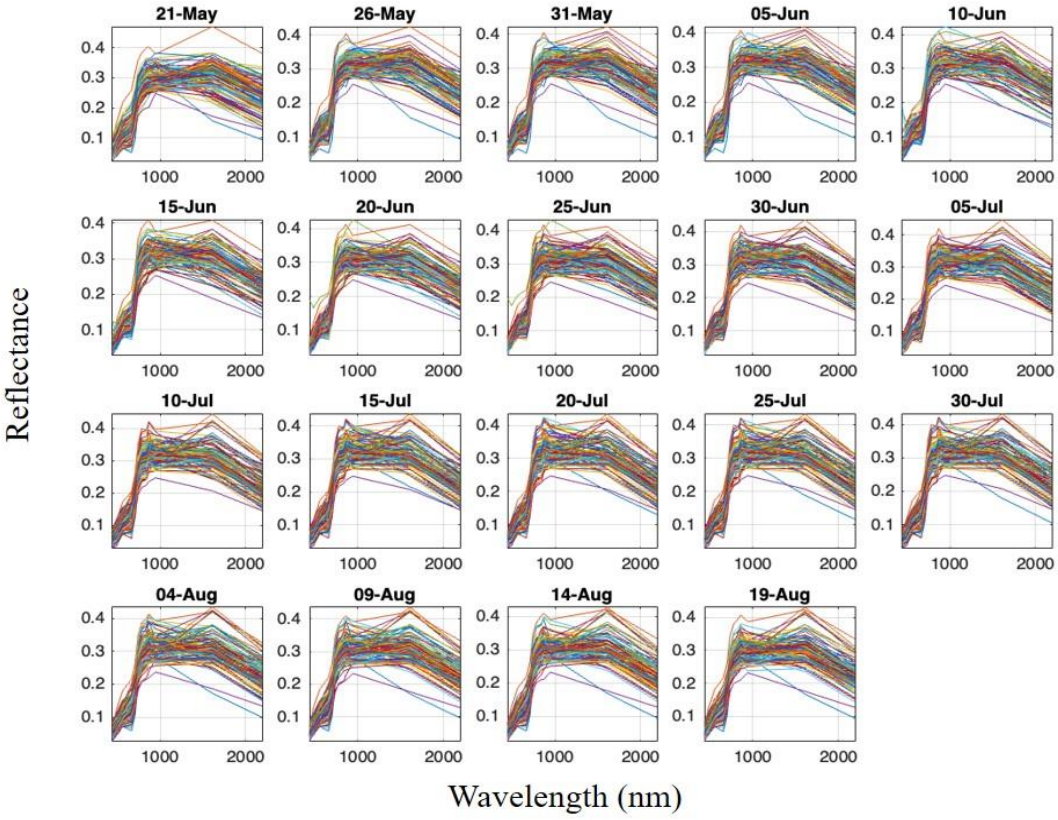


FIGURE 5. 4 - Interpolated spectra on the J =19 dates, for the I = 107 plots.

5.3.2 Quality of the regression models

Figure 5.5 shows the evolution of the SEC and SECV of an N-PLS for a cross-validation of 8 blocks repeated 20 times of a N-PLS calibrated on the temporal (Figure 5.5.a) and the spectral (Figure 5.5.b) slices selected by N-CovSel algorithm. It should be noted that the selected features, either dates or wavelengths, were ordered by the N-CovSel algorithm from highest to lowest covariance between the calibration set ($\underline{\mathbf{X}}_c$) and the **y**-vector (ground truth data).

This figure highlights a classical phenomenon for both graphs: a phase of decrease of the SEC, which corresponds to an improvement of the explanatory value of features, then a phase of increase of the SECV (while the SEC keeps on decreasing), which corresponded to the overlearning phase. On the basis of this joint analysis, the appropriate number of features for the two different N-PLS models were 6 temporal slices (Figure 5.5.a) and 7 spectral slices (Figure 5.5.b).

(a) (b)

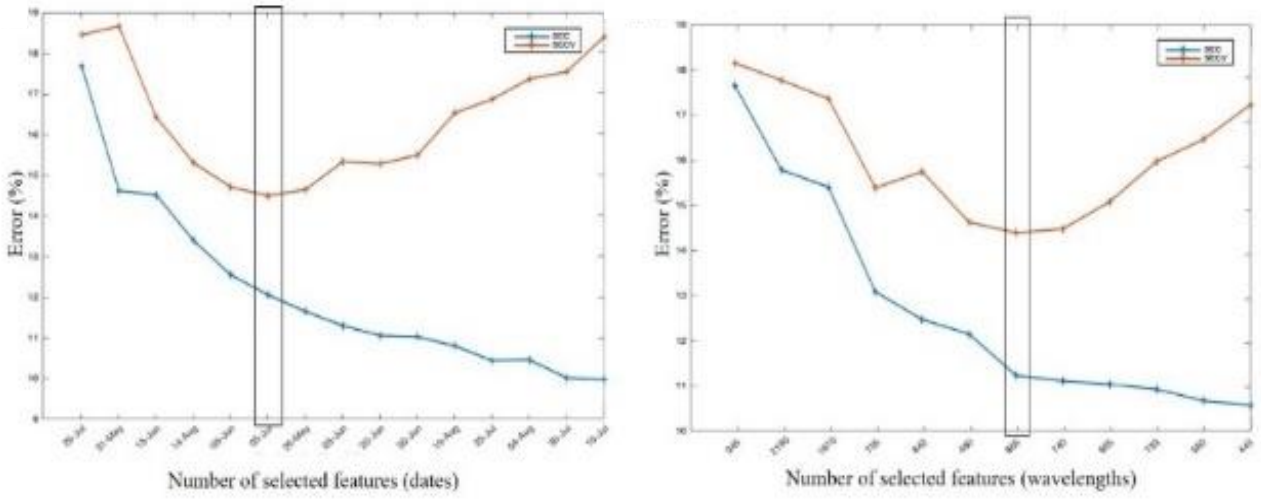


FIGURE 5. 5 -. Evolution of the SEC and the SECV criteria for an 8-block, 20-fold cross-validation of a N-PLS between (a) the temporal features (slices) selected by N-CovSel and the **y** losses and (b) the spectral features (slices) selected by N-CovSel and the y losses. The black frame indicates the optimal number of (a) temporal features (F = 6) and (b) spectral features (F = 7) selected.

Figure 5.6 presents the evolution of the SEC and SECV criteria for a cross-validation of 8 blocks repeated 20 times of a PLS calibrated on the date-wavelength columns selected and sorted by N-CovSel algorithm. Based on this joint analysis, the suitable number of features for the PLS model w 9 date-wavelength columns.

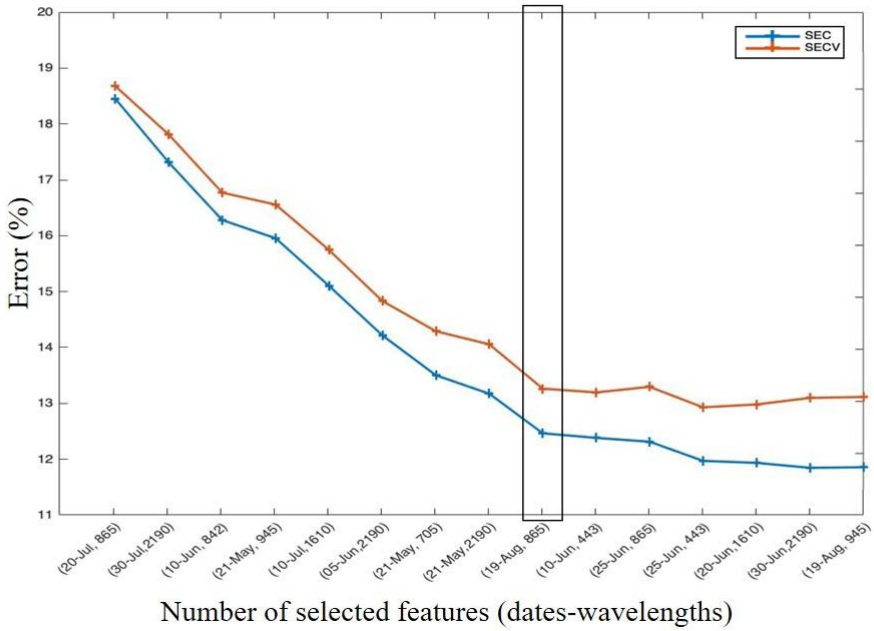


FIGURE 5.6 - Evolution of the SEC and the SECV criteria for an 8-block, 20-fold cross-validation of a PLS between the date-wavelength features (columns) selected by N-CovSel algorithm and the **y** losses. The black frame indicates the optimal number of columns (F= 9) selected.

The quality and performance of the temporal and spectral N-PLS models and of the date-wavelength-pair PLS model are presented in terms of the standard error of the calibration (SEC), the standard error of cross-validation (SECV) in the calibration set, the standard error of prediction of losses (SEP) in the test set, R^2 and the bias (Table 5.2).

A SEP over the predictions set between 11 and 13 % (Table 5.2) was consistent with the initial variability of the ground truth data (Section 5.2.2.1) due to the information required by the winegrowers to correctly characterise the level of the heatwave impact on a vineyard block.

TABLE 5. 2 - (a) N-PLS yield loss prediction results using the first 6 slices (temporal) selected by N-CovSel algorithm on individuals in the calibration set, with 80 vineyard blocks and in the test set, with 27 vineyard blocks. (b) N-PLS yield loss prediction results using the first 7 slices (spectral) selected by N-CovSel algorithm on individuals in the calibration set, with 80 vineyard blocks and in the test set, with 21 vineyard blocks. (c) PLS yield loss prediction results using the first 9 pairs (date-wavelength) selected by N-CovSel algorithm on individuals in the calibration set, with 80 vineyard blocks and in the test set, with 27 vineyard blocks.

Features	**F optimal**	**SEC (%)**	**SECV (%)**	**R^2 (%)**	**Bias (%)**	**SEP (%)**
(a) Temporal slices	6	12.1	14.2	0.62	-1.1	11.4
(b) Spectral slices	7	11.3	14.2	0.61	-1.4	13.0
(c) Date-wavelength columns	9	1.3	13.1	0.63	-2.3	11.7

5.3.3 Interpretation of the selected features

5.3.3.1 Extraction of 2-D features

Temporal slices

Figure 5.7 illustrates the operation of the N-CovSel algorithm, searching for 2-D features along temporal mode. Each sub-figure corresponds to the selection of a 2-D feature, i.e. a date. Each subplot represents $cov^2(\mathrm{X}_{.j}, \mathbf{y})$ as a function of the temporal dimension; the maximum of each curve corresponds to the selected features of $\underline{\mathbf{X}}$ in the second dimension with

the highest squared covariance with **y**. Each subplot (round) shows clear peaks, allowing the identification of dates involved in the prediction of yield losses. It should be noted that, for each round, the algorithm highlighted a different date of the time-series. Indeed, each curve showed a low value area around the previously selected variable and the overall amplitude of the curves for each round decreased as the features were extracted. These two particularities ensured that the selected features were at most complementary, i.e. at least correlated.

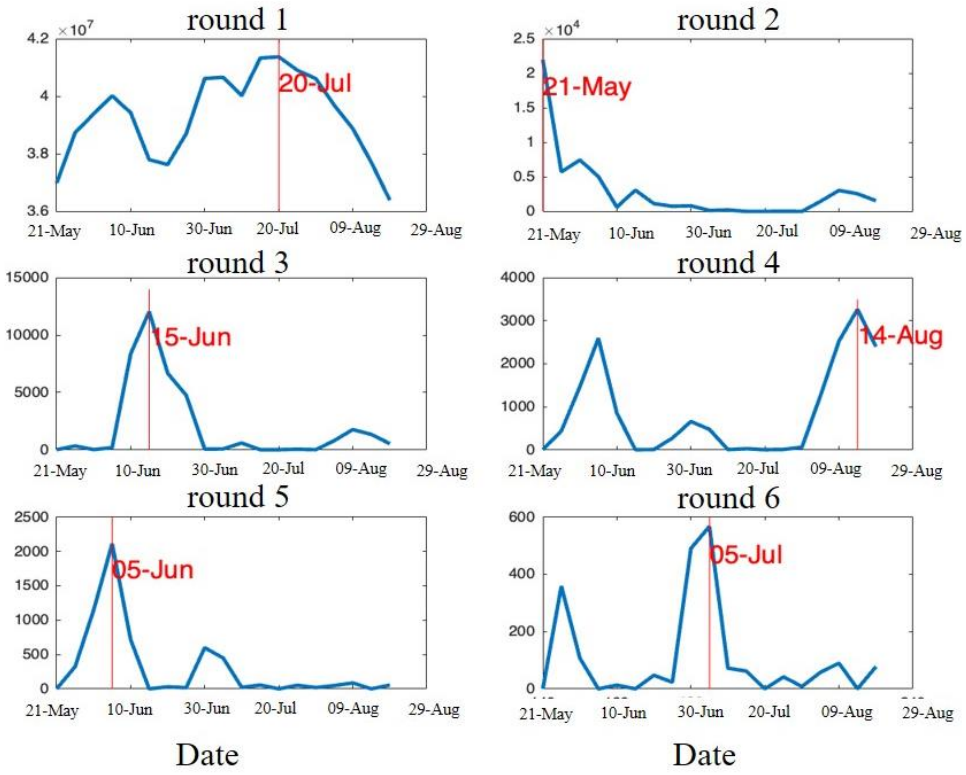


FIGURE 5. 7 - Evolution curves of $\boldsymbol{cov^2(X_j, y)}$ as the first 6 temporal slices were selected by N-CovSel. The selected feature corresponds to the maximum of each curve; the corresponding dates are shown in red.

The first round showed three local peaks (5th June, 30th June and 20th July) which did not appear in the subsequent rounds until the fifth one (5th June), meaning that the peaks, as well as their information, were correlated with each other. Thus, the information retained for the 20th (the global maximum peak) of July translated the information of a continuous spectral phenomenon from the beginning of June to the end of July that conditioned the final yield losses of the vineyard blocks, i.e. round 1 showed a phenomenon independent of heat stress. The second round represented the first available date of the study period and the third round (15th June) highlighted a date prior to heat stress. This indicated that the initial conditions of the vineyard blocks (before the extreme weather event) were also related to the observed final yield losses. The sixth round was the most indicative of the heatwave that occurred between 23rd June

and 8th July 2019 in view of their time frame. The fourth round showed two local peaks, on 14th August and 5th June and the fifth round showed the global peak on one same date, the 5th of June. This implied that as these were two consecutive rounds, the information contained in the 14th August (round 4) was independent from the 5th June round (round 5) in terms of final yield losses, i.e. the two dates were not correlated.

Spectral slices

Figure 5.8 illustrates the operation of the N-CovSel algorithm, searching for 2-D features along spectral mode. Each sub-plot corresponds to the selection of a 2-D feature, i.e. a wavelength. Each subplot represents $cov^2(\mathbf{X}_{.k}, \mathbf{y})$ as a function of the spectral dimension; the maximum of each curve corresponds to the selected features of $\underline{\mathbf{X}}$ in the third dimension with the highest squared covariance with **y**.

Each subplot (round) showed clear peaks, allowing the identification of wavelengths involved in the prediction of yield losses. It should be noted that, for each round, the N-CovSel algorithm highlighted a different wavelength of the spectrum. As mentioned for Figure 5.7, the particularities also shown in the Figure 5.8 ensure at most complementarity.

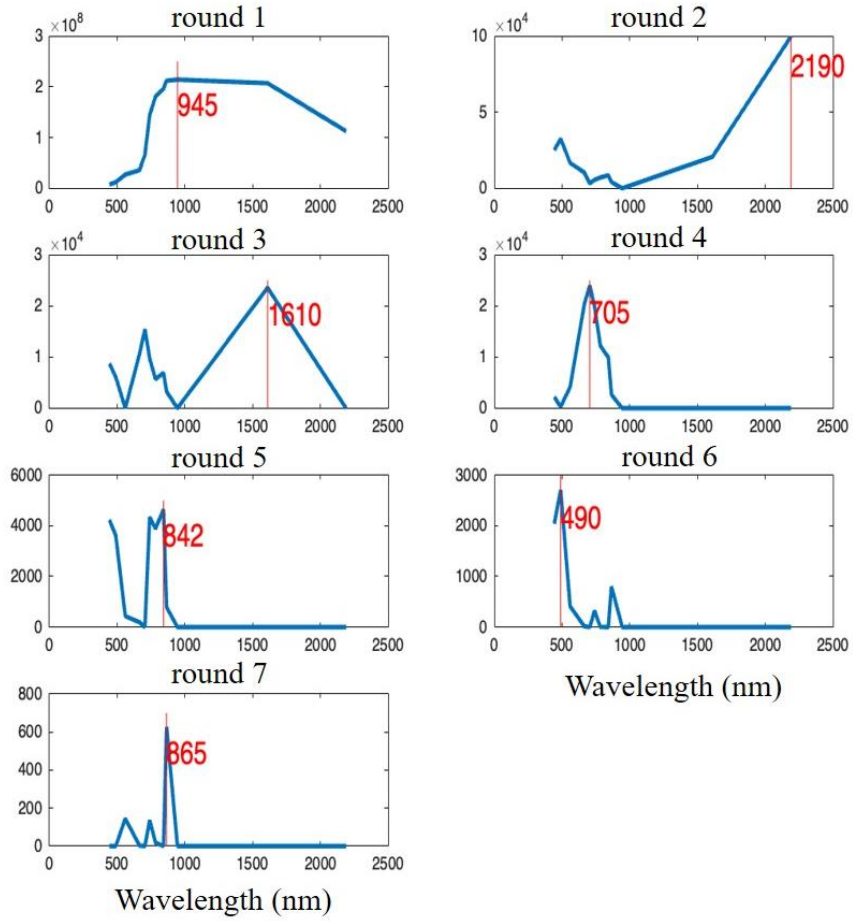


FIGURE 5. 8 - Evolution curves of $\boldsymbol{cov^2(X_{.k}, y)}$ as the first 7 spectral slices were selected by N-CovSel algorithm. The selected feature corresponds to the maximum of each curve; the corresponding wavelengths are shown in red.

It was noted that in round 1 (945 nm), the spectrum shown was the average spectrum of the vegetation. Although Clevers et al. (2008) determined that when looking through the atmosphere, the water band absorption in the 940 nm region should be considered to obtain information on the canopy water content, the shape of the displayed spectrum suggests that the 945 nm spectral band represents more of a multiplicative effect in the data. The 945 nm wavelength region had the highest covariance, i.e. the highest overall reflectance intensity, and this is the reason why the algorithm selected and sorted it in the first round. Regarding the spectral slices selected in the second and third rounds, with the range between 1600 nm and 2500 nm, i.e. in the shortwave infrared (SWIR) domain, it is well-known that the reflectance in this region of the spectrum is strongly correlated with vegetation water content (Jopia et al., 2020; Holzman et al., 2021). However, the second round (2190 nm) showed a baseline additive-type trend profile that reveals, as the first round, possible effects derived from the remote

sensing spectral measurement. It should be noted that these determinations of possible effects do not mean that the two rounds cannot provide information that could explain the changes in water concentration in the vineyard blocks. The following rounds highlighted spectral slices including the red-edge band at 705 nm (round 4) and the near-infrared (NIR) band at 842 nm and 865 nm (rounds 5 and 7) with round 6 determining the 490 nm wavelength, known as the blue band. Red-Edge region is related to leaf chlorophyll concentration (Laroche-Pinel et al., 2021b) and the reflectance in the NIR region is mainly affected by leaf and canopy structure (Slaton et al., 2001). The higher reflectance at 490 nm may be due to a strong reflection from dead biomass (Lorenzen & Jensen, 1988).

5.3.3.2 Extraction of 2-D features

Date-wavelength columns

Figure 5.9 shows the operation of the N-CovSel algorithm searching for 1-D features, i.e. pairs of dates-wavelengths. Each sub-plot represents the map of $cov^2(\mathbf{x}_{.jk}, \boldsymbol{y})$ as a function of the temporal and spectral domains; the global maximum in each sub-map corresponds to the selected features of $\underline{\mathbf{X}}$ in the second and third dimension with the highest squared covariance with **y**.

Each subplot (round) highlighted a different region of the temporal-spectral domain, allowing the identification of date-wavelength pairs involved in predicting yield losses. As for the 2-D extraction, for each round, the information correlated with the previous selected variables was removed, thus significantly decreasing the variance of the neighbouring variables in the following steps (Biancolillo et al., 2022).

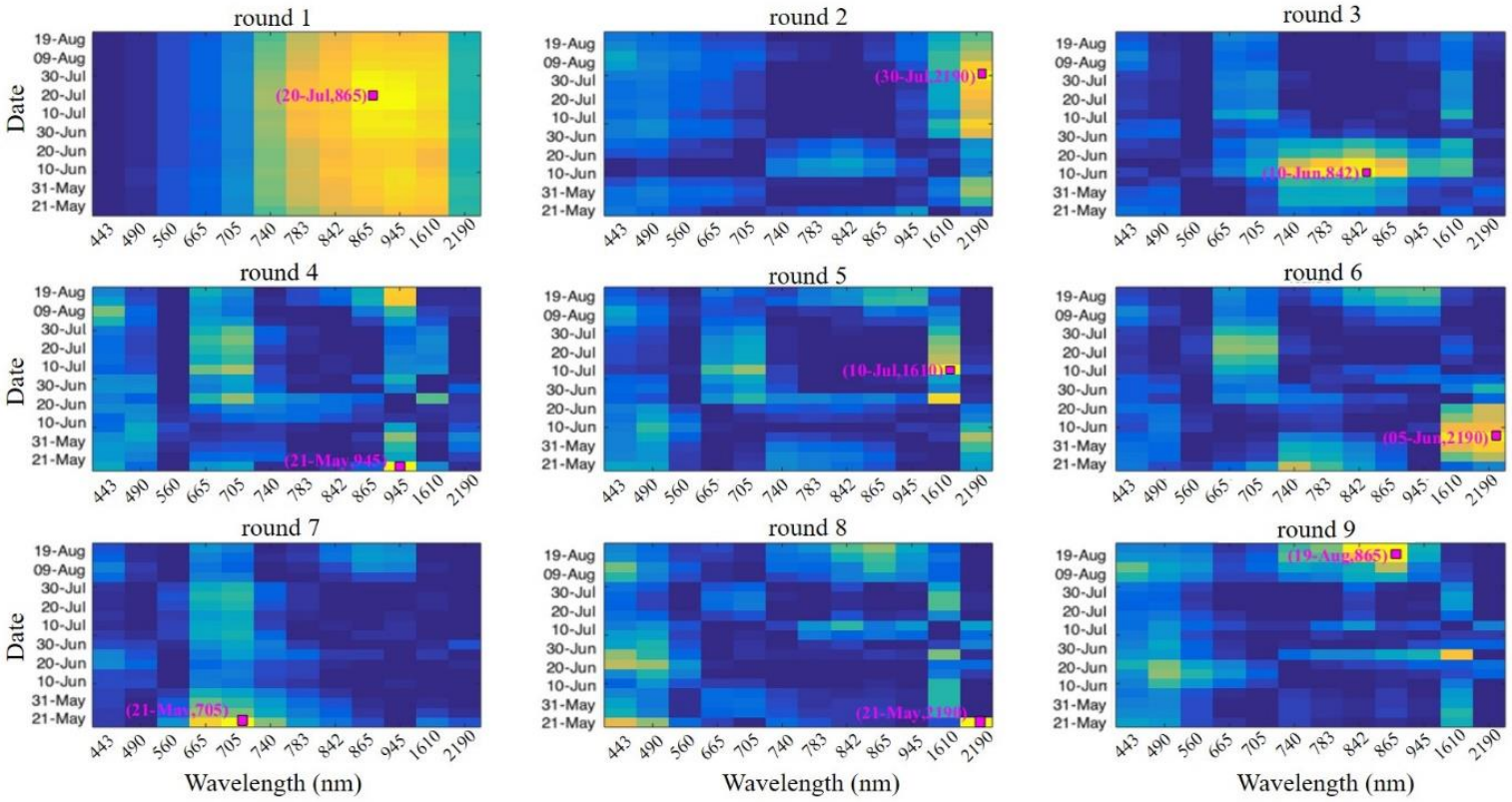


FIGURE 5. 9 - Evolution map of $\boldsymbol{cov^2(x_{.jk}, y)}$ as the first 9 pairs (date-wavelength) selected by N-CovSel. For each of the 9 rounds, the date-wavelength selected columns are highlighted by a red square. Dates and wavelengths are texted in pink. The colour gradient represents from yellow to blue, the highest and the lowest values of covariance between the date-wavelength pair (column) and the **y**-vector respectively.

The first round, unlike the others, indicates a spectral region as well as consecutive dates, thus determining an overall reflectance effect. It was observed that both the temporal and spectral dimensions had a low frequency i.e. the N-CovSel algorithm highlighted the entire study period containing the spectral region from 783 nm to 1610 nm (yellow). This result indicates that the overall effect of the reflectance, i.e. all-season vegetative profile that was related to the estimation of yield losses observed by the wine growers and advisors by means of maximum values of covariance. The remaining rounds showed high frequencies but in two of the different ways: (i) the second and third rounds showed high frequencies but which were prolonged either in the temporal dimension (round 2) or in the spectral dimension (round 3). For example, focusing on the second round, it should be noticed that the yellow colour appears from the 30th of June, with a maximum peak on the 30th of July but lengthening the high frequency until the 9th of August, at the wavelength 2190 nm; (ii) the remaining rounds from the fourth to the ninth, showed high local frequencies, i.e. covariance peaks, which highlighted more clearly a single date paired to a single wavelength.

Regarding the 1-D feature specificity of each round, the second, sixth and eighth rounds highlighted the wavelength 2190 nm, but with different dates, 30th July (round 2), 5th June (round 6) and 21st May (round 8). Laroche-Pinel et al. (2021a) demonstrated that the wavelength 2190 nm was one of the most discriminating for vine water status monitoring on a large scale. The three widely temporally spaced rounds may have been indicative of different responses of the various vine growth stages to water variation. The SWIR region is known to be sensitive to cell structure and water vegetation content (Huo et al., 2021). The date of 21st May, which also contained the wavelength in the SWIR region, was previously selected by the algorithm in the fourth and seventh rounds with the wavelengths 945 nm and 705 nm respectively. From the different wavelengths selected for the same date, it was determined that the initial vineyard blocks conditions related to the water status (2910 nm) as well as their chlorophyll concentration (705 nm) were related to the final yield losses. Besides the above, on this same date (21st May) the wavelength of 945 nm was highlighted. As shown in Figure 5.8, this wavelength could be representative of a multiplicative effect of the database and not represent agronomic information of interest. Just as the initial conditions of the set study period (May) were selected by the algorithm, so were the conditions at the end of the study period (August). The ninth round highlighted the pair on 19th August and the wavelength 865 nm. Reflectance between 685 nm and 700 nm has been established as one of the most sensitive for detecting plant stress (Gitelson et al., 1996).The third and fifth rounds were the closest in time to the heat episode that occurred between the 23rd June and the 3rd July 2019. The selected date-wavelength pairs were as follows: 10th June - 842 nm (round 3) and 10th July - 1610 nm (round 5). Reflectance at 842 nm is mainly related to leaf internal structure (Raddi et al., 2021). The high reflectance at this wavelength may have indicated a relevant change in morphology and canopy structure. As demonstrated by Raddi et al. (2021), the reflectance around 850 nm increases with season and severe stress factors. Regarding reflectance at 1610 nm, many studies reported the strong correlation of leaf water content with reflectance at wavelengths ranging from 1400 and 1900 nm (Champagne et al., 2003; Das et al., 2018). Thus, round 3 (10th June - 842 nm) could represent the water restriction (in the absence of irrigation) just before the heat stress during a period of high plant growth in the LR zone. This indicates that water stress in vineyard blocks before an extreme heat episode could have been an aggravating factor for yield loss. Meanwhile, round 5 (10th July - 1610 nm) could represent the subsequent effect of a sudden and strong heatwave on the water status of the vines.

5.4 Discussion

A generic example of the application of N-CovSel algorithm for variable selection was provided in the form of a time-series study in multispectral images. This paper showed the potential of methods originally developed in the analytical chemistry domain when applied to larger scales, e.g. in the life science domain. The application demonstrated the value of considering the feature reduction approach in the temporal and spectral dimensions for interpretation purposes in order to understand which variables contributed the most in the life science context presented.

In order to predict and estimate yield losses caused by a heatwave on vineyards fields, the N-CovSel algorithm was used. Based on a variable selection procedure according to their global covariance, the contributions of the temporal and spectral parts and their joint effect in the prediction of yield losses were characterised through three regression models. The performance of the models are as follows: for the temporal N-PLS model ($R^2 = 0.62$ – SEP = 11.4 %), for the spectral N-PLS model ($R^2 = 0.61$ - SEP = 13 %) and the temporal-spectral PLS model ($R^2 = 0.63$ - SEP = 11.7 %). From a predictive point of view, Lopez-Fornieles et al. (2022) already demonstrated that the application of the multidirectional regression method such as the N-PLS algorithm is appropriate to characterise and estimate the impact of an extreme event on grapevine. However, the interpretability offered by N-CovSel proved to be a very useful tool for understanding the agronomic processes underlying the spectral response of the crops over the time. It is well documented in the scientific literature that satellite monitoring of interactions between plants and light reflectance, in situations where crops interact with any aspect of their environment (e.g. extreme weather events), results in changes in plant signal (Knipling, 1970; Segarra et al., 2020). The variable selection approach identified the most significant features in a multidirectional environment, i.e. in a 3-way array, by selecting 2-D features (temporal and spectral slices) or 1-D features (date-wavelength columns) to be implemented within the model construction. Previous studies have shown similar results regarding the effects of heat stress from reflectance data in viticulture (Cogato et al., 2019b; Lopez-Fornieles et al., 2022), but notably, in the presented approach, the subset of features was selected simultaneously in two dimensions of the satellite information, i.e. temporal and spectral (Lopez-Fornieles et al., 2022). This selection procedure allows not only to identify the most significant wavelengths of the extreme weather episode but also knowledge on its a priori and a posteriori impact by integrating the temporal analysis from the N-way feature selection algorithm.

Since N-CovSel algorithm eliminates the correlation between variables by projecting the data orthogonally to the selected variable for the neighbouring variables in the following steps (Biancolillo et al., 2022), it is ensured that all selected features are at most complementary to each other. Furthermore, it is possible to sort the selected variables from the highest to the

lowest covariance related to observed yield losses. From the temporal slices (2-D features) selected and sorted, three important periods were observed that defined the data to be predicted i) the initial dates of the study period, centred on 21 May, ii) the dates close to the heatwave that occurred between the 23 June and 8 July and iii) the end of period dates, centred on 14-19 August. For the spectral slices (2-D features), the most important wavelengths (maximum covariance values) that were selected are known to be related to water absorption (Segarra et al., 2020) which may be indicative of the water status as the main factor affecting vine development. Wavelengths corresponding to the SWIR domain were observed from the date-wavelength columns (1-D feature) for the following dates ordered from highest to lowest covariance: 30th July - 2190 nm, 10th July - 1610 nm, 5th June - 2190 nm, 21th May - 2190 nm. As reflectance at 2190 nm is known to be relevant for monitoring vine water status at large spatial scale (Laroche-Pinel et al., 2021b), its selection at different dates throughout the study period shows the inconsistency of considering that yield loss is only due to heat stress. Indeed, the initial conditions (21st May) of water stress (2190 nm) (Laroche-Pinel et al., 2021b) as well as the information on the characteristics of the plant physiology in the Red Edge (700 nm) (Lopez-Fornieles et al., 2022) were already decisive for the final prediction. Given the proximity of the dates to the extreme weather event and that the detection of severe drought stress is centred at the wavelength 1610 nm (Cogato et al., 2019b), the date-wavelength pair of 10th July - 1610 nm was considered by the N-CovSel algorithm concerning the heatwave episode. The theory that the spectral response of the canopy representing the physiological behaviour of the grapevine, is affected by stress conditions due to fluctuations in ambient temperature, is well demonstrated in scientific literature (Cogato et al., 2021). In the final period of the study (still in full production), although after the collection of ground truth data on the condition of the vineyard blocks after the heatwave, the N-CovSel algorithm emphasised the 19 August - 865 nm pair. The reflectance in the Vegetation Red-Edge region (865 nm) is known in the literature as one of the most discriminating bands for water status (Laroche-Pinel et al., 2021b). The results of the present analysis confirm, with a variable selection approach, that a combination of SWIR (1610-2190 nm) (Das et al., 2018), Red-Edge (705 nm) (Ballester et al., 2018) and Red-edge Vegetation (865 nm) (Maimaitiyiming et al., 2017), is a valuable indicator for monitoring water status (Laroche-Pinel et al., 2021a).

The main advantage of using the N-CovSel algorithm for the remote sensing images is that being a methodology adapted for N-way arrays, the temporality and the spectral information are considered simultaneously. In the context of the life science case study, this allowed to establish that the heatwave was not the only explanatory factor of the final yield losses observed by the winegrowers and advisors. By temporally discriminating the most appropriate spectral information to characterise the beginning or end of the development season, as well as extreme events, it was observed that these were the integrating result of a series of factors that were mainly related to water restriction in key periods for plant development.

It is essential to place the results presented in this paper within the reality of multitemporal satellite data as they are sensors that measure reflected energy within several specific bands of the electromagnetic spectrum (Pettorelli et al., 2014). This implies that, as for the field of NIR spectroscopy (Isaksson & Næs, 1988), effects related to the reflectance of the spectrum are present in the analysis. The N-CovSel algorithm allowed the identification of multiplicative and additive effects in the selection of 2-D features. The choice to retain the observed effects was taken, as they could be important information in the interpretation of the model (e.g. wavelength 2019 nm). However, their removal at an early stage could have prevented the occurrence of effects in the covariance-based selection, e.g. wavelength 945 nm, mainly dedicated to atmospheric features detection (Verrelst et al., 2012). It should be noted that, due to the type of approach, the model should only be suitable for the year (2019) and the region (LR) considered. Thus, subsequent models remain specific to the learning base used for the calibration and their generalisation to other crops and/or other agricultural regions is rather limited.

Further applications are required before confirming the operational reliability of the N-CovSel method, in particular to provide spectral-temporal features to identify areas with different water restriction dynamics. For this, it will be necessary to complete the results of this study by extending the variable selection analysis to other types of phenomena, both those with a strong temporal evolution (e.g. extreme weather event such as hail) and those without (e.g. water scarcity in summer season), in order to better determine the dynamics of crop development and thus the reasons for its main cause-effects. As it appears that N-CovSel could be an efficient method addressing multiple response cases, an application study case to be studied would be its direct application to multispectral images, thus taking into account the spatial dimension.

Conclusion of Chapter 5

In this chapter it was decided to deal with complex datasets using a variable selection (or feature extraction) method to select a relatively small number of total variables and restrict it to the most significant ones for subsequent application in regression modelling. The approach selected was N-CovSel, which was designed to assess the relevance of variables in terms of their covariance with the response(s). The use of this method needs to be repositioned in the context of high-dimensional data as, to the authors' knowledge, there are few variable selection methods that account for more than one dimension at the same time, i.e. that consider data collected in a higher order structure. Therefore, the main objective of this chapter was to propose a formalism to apply the N-CovSel approach to a time-series of images at the regional scale in order to predict a variable of interest considering simultaneously the spectral and temporal dimension. More specifically, the N-CovSel approach was used to determine the spectral domains and time periods that appeared to be most relevant for assessing vine yield loss as determined by consultants and grape growers a few weeks after a heatwave event in 2019. The selected variables were used for subsequent application to a N-PLS or PLS model, depending on the structure of the variables selected (slices or columns). The prediction results were very similar with respect to those obtained in Chapter 4. Nevertheless, the results obtained in terms of information on the involvement of the temporal dimension in the analysis are very different, with the N-CovSel method being considerably simpler and easier to interpret.

The results showed the potential of N-CovSel as the spectro-temporal variables selected in the modelling process helped to understand the phenomenon studied. In fact, results related to the characterisation of yield losses revealed that although the heatwave had an impact on vineyards, its cause-consequence relationship was not direct. Other factors, such as initial crop conditions, proved to be important in characterising the significance of the impact suffered by the vineyards. The combination of SWIR (1610-2190 nm), Red-Edge (705 nm) and Red-edge Vegetation (865 nm) were the spectral domains that best characterised yield losses over time. This led to an assumption that water restriction is the main factor which affects plant development throughout its vegetative cycle in the LR region. The important contribution of this chapter and the method used is to highlight the implication of temporality in the expression of vegetative growth is fundamental to characterise yield losses induced by a heatwave.

Chapter 6. General conclusion and perspectives

6.1 Conclusions

Independent analysis of the spectral and temporal dimensions of remote sensing data does not allow a detailed analysis of crop dynamics at the field level over large scales. Despite this, it is hypothesised that high-resolution time series (i.e. less than a week of revisit time) of satellite imagery could be a suitable option for large-scale crop monitoring. Nowadays, the availability of time series of satellite images acquired from different satellite platforms, such as Sentinel-2, allows for near-real time monitoring. Nevertheless, it must be taken into account that the exploitation of multitemporal and multispectral remote sensing data for large-scale applications generates additional problems due to the large amount of different information that is collected. That is to say, when the time dimension is added as an explanatory variable of the spectral variation to be characterised, higher-dimensional data management is needed. Therefore, the exploitation of image time series in this manner raises methodological problems, as it involves a high volume of complex, heterogeneous and generally incomplete data.

Based on this statement, this thesis focused on the use of chemometrics methods to evaluate the integration of multi-spectral remote sensing time series as a potential source of relevant information for monitoring crop dynamics at a regional decision scale. To this end, high-dimensional Sentinel-2 satellite time series data were used to characterise the physiological properties of grapevines during their growing season from their spectro-temporal signatures for specific years. Two clear scientific issues were identified, which are recalled here:

- **How to integrate/consider the temporal dimensions in the analysis of multispectral time series images for crop monitoring in agriculture?**
- **Does the simultaneous consideration of the spectral and temporal dimension in time series provide more relevant and detailed way of capturing crop changes?**

The answers to these questions are summarised below, before outlining some perspectives (6.2).

In Chapter 2, the potential of remote sensing time series data to assess water restriction in vineyards using a conventional approach, such as NDVI was assessed. This first step to verify the potential of multispectral satellite imagery showed that an approach based on the empirical cumulative distribution function of NDVI values was relevant to classify vineyard blocks according to their water restriction. However, it also showed limitations in terms of temporal analysis and interpretation for crop monitoring derived from such vegetation indices, especially for crops, such as grapevines, where the changing presence/absence of the soil due to different management practices has a strong implication for the characterisation of changes in the spectral response within the imagery as the season and crop develops. The difficulty of visualising and interpreting raw data from vegetation indices was the starting point to expose

the current limitations (spectral and temporal) of these approaches and to determine why there is a demand for a paradigm shift in time-series vegetation monitoring.

Based on the assumption that the use of temporal variations in satellite imagery, in addition to spectral variations, will allow a more complete analysis and monitoring of the dynamics of the spectral response of crops, a move towards the use of chemometric methods was chosen in order to address the methodological challenge of managing the high dimensionality of these spectro-spatio-temporal data. For this purpose, unsupervised approaches (Chapter 3), supervised approaches (Chapter 4) and approaches based on variable selection (Chapter 5) were evaluated to determine if they were more holistic for the processing of multispectral images time series.

In Chapter 3, starting from the exploration of the potential of multispectral time series imagery for crop monitoring at regional scales with multi-way methodologies that are able to handle the high dimensionality of RS data, the objective was to see if the addition of the temporal dimension in the analysis facilitated the interpretation of crop dynamics. Within the context of a study applied at the field (vineyard block) scale, but representative of large scales, it is important to be aware of the operational limitation of the lack of field data (real observations) inherent to these decision scales. For this reason, it was considered appropriate to use, as a first approach, unsupervised chemometric methods that deal with high dimensionality, which in this case involved the temporal dimension, to see if they enhanced the understanding gained from crop monitoring. The results obtained with the PARAFAC methodology highlighted the interest of the simultaneous integration of the temporal and spectral dimension for identifying and interpreting spatial regional/local specificities (e.g. the presence/absence of irrigation sectors) in vineyard blocks for the years 2019 and 2020. The spectro-temporal profiles obtained were useful to understand the spatial variation of the spectral response of vine cultivation (soil-vegetation dynamics) over time through the evaluation of the score maps by viticulture experts. Given the context of this large-scale study, the methodology used to explore the multitemporal and multispectral data relied on expert knowledge as an alternative validation framework. However, although the PARAFAC methodology showed potential for characterising large-scale vineyard cultivation, the assumption of the trilinear structure of the data forced a temporal interpolation that may eventually mask relevant spectral variations for the proper monitoring of the crop evolution. Therefore, as an alternative method to solve the multi-way data analysis problem that does not require a continuous cubic structure (i.e. no temporal interpolation), the MCR-ALS methodology was proposed. The application of MCR-ALS did not provide further knowledge related to the characterisation of the grapevine crop, but aimed to provide a different approach to the PARAFAC method and the treatment of the raw temporal satellite data, opening the possibility of a more robust and interpretable characterisation of the evolution of the spectral response of grapevines. Both approaches were proven to be convenient methods to improve the large-scale understanding of the different

spatial behavioural patterns of grapevine cultivation by integrating their temporal dynamics. In addition, both methods showed the potential of including the spatial dimension for a better visualisation and interpretation of the results.

In Chapter 4, in order to investigate the integration of the temporal dynamics of crop reflectance, the potential of a supervised approach that could link the high-dimensional information of the RS data to the ground truth data of interest was evaluated. The results validated the effectiveness of the proposed N-PLS method to analyse complex multispectral data with different spectral domains and a clear temporal evolution. In particular, its potential to capture more relevant and detailed vineyard dynamics for better crop monitoring at the regional scale was shown. By proposing a formalism to apply the N-PLS approach to a time series of RS images in order to predict a limited number of ground data, the interest of adopting a systemic analysis that simultaneously accounts for the spectral and temporal characteristics of the data under consideration was also demonstrated. The application of N-PLS to characterise and estimate the impact of an extreme weather event, a heat wave, in 107 vineyards located in the Languedoc-Roussillon (LR) wine region, showed that it was possible to predict yield losses with a R^2 of 0.66 and an RMSE of 10.7 %. In addition to its predictive capability, the increased interpretability of the approach allowed the identification of the spectral bands around the red edge (700 nm) and the SWIR region (1600 nm) as being the most relevant. This was because their spectral profile showed a deepening of the time profile during the heat stress period, allowing the band information to be related to the recorded yield losses.

The results of this study were complemented by further research, extending the analysis to the spatial distribution of the effect of an extreme weather phenomenon to 4978 vineyard blocks to determine its regional dynamics on yield loss. To this end, N-PLS was applied to a specific viticulture-based case to examine the effect of the heat wave that occurred on the 28th of June 2019 in the LR region. Although the validation of the model at the regional scale was performed on the basis of the maximum temperature observed on the 28th of June, 2019, which is a relatively simple approach to address a very complex phenomenon, the results showed the robustness of the methodology and its ability to map the spatial footprint of the heatwave that affected the south of France. Thus, it was shown that the high dimensionality of the data could be accounted for by using a systemic methodology that simultaneously took into account both the spectral and temporal characteristics to discriminate and characterise the impact of a heat wave on vineyards at a regional scale.

Regarding this type of supervised approach, it must be noted that it was not able to handle temporal gaps in the data due to cloud/shading effects and/or the irregular availability of satellite images during the acquisition period. Temporal interpolation of the data to obtain a continuous three-way data structure was mandatory for the development of the method. In fact, the application of N-PLS would not be suitable for areas or seasons where image availability is strongly affected by cloudiness.

In Chapter 5, in order to further explore what complementary insights the temporal dimension can provide for the identification of the most discriminating spectral domains in relation to the development of a crop, the potential of a multi-way variable extraction/selection method was assessed: the N-CovSel method. As with the N-PLS approach, a continuous cube was also necessary to apply the methodology. The N-CovSel approach aimed to identify and characterise the most significant temporal (dates or periods) and spectral (spectral bands or domains) responses of the vine canopy to a stress period. For this purpose, the study aimed at identifying and interpreting the different effects caused by the same heat wave in 2019 (between 23 June and 8 July) on the percentage of yield loss observed on 107 vineyard blocks in the LR region. The contributions of the temporal and spectral slices and their joint effect (date-wavelength columns) in the prediction of yield losses were characterised through three regression models. The results of this study showed that the performance of the models defined according to a) spectral slices, b) temporal slices and c) spectral-temporal columns, offered equivalent quality and robustness of prediction with R^2 values between 0.62, 0.61 and 0.63 and a SEP of 11.4 %, 13 % and 11.7 %., respectively. Moreover, the results demonstrated the value of considering the variable selection approach in the temporal and spectral dimensions for interpretation purposes, as this allowed the importance of taking into account temporal windows to be established. It was determined that the temperature was not the only explanatory factor for the observed final yield losses, but rather it was the integrating result of a series of factors, including temperature, that were mainly related to water restriction in key periods of plant development.

In conclusion, regarding the methodologies used, the innovative part comes from the explicit consideration of the temporal nature of the RS data with multi-way methods. As agricultural crops evolve over time within a season, relying on the analysis of the temporal dimension seems to be a relevant and interesting idea to better detect spectral variations related to crop changes. However, the characterisation of crop behaviour over time for each spectral band is meaningless if independently extracted for each image. The overall temporal ‘profile’ of the spectral reflectance curve of the crop needs to be taken into account to distinguish regional scale variations induced by climate and management practices at certain points in the growth cycle. Therefore, methodologies have been proposed and used that can handle high-dimensional data that simultaneously incorporate spectral-temporal data. Moreover, with these chemometric methodologies that provides useful information, such as scores, loadings and regression coefficients, the results obtained can facilitate the interpretability of the evolution of agronomic traits over time.

In general, the integration of multi-spectral remote sensing time series proved to be a promising way to derive multi-way attributes that offered the possibility to characterise the growth dynamics of a crop, such as grapevines, and the impacts on it, linked to climatic phenomena that cause extreme stress in the crop.

6.2 Perspectives

Several perspectives can be drawn from this research, which are presented hereafter.

In this research, the analyses were restricted to an *a posteriori* integration of the spatial dimension by geostatistical methods. The spatial dimension was not integrated into the main (spectro-temporal) analysis because the nature of the spatial data did not coincide with the nature of the spectral or temporal data, which can be represented as a 'spectrum' adapted to chemometric analysis. Thus, a next step would be to deal simultaneously with the spectral-temporal-spatial dimensions of the RS data. In this perspective, two different paths to consider have been identified. A first approach could be based on the integration of the spatial dimension by means of spatial constraints in the optimisation process of the algorithm to be applied. A clear example of this type of procedure is the MCR-ALS methodology, which applies appropriate constraints to impose specific spatial characteristics on the data set. The application of spatial constraints based on image processing approaches (e.g., segmentation, smoothing or spatial filtering), image modelling techniques (e.g. spatial model fitting, super-resolution or deconvolution), or spatial attributes (e.g. texture), have great potential for adaptation to the domain of time-series analysis of multispectral images. In particular, in the agricultural context, this type of approach would make it possible to exploit the spatial variability in the imagery to better understand chemical and physical events or phenomena that affect site specifically crops over time. However, an adaptation of this kind of methodology would require a redesign of the relevant spatial variables to be considered at high spatial resolution scales of multispectral imagery. The second possible path to follow would be a paradigm shift in the integration of the spatial dimension. In fact, the main challenge of simultaneously considering the spectral, temporal and spatial dimension in a single chemometric methodology is the differing nature of the data in the spatial dimension relative to the spectral or temporal data dimensions. Therefore, this integration should represent the spatial dimension as a continuous 'spectrum', rather than as a simple discontinuous attribute, i.e. the spatial data state cannot be limited to a specific set of values, but needs to vary, in a progressive way, along a continuum. Approaches to simultaneously exploit spectral, temporal and spatial dimensions in multispectral remote sensing images remain a current issue and a challenge for the coming years.

Likewise, the spectro-temporal methodological approaches presented throughout the manuscript also present potential perspectives *per se* in different fields of application. Given the potential to identify and characterise phenomena with a clear temporal evolution (e.g. a heat wave) that represent a rapid change in the spectro-temporal profile of crop vegetation, it would be interesting to determine whether other types of phenomena that vary in a less pronounced way (e.g. the onset of water stress) could also be assessed. It would also be interesting to further investigate and validate the proposed approaches in (i) other wine growing regions (not the

Mediterranean arc), (ii) other crops, within and/or outside the domain of arboriculture and (iii) other applications of earth observation, other than crop monitoring, as these methods may also be adapted to more general domains in agriculture and environmental modelling, for example land cover and land use monitoring, or disaster modelling and monitoring. In addition, in this research, only time series of Sentinel-2 remote sensing images have been used. However, there are many other (and many new) earth observation sources (e.g. other satellite platforms, UAVs, etc.) that now present an unprecedented amount of possibilities in terms of data availability as well as unique characteristics in terms of revisit time, spatial resolution, diverse information to be provided, cost and spatial coverage. Therefore, a next step would be to consider the use of these multi-way methodologies with other sources of information to study and characterise in more detail the dynamic spatial monitoring from a plot/field to a regional scale to better match the specific conditions of all crop types.

Going a step further, it should be noted that the methods presented are dedicated to a single spatial scale. In general, remote sensing algorithms are still dedicated to a single given scale, leading to large discrepancies between data sources and models at different scales. However, in line with the environmental context in which humans find themselves (e.g. recurrent extreme weather events), data processing is crucial, both at the global scale (e.g. food security), at the regional scale (e.g. management of terroir specificities) and at the field scale (e.g. management practices in agriculture). Thus, there is a need to adapt the algorithms used in remote sensing to multiple scales, where the methodologies shown would be interesting for their ability to deepen our fundamental understanding of the evolution of the process of crop functioning in relation to the effects of different weather conditions.

These sets of perspectives represent necessary short-term research that is needed in order to set up operational tools for diagnosing, monitoring spatial dynamics and steering crops at different scales in the face of a changing climate and the expected increased prevalence of extreme weather events (heat, hail, etc.). In the long term, this would provide winegrowing (and cropping) professionals (producers, advisors, consular bodies, water distributors) with decision support tools (for expert appraisal or as a model input variable) to enable them to make management decisions and take differential actions. The implementation of such services in agriculture, and particularly in the wine industry, seems both plausible and opportune. However, it should be noted that in order to work on the proper monitoring of a crop (not only grapevines) on a large scale, it is necessary to set up a collective database that must go from the small farmer to regional public bodies. New forms of data collection at user level, such as crowd sourcing, or via the digitization of agriculture, should be a source of reflection to facilitate and open the way to collate all available data types to improve the collective knowledge necessary for the proper monitoring of crops.

Résumé étendu de la thèse

Chapitre 1. Introduction

Le nombre de satellites en télédétection ainsi que la qualité et la portée des informations collectées ne cessent d'augmenter, offrant une plus grande couverture des cultures avec des temps de revisite plus courts et avec la disponibilité de données d'archives. Le besoin d'algorithmes efficaces capables de traiter une telle quantité de données exigera de plus en plus de nouvelles approches méthodologiques à utiliser dans la surveillance agricole (Dalla Mura et al., 2015). Le suivi de cultures annuelles ou pérennes à l'aide de technologies de télédétection est donc un vaste sujet qui a été largement abordé sous de multiples angles. La télédétection est discutée, par exemple, pour des applications agricoles spécifiques (i.e. la prévision du rendement, la surveillance de l'irrigation, etc.), sur des plateformes de télédétection spécifiques (i.e. des satellites, des drones), des capteurs spécifiques (i.e. des systèmes de détection actifs ou passifs) ou des emplacements (i.e. différentes échelles spatiales de résolution, et dans des contextes climatiques spécifiques (i.e. biomes) (Weiss et al., 2020). Les informations d'intérêt pour le suivi des cultures sont basées sur les caractéristiques des cultures ou des systèmes agricoles et, surtout, sur la façon dont ces derniers varient dans l'espace et dans le temps. En fait, la relation entre ce qui est directement mesuré par les instruments de télédétection et les caractéristiques agronomiques elles-mêmes doit être modélisée d'une manière ou d'une autre (plus ou moins substantielle) pour déduire les secondes à partir des premières (Weiss et al., 2020).

Actuellement, la grande majorité des technologies de télédétection, c'est-à-dire les plateformes et les capteurs, sont accessibles et relativement abordables pour le secteur agricole. Toutefois, le progrès technologique n'a pas suivi le rythme de développement du secteur agricole en question, de sorte qu'il n'a pas encore été possible d'utiliser au maximum de leurs potentielles ces technologies. Cela s'explique en raison de contraintes opérationnelles telles que l'efficacité à répondre de manière adéquate à la nature du problème agronomique, du coût élevé en temps et en moyens nécessaires pour traiter ces données à haute résolution (Khanal et al., 2020). Cette dernière limitation conduit à des défis méthodologiques, tels que la gestion du grand volume, de la diversité et de la complexité des données de télédétection, en particulier lorsqu'on considère la haute dimensionnalité (qu'elle soit spectrale, temporelle ou spatiale) (Ma et al., 2015). Par exemple, en raison de la disponibilité accrue de données temporelles provenant de satellites, une attention particulière dans le domaine du suivi des cultures a été accordée au développement de nouvelles approches qui prennent en compte des données à haute dimension, permettant aux chercheurs d'examiner les tendances temporelles (Southworth et Muir, 2021).

En effet, les séries temporelles mettent en évidence la possibilité de découvrir la tendance du changement en construisant des 'profils' de données spectro-temporelles, ce qui est essentiel pour un suivi précis des cultures, car chaque type de culture a un calendrier bien défini avec des périodes d'exploitation spécifiques et des modèles de croissance saisonniers uniques (Loew et al., 2013). Ce fait conduit à la formulation du problème de recherche général de la thèse, qui est le suivant :

Les informations des séries temporelles issues de l'imagerie de télédétection multi-spectrale offrent-elles de nouvelles perspectives en agriculture par la proposition de descripteurs évaluant la performance des cultures ?

Plus précisément, ce projet vise à déterminer si la prise en compte simultanée de la dimension spectrale et temporelle dans les séries temporelles peut fournir une vision plus pertinente et plus détaillée des changements dans les cultures. L'hypothèse est que l'intégration de caractéristiques multi-spectrales dans une analyse multi-temporelle permettant d'obtenir des signatures spectrales-temporelles qui caractérisent la variabilité spatiale peut offrir de nouvelles perspectives dans la surveillance des cultures pour des applications telles que l'estimation du rendement des cultures, l'évaluation de l'état des cultures et la détection du stress des cultures.

Cependant, cette hypothèse soulève une question de gestion des données multidimensionnelles :

Comment intégrer/considérer pertinemment la dimension temporelle dans l'analyse des images multi-spectrales de séries temporelles pour le suivi des cultures en agriculture ?

L'une des approches intéressantes pour y répondre semble être l'utilisation de méthodes chimiométriques. Ces méthodes permettent d'analyser des ensembles de données complexes et hautement dimensionnels obtenus à partir de sources de données RS. L'adaptation de ces méthodes chimiométriques au domaine de l'agriculture permet donc d'intégrer des informations spectrales et temporelles pour identifier et décrire l'interrelation des facteurs environnementaux ou humains et leur impact potentiel sur l'agriculture.

Outre la décision d'utiliser uniquement des méthodes chimiométriques, deux autres décisions ont été prises préalablement afin d'aborder les questions présentées plus en détail. La première était de se concentrer sur une culture particulière, la viticulture, et à l'échelle d'une région, l'Occitanie. Plus précisément dans la région viticole anciennement connue sous le nom de Languedoc-Roussillon. Cette région possède le plus grand vignoble français en surface avec 2 bassins de production, l'un en zone méditerranéenne et l'autre dans le sud-ouest (Filippi, 2012) et se caractérise par le caractère mixte de ses vignobles, car ils sont situés dans une grande

diversité de paysages et de conditions climatiques contrastés, ce qui la rend intéressante à étudier. Compte tenu de l'échelle spatiale choisie dans ces travaux de recherche, seuls les satellites Sentinel-2 (A/B) ont été considérés en raison de leurs caractéristiques intéressantes et représentatives en termes de résolution temporelle (fréquence de revisite globale de 5 jours), spatiale (4 bandes spectrales à 10 m, 6 bandes à 20 m et 3 bandes à 60 m) et spectrale (13 bandes spectrales du visible à l'infrarouge à ondes courtes).

Chapitre 2. Principales limites de l'utilisation exclusive de la dimension spectrale des séries chronologiques multi-spectrales pour le suivi des cultures

Les indices spectraux de la végétation ont été conçus et utilisés comme indicateurs des variations temporelles et spatiales de la structure des propriétés biophysiques et de la vigueur de végétation (Xue and Baofeng, 2017). De ce fait, les indices de végétation comme une opportunité majeure pour étudier et surveiller la végétation et sa dynamique (Silleos et al., 2006). Compte tenu de ce contexte, il est naturel que l'analyse des séries temporelles se soit concentrée sur l'analyse de l'évolution des indices de végétation dans le temps (Hatfield et al., 2019 ; Li et al., 2021). Dans cette perspective, l'objectif de ce chapitre est de montrer l'intérêt et les limites des approches classiques (c'est-à-dire basées sur les indices de végétation) pour exploiter le potentiel des séries temporelles pour le suivi agricole à l'échelle régionale.

Pour cela, une approche originale permettant de traduire la dynamique temporelle de NDVI obtenue par Sentinel-2 sur l'ensemble d'une saison de végétation a été développée. Cette approche a été construite en considérant la fonction de distribution cumulative empirique des valeurs NDVI observées. Cette approche a été appliquée dans une étude de cas liée à la dynamique de l'état hydrique de la vigne. L'approche des séries temporelles NDVI s'est avérée capable de classifier les parcelles en fonction de leur restriction hydrique permettant ainsi d'identifier la dynamique du stress hydrique à l'échelle régionale. Toutefois à travers ces travaux, plusieurs limitations ont été mises en évidence. En effet, Maynard et al. (2007) ont résumé trois points problématiques pour une approche qui utilise une seule ou un ensemble limité de bandes spectrales pour calculer des indices de végétation : (i) elle ne tient pas compte du fait que d'autres bandes spectrales pourraient contribuer et améliorer l'interprétation ; (ii) elle limite la capacité à modéliser les effets des différentes interactions végétation-rayonnement dans différentes parties du spectre électromagnétique ; et (iii) l'impact de l'hétérogénéité du sol sur les réponses spectrales est particulièrement évident, notamment dans les canopées hétérogènes comme celles des vignobles. Ces questions ont des implications importantes pour l'identification des changements et le suivi de la dynamique des surfaces végétalisées. D'où la nécessité d'un changement d'approche pour intégrer correctement toutes les informations

spectrales fournies par les images multi-spectrales. Par exemple, une approche utilisant des données spectrales non indexées devrait être considérée comme une alternative intéressante aux indices de végétation pour éviter les problèmes discutés.

En plus de la résolution spectrale, il faut tenir compte du fait que les satellites fournissent régulièrement des images avec une résolution temporelle et spatiale différente. L'approche dominante du suivi des cultures, basée sur des séries temporelles d'indices de végétation, ne prend en compte aucune de ces deux dimensions, c'est-à-dire ni la dimension temporelle ni la dimension spatiale dans l'analyse. Par conséquent, ces approches basées sur les indices de végétations ne sont pas adaptées et judicieuses pour extraire et traiter toutes les informations pertinentes qui peuvent être présentées dans ce flux de données massif. Cependant, fusionner ou combiner ces informations pour une interprétation de meilleure qualité est un défi qui n'est pour le moment pas toujours bien résolu (Militino et al., 2018).

Chapitre 3. Potentiel des méthodes multidimensionnelles non supervisées pour l'exploration régionale des données sur les cultures à partir de séries temporelles multi-spectrales

Dans le chapitre précèdent, on a pu mettre en évidence certaines limites de l'approche conventionnelle qui reposent uniquement sur les changements des indices de végétation au cours du temps pour caractériser la dynamique de croissance d'une culture. Or intrinsèquement, les images de télédétection fournissent des informations plus riches sur la variabilité spatio-temporelle de la signature spectrale du couvert végétal. Par conséquent et dans l'objectif d'exploiter toute cette richesse d'information, la nécessité d'utiliser des approches d'analyses de données plus complètes permettant de considérer simultanément les dimensions spectrale et temporelle pour rendre compte de toutes les informations potentielles fournies par les séries temporelles d'images multi-spectrales est un prérequis. Cependant, l'ordonnancement temporel des données de télédétection pour fournir une autre dimension à l'espace de données créant ainsi non seulement un défi en termes d'extraction et d'analyse, mais aussi en termes de structuration de ces ensembles de données à haute dimension (Chi et al., 2016). Dans ce qui suit, nous sommes parties du postulat que l'organisation des données issues d'instruments multi-spectraux sous la forme d'une structure multidimensionnelle (cube de données) dont les constituants principaux sont des dimensions était une proposition pertinente pour extraire le meilleur potentiel des séries temporelles d'images multi-spectrales. Cette façon originale de structurer les données permet de considérer la dimension temporelle comme un facteur inhérent à l'analyse (Ferreira et al., 2020). Ce chapitre vise donc à explorer le potentiel de séries

temporelles d'images multi-spectrales à l'échelle régionale pour l'extraction de connaissances agronomiques tout en tenant compte simultanément de la dimensionnalité spectrale et temporelle des données.

La variabilité spatio-temporelle de la signature spectrale d'un couvert végétal à l'échelle spatiale de résolution d'une région peut révéler des informations très utiles pour le suivi de la dynamique croissance des cultures. Cependant, si l'étude de l'échelle régionale est intéressante par sa diversité d'informations (conditions pédoclimatiques, cépages, pratiques culturales…), il est très difficile de disposer de données auxiliaires de référence provenant du terrain. Dans ce contexte, les approches de modélisation multidimensionnelle non supervisées utilisées en chimiométrie ont été considérées comme une première tentative d'exploration des données, car elles ne font intervenir que les propriétés intrinsèques des séries temporelles recueillies dans les images multi-spectrales. Par conséquent, des approches de modélisation multidimensionnelle non supervisées sont utilisées pour fournir des informations pertinentes sur une étude de cas consacrée aux séries temporelles du satellite Sentinel-2 et à la culture de la vigne à l'échelle régionale de la région Languedoc- Roussillon pour fournir plusieurs nouvelles connaissances sur le suivi des cultures grâce à l'analyse simultanée de la dimension spectrale et temporelle.

Deux approches de modélisation multidimensionnelle non supervisée différentes de résolution à trois dimensions ont été utilisées pour intégrer la dimension temporelle dans la caractérisation de la variabilité de la phénologie de la vigne à l'échelle régionale. La première approche utilisant la méthodologie PARAllel FACtor (PARAFAC) a souligné à quel point les études exploratoires peuvent être enrichissantes en tenant compte de l'inclusion de la dimension temporelle pour une meilleure compréhension du suivi des cultures à l'échelle régionale, sans pour autant négliger les limites évidentes qu'elles impliquent, comme la faible connaissance des spécificités d'interprétation à grande échelle. La méthode PARAFAC a permis de fournir des informations sur les profils spectraux temporels, qui, à leur tour, ont permis la caractérisation spatiale au moyen de cartes des scores obtenus à l'échelle de la région viticole Languedoc-Roussillon sur les périodes de mai à aout pour les années 2019 et 2020. Ces cartes factorielles de scores ont montré des motifs spatiaux qui représentaient les différences et les similitudes de la dynamique de composants liés aux signatures spectrales de la végétation et au sol. Au vu de ces résultats (Figure R.1), ce type d'approche de modélisation multidimensionnelle non supervisée s'est avéré judicieuse pour spatialiser et caractériser des phénomènes ayant une évolution temporelle, par exemple la réponse de la signature spectrale du couvert végétal de la vigne, en fournissant des 'motifs' spectro-temporelles capables de mettre en évidence des différences de comportement induit par des par des facteurs climatiques (i.e. la saisonnalité, les événements météorologiques extrêmes) ou non climatiques (i.e. pratiques culturales et de récolte). Cependant, dans cette étude de cas, probablement en raison de l'échelle de résolution, son application nécessitait une connaissance experte *a posteriori* du phénomène observé, limitant ainsi son applicabilité immédiate. Le cadre conceptuel sur lequel la méthodologie PARAFAC a été proposée repose sur une validation basée sur des observations

d’experts externes. Selon l’approche proposée par Shadish et al. (2002), la validité externe pourrait être obtenue en sélectionnant un petit nombre d’experts délibérément très diversifiés. Par conséquent, bien que la limitation majeure de l’approche soit due à la rareté des données auxiliaires de référence provenant du terrain, l’application validée avec un cadre pratique d’opinions d’experts vignerons a démontré que dans les cartes factorielles de scores, différentes structures spatiales étaient visuellement mises en évidence qui pourrait potentiellement révéler quelques indices interprétatifs en fonction de la situation du vignoble (zone géographique, caractéristiques du sol et capacité hydrique disponible), des différents stades phénologiques des vignobles (différents label européen, cépage et itinéraire technico-culturel) et de l’impact du recours à l’irrigation dans certaines zones.

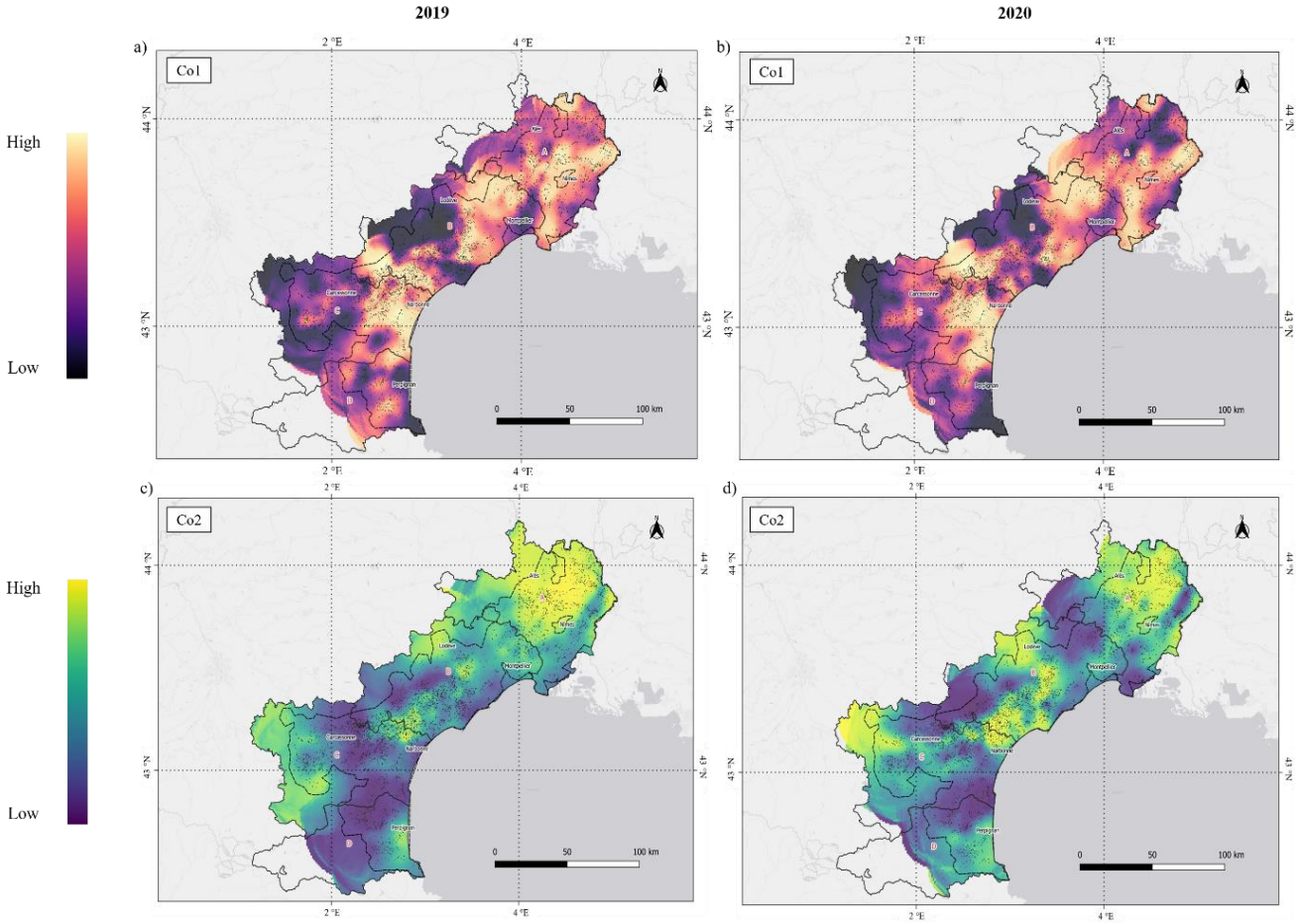


FIGURE R. 1 - Cartes krigées des valeurs de scores obtenus pour les blocs de vignobles à l’échelle régionale pour les années 2019 et 2020 en distinguant pour a) la composante 1 identifiée comme le profil spectro-temporel du sol et b) la composante 2 identifiée comme le profil spectro-temporel de la végétation. Les couleurs claires représentent des valeurs de score élevées et les couleurs foncées des valeurs de score faibles pour les deux composantes.

Dans cette première approche, nous avons émis l'hypothèse que l'organisation des données issues d'instruments multi-spectraux sous la forme d'une structure multidimensionnelle (cube de données) était une proposition pertinente pour extraire le meilleur potentiel des séries temporelles d'images multi-spectrales. Toutefois, l'utilisation de la méthode PARAFAC nécessite de réaliser une interpolation temporelle des données pour obtenir une structuration multidimensionnelle continue des informations, compte tenu du caractère irrégulier temporellement et spatialement des images Sentinel-2 dues à des conditions atmosphériques défavorables. Par conséquent, pour la deuxième approche, nous avons proposé l'application d'une approche de modélisation multidimensionnelle non supervisée capable de gérer la haute dimensionnalité des données, mais sans avoir besoin d'appliquer une interpolation temporelle qui pourrait masquer des informations pertinentes pour un bon suivi des vignobles à l'échelle régionale. Pour tenter de résoudre ce problème, la méthode Multivariate Curve-Resolution Alternated Least Squares (MCR-ALS) a été identifiée comme une approche prometteuse pour reconfigurer la dimension temporelle dans l'analyse. En fait, la méthodologie MCR-ALS pourrait fonctionner avec les séries temporelles originales des données Sentinel-2. En travaillant avec les données temporelles originales, elle permet d'obtenir des informations réelles date par date puisqu'aucun processus d'interpolation n'est nécessaire. L'application de la méthodologie a été axée sur 4978 blocs de vignes dans la région du Languedoc-Roussillon pendant la période de mai à août 2019. D'un point de vue exploratoire, cette approche s'est révélée capable de (i) de représenter la variabilité spatiale des composantes (sol-végétation) pour les dates de revisite des satellites Sentinel-2, c'est-à-dire sans interpolation temporelle et qui (ii) de visualiser dans l'espace puisqu'elle aborde deux dimensions spatiales (pixels x-y), en créant directement des 'images'. Cependant, la méthode MCR-ALS n'est pas allée au-delà de la détermination d'une alternative potentielle pour s'adapter aux caractéristiques temporelles de l'ensemble de données, car aucune validation selon des critères agronomiques n'a été effectuée.

Par conséquent, le potentiel des approches de modélisation multidimensionnelle non supervisées capables de considérer le temps comme une variable supplémentaire dans l'analyse a été établi. De plus, la première approche (PARAFAC) a établi l'importance de prendre en compte la dimension temporelle afin de fournir des informations pour le suivi des cultures à l'échelle régionale.

Chapitre 4. Potentiel des méthodes multidimensionnelles supervisées pour l'exploration régionale des données sur les cultures à partir de séries temporelles multi-spectrales

Aux grandes échelles spatiales d'étude, il est souvent difficile de mesurer toutes les données nécessaires et pertinentes pour appliquer des fonctions de modélisation descriptives et prédictives. Cependant, lorsqu'il est possible de mesurer certaines variables de référence provenant du terrain, la caractérisation d'un phénomène agronomique peut s'avérer être une tâche moins complexe. Dans ce contexte, une approche de modélisation multidimensionnelle supervisée a été proposée, comme modèle prédictif, pour estimer une information agronomique, ici le rendement d'une culture à partir de diverses données mesurées. En particulier, une méthode supervisée appelée N-way Partial Least Squares (N-PLS) (Bro, 1996) a été utilisée pour aborder la question de la réponse multivariée possible des vignobles à l'échelle régionale à une vague de chaleur, sur la base de séries temporelles satellitaires. Plus précisément, l'étude de cas s'est basée sur l'impact d'une vague de chaleur à la fin du mois de juin 2019 sur les rendements de récolte de la vigne dans la région Languedoc-Roussillon.

Puisque les données de télédétection peuvent être organisées de manière significative dans une structure multidimensionnelle (par exemple un cube, voir chapitre 3), l'hypothèse était que le suivi de l'évolution de certains phénomènes, comme la croissance végétative, en considérant simultanément la dimension spectrale et temporelle, pourrait avoir le potentiel de révéler une situation de stress sur des échelles de temps, car le déclencheur de stress et son effet varient considérablement dans le temps. Par conséquent, dans un premier temps, le potentiel de la méthodologie N-PLS en tant que technique de modélisation utile pour l'analyse de séries temporelles d'images multi-spectrales a été présenté dans le cadre d'une étude de cas portant sur 107 parcelles de vignes présentant différents degrés de pertes de rendement (signalées par les viticulteurs et les conseillers) dues à l'impact d'une vague de chaleur. La température maximale de la canicule a été enregistrée le 28 juin 2019, avec 45°C. Dans un deuxième temps, le modèle qui a été calibré à partir des 107 parcelles de vignes a été appliqué pour caractériser d'un point spectral, temporel et spatial les effets de cet épisode de canicule sur 4978 autres parcelles de vignes dans la région Languedoc-Roussillon en capturant efficacement les relations causales entre la réponse à un facteur spécifique à la culture, c'est-à-dire la canicule (facteur) et la perte de rendement de la vigne (réponse) à une échelle régionale.

L'application de la méthode N-PLS pour caractériser et estimer l'impact d'un événement extrême, tel qu'une vague de chaleur sur 107 parcelles de vignes, a montré que la performance des modèles de calibration, lorsqu'ils ont été appliqués aux données de validation, offrait une qualité de prédiction (R^2) des pertes de rendement avec une performance de 0,66 et une erreur (RMSE= de 10 %. Outre le modèle d'estimation des pertes de rendement, l'analyse

N-PLS a montré l'intérêt d'adopter une analyse systémique qui tient compte simultanément des caractéristiques spectrales et temporelles des données considérées. Notamment, la nature linéaire de la réduction de la dimensionnalité dans la méthode présentée a permis une interprétation simple en utilisant les *b*-coefficients calculés, qui sont directement liés à l'importance des variables explicatives, c'est-à-dire à la meilleure signature spectrale liée à l'événement étudié. Par l'intermédiaire de l'analyse des *b*- coefficients, il a été possible de déterminer que les bandes spectrales autour du 'Red-Edge (700 nm) et de la région 'Short-wave InfraRed' (1600 nm) à la fin du mois de juin et au début du mois de juillet étaient les plus pertinentes pour quantifier l'effet d'une vague de chaleur survenue à la fin du mois de juin, car leur profil spectral a montré un approfondissement du profil temporel pendant la période de stress thermique (Figure R.2). Ces résultats ont démontré la pertinence des séries temporelles multi-spectrales par satellite pour évaluer l'impact d'une vague de chaleur sur les pertes de vignes lorsqu'elles sont combinées au modèle N-PLS. Cependant, ce modèle a été calibré qu'avec un petit nombre de parcelles de vignes réparties sur une petite partie représentative de la région. Il a donc été jugé approprié d'étudier si le modèle calibré sur ce petit ensemble de données pouvait être appliqué à l'ensemble de la région afin de tester sa capacité à révéler les empreintes spatiales des vagues de chaleur à l'échelle régionale (4978 parcelles de vignobles). L'application de l'approche N-PLS à l'échelle régionale, bien que calibrée sur quelques parcelles, a permis de démontrer la robustesse de la méthodologie et sa capacité à cartographier l'empreinte spatiale de la perte de rendement due à la canicule qui a touché le sud de la France en 2019 (Figure R.3). En conclusion, l'approche multidirectionnelle N-PLS a montré un grand potentiel pour prédire, identifier et caractériser des phénomènes spectraux, temporels et surtout dépendants du temps, tels que l'impact du stress sur des parcelles de vignes à une échelle régionale.

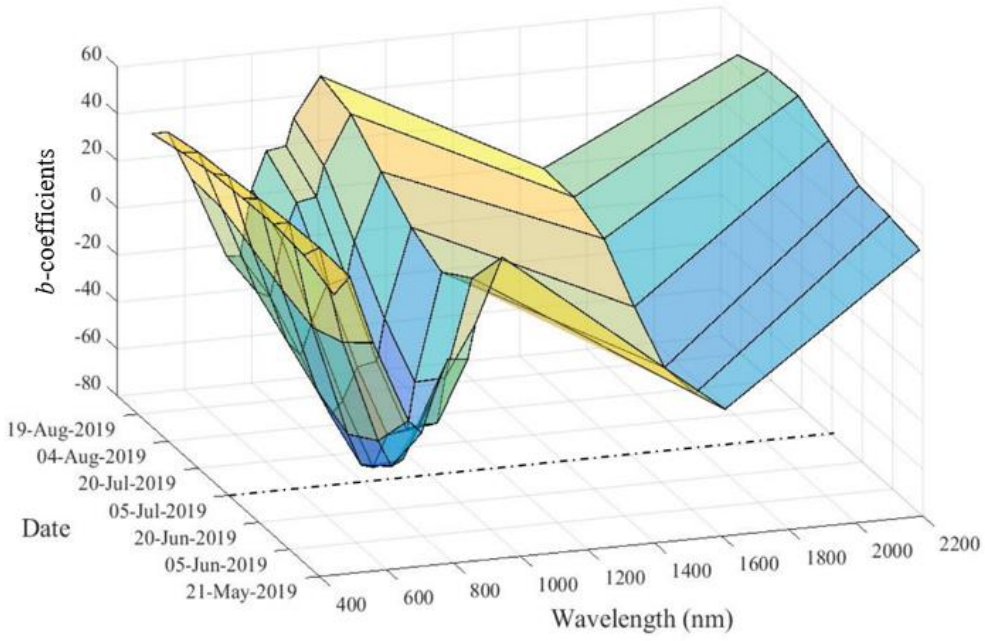


FIGURE R. 2 - Représentation combinée des profils temporels et spectraux sous la forme d'une vue 3D des *b*-coefficients N-PLS correspondants. Ligne pointillée noire représente la date la plus pertinente de la canicule.

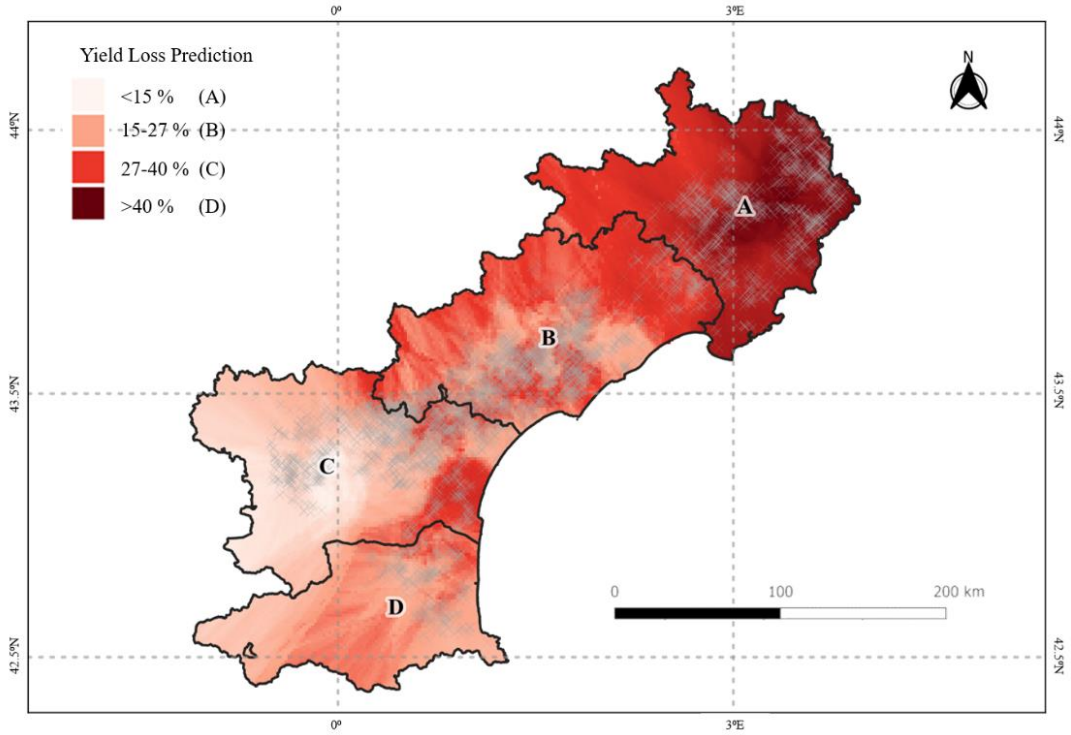


FIGURE R. 3 - Carte krigée des prédictions de perte de rendement à une échelle régionale dérivée du modèle N-PLS.

Chapitre 5. Détermination, à partir de séries temporelles multi-spectrales, des domaines spectraux et temporels les plus discriminants pour la caractérisation de la croissance végétative de la vigne en relation avec un événement climatique extrême

Du point de vue de la visualisation et de l'interprétation des données, l'analyse multidirectionnelle permet de simplifier les résultats, en fournissant des modèles plus adéquats et robustes utilisant relativement peu de paramètres d'entrées (Salvatore et al., 2013). En effet, il a été montré qu'à partir d'une approche de modélisation multidimensionnelle supervisée issue de la chimiométrie qui considère simultanément la dimension spectrale et temporelle, il est possible d'identifier les signatures spectrales-temporelles ainsi que l'empreinte spatiale capable d'expliquer un événement météorologique extrême tel qu'une vague de chaleur. Cependant, à un niveau plus opérationnel, il est important de spécifier plus en détail quels domaines spectraux et temporels sont affectés par les changements de croissance de la végétation pendant une période de stress afin d'optimiser le suivi de croissance des cultures. À cette fin, il a été proposé d'approfondir les domaines spectral et temporel à l'aide d'une méthode de sélection des variables à plusieurs voies, N-way Covariance Selection (N-CovSel), qui a été développé pour évaluer la pertinence des variables explicatives en fonction de leur covariance avec la ou les réponses, c'est-à-dire la variable d'intérêt à prédire. La méthodologie N-CovSel a été mise en œuvre pour l'extraction d'éléments satellitaires multi-spectraux afin de caractériser les bandes ou domaines spectraux, ainsi que les dates ou périodes déterminantes pour comprendre et identifier les différences d'effets au sein de 107 parcelles de vignes touchés par la canicule survenue entre le 23 juin et le 8 juillet 2019 dans la région Languedoc-Roussillon. La méthodologie N-CovSel proposée a permis de gérer la sélection de variables hautement dimensionnelles, en tenant compte de la relation inhérente entre les variations de réflectance et les fenêtres temporelles dans lesquelles elles se sont produites. Il a donc été démontré par une modélisation prédictive ultérieure de ces variables (c'est-à-dire un modèle pour chaque type de variable) que les variables spectrales, temporelles ou spectro-temporelles sélectionnées par la méthodologie N-CovSel étaient pertinentes pour caractériser la disparité des pertes de rendement de récolte à une échelle régionale pouvant résulter d'un événement climatique extrême parvenu au début de l'été 2019. La performance des modèles a été évaluée par le R^2 et l'erreur standard de prédiction (SEP) comme suit : pour le modèle N-PLS temporel (R^2 = 0,62 - SEP= 11,4 %), pour le modèle N-PLS spectral (R^2 = 0,61 - SEP = 13 %) et le modèle PLS temporel-spectral (R^2 = 0,63 - SEP= 11,7 %).

En outre, les domaines spectraux les plus liés à la perte de rendement des parcelles de vigne (observée par les consultants et les viticulteurs) étaient la combinaison du 'SWIR' (1610-

2190 nm), du ‘Red-Edge’ (705 nm) et du ‘Vegetation Red-Edge’ (865 nm) (Figure R.4). Les régions spectrales sélectionnées par la méthodologie sont utiles pour la détection du stress hydrique dans les vignobles (Laroche-Pinel et al., 2021b). En effet, la figure R.4 illustre que bien que la canicule ait eu un impact sur les vignobles, sa relation de cause à conséquence n’est pas directe. D’autres facteurs, tels que l’état initial des cultures, se sont avérés importants pour caractériser l’importance de l’impact subi par les vignobles. Par conséquent, il a été conclu que l’implication de la saisonnalité dans l’expression de la croissance végétative est fondamentale pour caractériser les pertes de rendement.

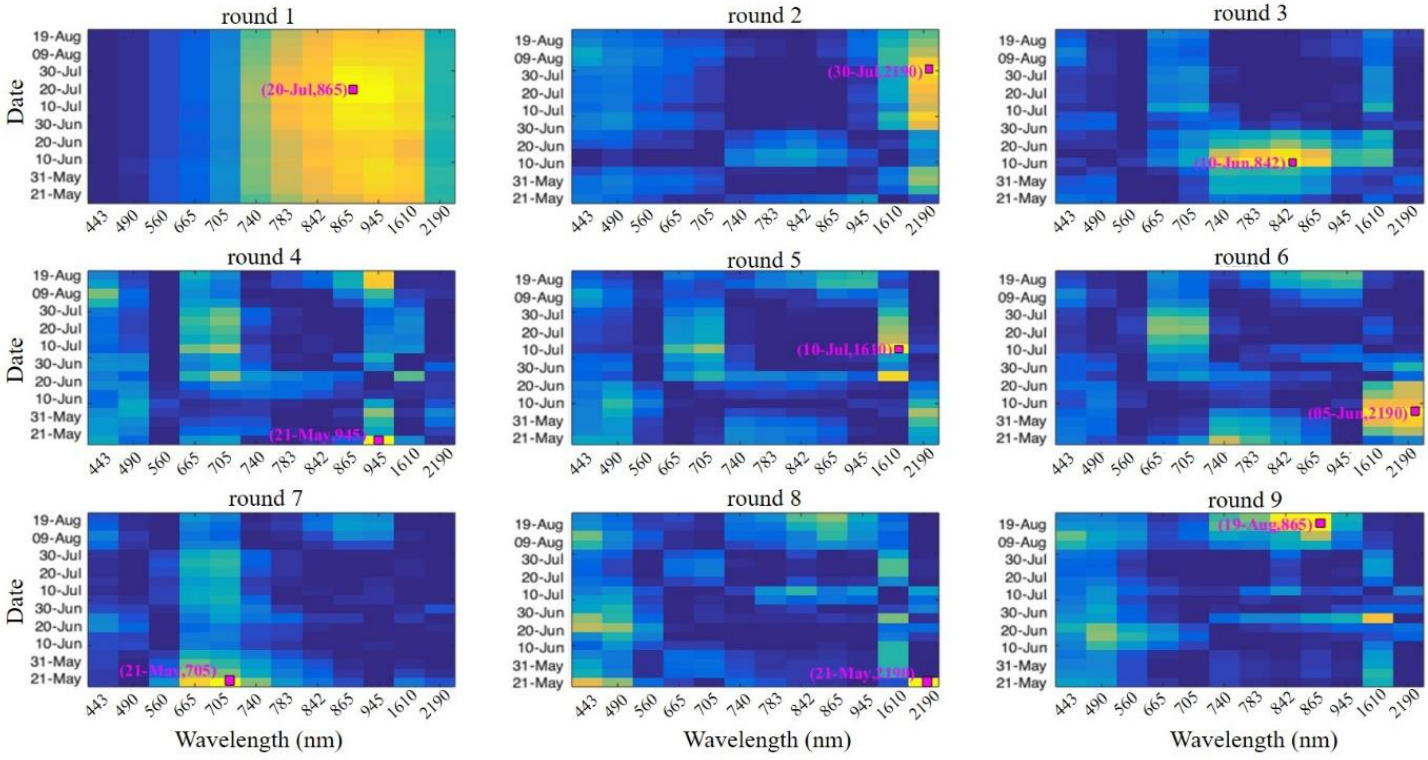


FIGURE R. 4 - Carte d’évolution du paramètre $\boldsymbol{cov^2(x_{jk}, y)}$ pour les 9 premières paires spectro-temporelles (date-longueur d’onde) sélectionnées par N-CovSel. Pour chacune des 9 paires, les dates et longueurs d’onde sélectionnées sont mises en évidence par un carré rouge. Les dates et les longueurs d’onde sont écrites en rose. Le gradient de couleur représente, du jaune au bleu, les valeurs de covariance les plus élevées et les plus faibles entre la paire date-longueur d’onde sélectionnée et le vecteur **y**, respectivement.

Chapitre 6. Conclusion

L'analyse indépendante des dimensions spectrales-temporelles des données de télédétection ne permet pas une analyse détaillée de la dynamique de croissance des cultures au niveau des parcelles sur de grandes échelles spatiales. L'hypothèse formulée était que les séries temporelles à haute résolution (c'est-à-dire moins d'une semaine de temps de revisite) pourraient être une option appropriée pour le suivi des cultures. En effet, la disponibilité des séries temporelles d'images satellites qui existent aujourd'hui à partir de données acquises par différentes plateformes satellitaires, telles que Sentinel-2, permet un suivi en temps quasi réel. Néanmoins, il faut tenir compte du fait que l'exploitation de ces données de télédétection multi-temporelles et multi-spectrales pour des applications à grande échelle soulève des problèmes d'intensité de données en raison de la grande quantité d'informations hétérogènes qui sont collectées. Cela signifie que si la dimension temporelle est ajoutée comme variable explicative de la variation spectrale à caractériser, cela conduit au domaine de la gestion des données à haute dimension. De manière globale, les résultats de la thèse ont mis en évidence d'une part, la pertinence de l'utilisation d'approche de modélisation issue de la chimiométrie pour analyser des ensembles de données de télédétection spatio-temporelles et spectrales et d'autre part l'importance de prendre en compte la dimension temporelle de ces données lors de la génération d'informations pour le suivi des cultures à l'échelle régionale. Ce travail a démontré que les méthodes d'analyse qui exploitent les signatures temporelles et spectrales pour extraire des informations sur les variations de la croissance végétative à l'échelle régionale fournissent des informations précieuses et pertinentes pour évaluer le rendement de récolte d'une culture, telle que la vigne.

Plusieurs perspectives peuvent être tirées de cette recherche. En tenant compte de la possibilité d'identifier et de caractériser des phénomènes avec une évolution temporelle évidente (i.e. une vague de chaleur) qui se traduit par un changement dans le profil spectro-temporel de la végétation des cultures, il serait intéressant de déterminer si d'autres types de phénomènes qui varient de manière moins prononcée (i.e. le début d'un stress hydrique) peuvent également être évalués. Il serait également intéressant de poursuivre les recherches pour valider les approches proposées dans (i) d'autres régions viticoles (ii) d'autres cultures agricoles, et (iii) d'autres applications que le suivi de croissance des cultures, car elles peuvent également être adaptées à des domaines plus généraux de l'agriculture tels que le suivi de la couverture et de l'utilisation des sols. En outre, cette recherche n'a utilisé que des séries chronologiques d'images de télédétection Sentinel-2. Par conséquent, une prochaine étape serait d'envisager d'autres sources d'information pour étudier et caractériser plus en détail le suivi spatial dynamique, de la parcelle à l'échelle régionale, afin de mieux répondre aux conditions spécifiques de tous les types de cultures.

Par ailleurs, dans cette recherche, l'intégration de la dimension spatiale a été limitée à une intégration a posteriori de la dimension par des méthodes géostatistiques. La dimension spatiale n'est pas intégrée dans l'analyse principale (spectro-temporelle), car la nature des données spatiales ne coïncide pas avec la nature des données spectrales ou temporelles, qui peuvent être représentées comme un 'spectre', adapté à l'analyse par des approches de modélisation multidimensionnelle issues de la chimiométrie. Ainsi, une prochaine étape consisterait à traiter simultanément les dimensions spectrale-temporelle-spatiale des données de télédétection. Une première voie pourrait être basée sur l'intégration de la dimension spatiale au moyen de contraintes spatiales dans le processus d'optimisation de l'algorithme à appliquer. Un exemple clair d'une telle procédure est la méthodologie Multivariate Curve-Resolution Alternated Least Squares (MCR-ALS), qui applique des contraintes appropriées pour imposer des caractéristiques spatiales spécifiques à l'ensemble des données. La deuxième voie possible serait un changement de paradigme dans l'intégration de la dimension spatiale. En effet, le principal défi de la prise en compte simultanée des dimensions spectrale, temporelle et spatiale dans une seule approche de modélisation multidimensionnelle est que la nature des données spatiales ne coïncide pas avec la nature des données spectrales ou temporelles. Cette intégration doit donc représenter la dimension spatiale non pas comme un simple attribut discontinu, mais comme un 'spectre' continu, tout comme sont représentées les dimensions spectrale et temporelle.

Bibliography

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). Wiley interdisciplinary reviews. *Computational Statistics, 2*(1), 97–106

Abdi, H., & Williams, L. J. (2013). Partial least squares methods: partial least squares correlation and partial least square regression. In *Computational toxicology* (pp. 549-579). Humana Press, Totowa, NJ.

Acevedo-Opazo, C., Tisseyre, B., Guillaume, S., & Ojeda, H. (2007). Test of NDVI information for a relevant vineyard zoning related to vine water status. In *Precision agriculture'07. Papers presented at the 6th European Conference on Precision Agriculture, Skiathos, Greece, 3-6 June, 2007* (pp. 547-554). Wageningen Academic Publishers.

Acevedo-Opazo, C., Tisseyre, B., Guillaume, S., Ortega-Farias, S., & Ojeda, H. (2008). Is it possible to assess the spatial variability of vine water status? *Journal International des Sciences de la Vigne et du Vin, 42*(4), 203-219.

Ahmad Fadzillah, N., Rohman, A., A, I., Shuhaimi, M., & Khatib, A. (2013). Application of FTIR-ATR Spectroscopy Coupled with Multivariate Analysis for Rapid Estimation of Butter Adulteration. *Journal of Oleo Science, 62*(5), 555–62.

Ahmad, U., Nasirahmadi, A., Hensel, O., & Marino, S. (2022). Technology and Data Fusion Methods to Enhance Site-Specific Crop Monitoring. *Agronomy, 12*(3), 555.

Alam, M. S., Islam, M. N., Bal, A., & Karim, M. A. (2008). Hyperspectral target detection using Gaussian filter and post-processing. *Optics and Lasers in Engineering, 46*(11), 817–822.

Albanwan, H., & Qin, R. (2018). A novel spectrum enhancement technique for multi-temporal, multi-spectral data using spatial-temporal filtering. *ISPRS Journal of Photogrammetry and Remote Sensing, 142*, 51–63.

Allen, W. A., Gausman, H. W., Richardson, A. J., & Thomas, J. R. (1969). Interaction of Isotropic Light with a Compact Plant Leaf. *JOSA, 59*(10), 1376-1379.

Amigo, J. M., Martí, I., & Gowen, A. (2013). Chapter 9 - Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality. In F. Marini (Éd.), *Data Handling in Science and Technology* (Vol. 28, pp. 343-370). Elsevier.

Amorós-López, J., Gómez-Chova, L., Alonso, L., Guanter, L., Zurita-Milla, R., Moreno, J., & Camps-Valls, G. (2013). Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. *International Journal of Applied Earth Observation and Geoinformation, 23*, 132-141.

Andersson, C., & Bro, R. (2000). The N-way Toolbox for Matlab. *Chemometrics and Intelligent Laboratory Systems, 52*(1), 1–4.

Aparicio, N., Villegas, D., Casadesus, J., Araus, J. L., & Royo, C. (2000). Spectral Vegetation Indices as Nondestructive Tools for Determining Durum Wheat Yield. *Agronomy Journal, 92*(1), 83-91.

Arenas-Garcia, J., & Camps-Valls, G. (2007). Feature extraction from remote sensing data using Kernel Orthonormalized PLS. In *2007 IEEE International Geoscience and Remote Sensing Symposium, Barcelona, Spain, 23-27 July, 2007* (pp. 258-261). IEEE

Atzberger, C. (2013). Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sensing, 5*(2), 949-981.

Ballester, C., Zarco-Tejada, P. J., Nicolás, E., Alarcón, J. J., Fereres, E., Intrigliolo, D. S., & Gonzalez-Dugo, V. (2018). Evaluating the performance of xanthophyll, chlorophyll and structure-sensitive spectral indices to detect water stress in five fruit tree species. *Precision Agriculture, 19*(1), 178–193.

Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews, 13*(1-2), 95-120.

Baret, F., Houlès, V., & Guérif, M. (2007). Quantification of plant stress using remote sensing observations and crop models: The case of nitrogen management. *Journal of Experimental Botany, 58*(4), 869-880.

Basso, B., Cammarano, D., & De Vita, P. (2004). Remotely sensed vegetation indices: Theory and applications for crop management. *Rivista Italiana di Agrometeorologia, 1*(5), 36-53.

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., De Abelleyra, D., P. D. Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., & R. Verón, S. (2018). Remote Sensing and Cropping Practices: A Review. *Remote Sensing, 10*(1), 99.

Bergant, K., & Kajfež-Bogataj, L. (2005). N–PLS regression as empirical downscaling tool in climate change studies. Theoretical and Appied. *Climatoology, 81*(1), 11–23.

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., & Hank, T. (2018). Evaluation of the PROSAIL Model Capabilities for Future Hyperspectral Model Environments: A Review Study. *Remote Sensing, 10*(1), 85.

Biancolillo, A., Marini, F., & Roger, J.M. (2021). N-CovSel, a new strategy for feature selection in N-way data. In *17th Scandinavian Symposium of Chemometrics (SSC17), Aalborg (DK), 6-9 September, 2021*. https://ssc17.org/Program_abstracts.pdf

Biancolillo, A., Marini, F., & Roger, J.M. (2022). N-CovSel, a new strategy for feature selection in N-way data. Submitted to *Analytica Chimica Acta*.

Bishop, M. P. (2013). Remote Sensing and GIScience in Geomorphology: Introduction and Overview. In J. Shroder & M. P. Bishop (Eds.), *Treatise on Geomorphology. Remote Sensing and GIScience in Geomorphology* (Vol. 3, pp. 1–24). San Diego: Academic Press.

Blanchet, L. (2008). *Méthodes de résolution dédiées à l'étude spectroscopique de processus photoinduits. Adaptation aux spécificités des spectres résolus en temps* (Doctoral dissertation, Université des Sciences et Technologie de Lille-Lille I).

Bovolo, F., & Bruzzone, L. (2015). The Time Variable in Data Fusion: A Change Detection Perspective. *IEEE Geoscience and Remote Sensing Magazine, 3*(3), 8-26.

Bro, R. (1996). Multiway calibration. Multilinear PLS. *Journal of Chemometrics, 10*(1), 47–61.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems, 38*(2), 149–171.

Bro, R., Smilde, A., & Jong, S. D. (2001). On the difference between low-rank and subspace approximation: improved model for multi-linear PLS regression. *Chemometrics and Intelligent Laboratory Systems, 58*(1), 3-13.

Brunel, G., Pichon, L., Taylor, J., & Tisseyre, B. (2019). Easy water stress detection system for vineyard irrigation management. In *12th European Conference on Precision Agriculture, ECPA 2019* (pp. 935-942). Wageningen Academic Publishers.

Brunori, E., Maesano, M., Moresi, F. V., Antolini, A., Bellincontro, A., Forniti, R., Biasi, R., & Mencarelli, F. (2020). Using UAV-based remote sensing to assess grapevine canopy damage due to fire smoke. *Journal of the Science of Food and Agriculture, 100*(12), 4531-4539.

Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., & Konopka, A.E. (1994). Field-Scale Variability of Soil Properties in Central Iowa Soils. *Soil Science Society of America Journal, 58*(5), 1501–1511.

Carrão, H., Gonçalves, P., & Caetano, M. (2008). Contribution of multispectral and multitemporal information from MODIS images to land cover classification. *Remote Sensing of Environment, 112*(3), 986-997.

Carrere, V., Briottet, X., Jacquemoud, S., Marion, R., Bourguignon, A., Chami, M., Dumont, M., Minghelli-Roman, A., Weber, C., Lefevre-Fonollosa, M.-J., & Mandea, M. (2013). HYPXIM: A second generation high spatial resolution hyperspectral satellite for dual applications. In *5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (pp.1-4). IEEE.

Carvalho, L. C., Coito, J. L., Gonçalves, E. F., Chaves, M. M., & Amâncio, S. (2016). Differential physiological response of the grapevine varieties Touriga Nacional and Trincadeira to combined heat, drought and light stresses. *Plant Biology, 18*, 101–111.

Champagne, C., Staenz, K., Abdou, B., McNairn, H., & Deguise, J.-C. (2003). Validation of a hyperspectral curve-fitting model for the estimation of plant water content of agricultural canopies. *Remote Sensing of Environment, 87*(2-3), 148–160.

Chang, L., Peng-Sen, S., & Shi-Rong, L. (2016). A review of plant spectral reflectance response to water physiological changes. *Chinese Journal of Plant Ecology, 40*(1), 80-91.

Chen, B., Li, J., & Jin, Y. (2021). Deep Learning for Feature-Level Data Fusion: Higher Resolution Reconstruction of Historical Landsat Archive. *Remote Sensing 13*(2), 167.

Chen, D., Huang, J., & Jackson, T. J. (2005). Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands. *Remote Sensing of Environment, 98*(2), 225-236.

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big Data for Remote Sensing: Challenges and Opportunities. *Proceedings of the IEEE, 104*(11), 2207–2219.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S.V., Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment, 219*, 145–161.

Clevers, J. G. P. W., Büker, C., van Leeuwen, H. J. C., & Bouman, B. A. M. (1994). A framework for monitoring crop growth by combining directional and spectral remote sensing information. *Remote Sensing of Environment, 50*(2), 161-170.

Clevers, J. G. P. W., Kooistra, L., & Schaepman, M. E. (2008). Using spectral information from the NIR water absorption features for the retrieval of canopy water content. *International Journal of Applied Earth Observation and Geoinformation, 10*(3), 388–397.

Cogato, A., Meggio, F., De Antoni Migliorati, M., & Marinello, F. (2019a). Extreme Weather Events in Agriculture: A Systematic Review. *Sustainability, 11*(9), 2547.

Cogato, A., Pagay, V., Marinello, F., Meggio, F., Grace, P., & De Antoni Migliorati, M. (2019b). Assessing the Feasibility of Using Sentinel-2 Imagery to Quantify the Impact of Heatwaves on Irrigated Vineyards. *Remote Sensing, 11*(23), 2869.

Cogato, A., Wu, L., Jewan, S. Y. Y., Meggio, F., Marinello, F., Sozzi, M., & Pagay, V. (2021). Evaluating the Spectral and Physiological Responses of Grapevines (Vitis vinifera L.) to Heat and Water Stresses under Different Vineyard Cooling and Irrigation Strategies. *Agronomy, 11*(10), 1940.

Coppi, R. (1994). An introduction to multiway data and their analysis. Computational *Statistics & Data Analysis, 18*(1), 3–13.

Dalla Mura, M., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J., & Benediktsson, J. A. (2015). Challenges and Opportunities of Multimodality and Data Fusion. In Remote Sensing. *Proceedings of the IEEE, 103*(9), 1585-1601.

Das, B., Mahajan, G. R., & Singh, R. (2018). Hyperspectral remote sensing: use in detecting abiotic stresses in agriculture. In *Advances in Crop Environment Interaction* (pp. 317-335). Springer, Singapore

Davies, A. M. C., & Fearn, T. Back to basics: calibration statistics. *Spectroscopy Europe, 18*(2), 31-32

Dayal, B. S., & MacGregor, J. F. (1997). Improved PLS algorithms. *Journal of Chemometrics, 11*(1), 73–85.

De Boeck, H. J., Dreesen, F. E., Janssens, I. A., & Nijs, I. (2010). Climatic characteristics of heat waves and their simulation in plant experiments. *Global Change Biology, 16*(7), 1992–2000.

De Juan, A., & Tauler, R. (2001). Comparison of three-way resolution methods for non-trilinear chemical data sets. *Journal of Chemometrics, 15*(10), 749–771.

De Juan, A., & Tauler, R. (2019). Data fusion by multivariate curve resolution. In *Data handling in science and technology* (Vol. 31, pp. 205-233). Elsevier.

De Juan, A., Rutan, S. C., & Tauler, R. (2019). Two-Way Data Analysis: Multivariate Curve Resolution, Iterative Methods. S. Brown, R. Tauler, B. Walczak (Eds.), *Comprehensive Chemometrics,* (pp. 153-171). Elsevier, Oxford.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Bégué,

A., Dejoux, J.-F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P.S., Trore, S.S., & Koetz, B. (2019). Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sensing of Environment, 221*, 551-568.

Devaux, N., Crestey, T., Leroux, C., & Tisseyre, B. (2019). Potential of Sentinel-2 satellite images to monitor vine fields grown at a territorial scale. *OENO One, 53*(1), 52-59.

Dorigo, W. A., Zurita-Milla, R., de Wit, A. J. W., Brazile, J., Singh, R., & Schaepman, M. E. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation, 9(*2), 165–193.

Dowd, C. (2020). A new ECDF Two-Sample Test Statistic. *arXiv preprint arXiv:2007.01360*

Drost, E. (2011). Validity and Reliability in Social Science Research. *Education Research and Perspectives, 38*(1), 105–124.

Droulia, F., & Charalampopoulos, I. (2021). Future Climate Change Impacts on European Viticulture: A Review on Recent Scientific Advances. *Atmosphere, 12*(4), 495.

Dry, P., & Loveys, B. (1998). Factors influencing grapevine vigour and the potential for control with partial rootzone drying. *Australian journal of grape and wine research, 4*(3), 140-148.

Dubovik, O., Schuster, G. L., Xu, F., Hu, Y., Bösch, H., Landgraf, J., & Li, Z. (2021). Grand Challenges in Satellite Remote Sensing. *Frontiers in Remote Sensing, 2*, 619818.

Easterday, K., Kislik, C., Dawson, T. E., Hogan, S., & Kelly, M. (2019). Remotely Sensed Water Limitation in Vegetation: Insights from an Experiment with Unmanned Aerial Vehicles (UAVs). *Remote Sensing, 11*(16), 1853.

Egan, W. G. (1992). Polarization in remote sensing. *Polarization and Remote Sensing, 1747*, 2-48.

Esbensen, K., & Geladi, P. (1989). Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems, 7*(1-2), 67–86.

FAO. (2010). Global strategy to improve agricultural and rural statistics. *FAO.* https://www.fao.org/publications/card/en/c/4e42f708-0a9e-56fb-acb5-67f352273186/

Favilla, S., Durante, C., Vigni, M. L., & Cocchi, M. (2013). Assessing feature relevance in NPLS models by VIP. *Chemometrics and Intelligent Laboratory Systems, 129*, 76–86.

Fernandes-Silva, A., Oliveira, M., Paço, T. A., & Isabel-Ferreira. (2018). Deficit Irrigation in Mediterranean Fruit Trees and Grapevines: Water Stress Indicators and Crop Responses. In *Irrigation in Agroecosystems, London, UK, 2018* (pp. 1–35). IntechOpen

Fernández-Mena, H., Frey, H., Celette, F., Garcia, L., Barkaoui, K., Hossard, L., Naulleau, A., Métral, R., Gary, C., & Metay, A. (2021). Spatial and temporal diversity of service plant management strategies across vineyards in the south of France. Analysis through the Coverage Index. *European Journal of Agronomy, 123*, 126191.

Ferreira, K. R., Queiroz, G. R., Vinhas, L., Marujo, R. F. B., Simoes, R. E. O., Picoli, M. C. A., Camara, G., Cartaxo, R., Gomes, V.C.F., Santos, L. A., Sanchez, A. H., Arcanjo., J. S. Fronza, J.G., Noronha, C.A., Costa, R.W., Zaglia, M.C., Zioti, F., Korting, T.S., Soares, A.R, Chaves, M.E.D., & Fonseca, L.M.G. (2020). Earth Observation Data Cubes for Brazil: Requirements, Methodology and Products. *Remote Sensing, 12*(24), 4033.

Filella, I., Serrano, L., Serra, J., & Peñuelas, J. (1995). Evaluating Wheat Nitrogen Status with Canopy Reflectance Indices and Discriminant Analysis. *Crop Science, 35*(5), 1400–1405.

Fraga, H., Molitor, D., Leolini, L., & Santos, J. A. (2020). What is the impact of heatwaves on European viticulture? A modelling assessment. *Applied Sciences, 10*(9), 3030.

Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. (1965). Spectral Properties of Plants. *Appied Optics, 4*(1), 11–20.

Gausman, H. W., Allen, W. A., Myers, V. I., & Cardenas, R. (1969). Reflectance and Internal Structure of Cotton Leaves, Gossypium hirsutum L.1. *Agronomy Journal, 61*(3), 374-376.

Ge, Y., Zhang, X., Atkinson, P. M., Stein, A., & Li, L. (2022). Geoscience-aware deep learning: A new paradigm for remote sensing. *Science of Remote Sensing, 5*, 100047.

Gilbertson, J. K., & van Niekerk, A. (2017). Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Computers and Electronics in Agriculture, 142*, 50–58.

Giovos, R., Tassopoulos, D., Kalivas, D., Lougkos, N., & Priovolou, A. (2021). Remote Sensing Vegetation Indices in Viticulture: A Critical Review. *Agriculture, 11*(5), 457.

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment, 58*(3), 289–298.

Gitelson, A. A., Merzlyak, M. N., Zur, Y., Stark, R., & Gritz, U. (2001). Non-destructive and remote sensing techniques for estimation of vegetation status. In *Proceedings of the 3rd*

European Conference on Precision Agriculture (ECPA), Montpellier, France, 18–20 June 2001 (pp 205–210). Wageningen Academic Publishers

Goodarzi, M., & Freitas, M. P. (2009). On the use of PLS and N-PLS in MIA-QSAR: Azole antifungals. *Chemometrics and Intelligent Laboratory Systems, 96*(1), 59–62.

Greenwood, P. E., & Nikulin, M. S. (1996). *A Guide to Chi-Squared Testing* (Vol. 280). John Wiley & Sons.

Gremillion, K. J., & Piperno, D. R. (2009). Human Behavioral Ecology, Phenotypic (Developmental) Plasticity, and Agricultural Origins: Insights from the Emerging Evolutionary Synthesis. *Current Anthropology, 50(*5), 615-619.

Griffiths, G. H., & Lee, J. (2000). Landscape pattern and species richness; regional scale analysis from remote sensing. *International Journal of Remote Sensing, 21*, 2685–2704.

Hanafi, M., Ouertani, S. S., Boccard, J., Mazerolles, G., & Rudaz, S. (2015). Multi-way PLS regression: Monotony convergence of tri-linear PLS2 and optimality of parameters. *Computational Statistics & Data Analysis, 83,* 129–139.

Hansen, P. M., Jørgensen, J. R., & Thomsen, A. (2002). Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. *The Journal of Agriculture Science, 139*(3), 307–318.

Harshman, R. (1970). Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimodal factor analysis. *UCLA Working Papers in Phonetics, 16*, 1-84

Hassan-Esfahani, L., Torres-Rua, A., Jensen, A., & Mckee, M. (2017). Spatial Root Zone Soil Water Content Estimation in Agricultural Lands Using Bayesian-Based Artificial Neural Networks and High- Resolution Visual, NIR, and Thermal Imagery. *Irrigation and Drainage, 66*(2), 273-288.

Hatfield, J. L., Prueger, J. H., Sauer, T. J., Dold, C., O'Brien, P., & Wacha, K. (2019). Applications of Vegetative Indices from Remote Sensing to Agriculture: Past and Future. *Inventions, 4*(4), 71.

Héberger, K. (2008). Chapter 7 - Chemoinformatics - Multivariate mathematical–statistical methods for data evaluation. In K. Vékey, A. Telekes, & A. Vertes (Eds.), *Medical Applications of Mass Spectrometry* (p. 141-169). Elsevier.

Henrion, R. (1994). N-way principal component analysis theory, algorithms and applications. *Chemometrics and Intelligent Laboratory Systems, 25*(1), 1–23.

Herrero-Huerta, M., Hernández-López, D., Rodriguez-Gonzalvez, P., González-Aguilera, D., & González-Piqueras, J. (2014). Vicarious radiometric calibration of a multispectral sensor from an aerial trike applied to precision agriculture. *Computers and Electronics in Agriculture, 108*, 28-38.

Hird, J., & Mcdermid, G. (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment, 113*(1), 248-258.

Hollstein, A., Segl, K., Guanter, L., Brell, M., & Enesco, M. (2016). Ready-to-Use Methods for the Detection of Clouds, Cirrus, Snow, Shadow, Water and Clear Sky Pixels in Sentinel-2 MSI Images. *Remote Sensing, 8*(8), 666.

Holzman, M. E., Rivas, R. E., & Bayala, M. I. (2021). Relationship between TIR and NIR-SWIR as Indicator of Vegetation Water Availability. *Remote Sensing, 13*(17), 3371.

Hong, D., Gao, L., Yao, J., Zhang, B., Plaza, A., & Chanussot, J. (2020). Graph convolutional networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 59*(7), 5966-5978.

Hu, Q., Wu, W., Song, Q., Lu, M., Chen, D., Yu, Q., & Tang, H. (2017). How do temporal and spectral features matter in crop classification in Heilongjiang Province, China? *Journal of Integrative Agriculture, 16*(2), 324-336.

Hua, L., Wang, H., Sui, H., Wardlow, B., Hayes, M. J., & Wang, J. (2019). Mapping the Spatial-Temporal Dynamics of Vegetation Response Lag to Drought in a Semi-Arid Region. *Remote Sensing, 11*(16), 1873.

Huang, Y., Chen, Z., Yu, T., Huang, X., & Gu, X. (2018). Agricultural remote sensing big data: Management and applications. *Journal of Integrative Agriculture, 17*(9), 1915-1931.

Huo, L., Persson, H. J., & Lindberg, E. (2021). Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red & SWIR (NDRS). *Remote Sensing of Environment, 255*, 112240.

Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *International Journal of Photogrammetry and Remote Sensing, 80*, 91-106.

Inoue, Y., Sakaiya, E., Zhu, Y., & Takahashi, W. (2012). Diagnostic mapping of canopy nitrogen content in rice based on hyperspectral measurements. *Remote Sensing of Environment, 126*, 210–221.

Irons, J. R., & Petersen, G. W. (1981). Texture transforms of remote sensing data. *Remote Sensing of Environment, 11*, 359-370.

Isaksson, T., & Næs, T. (1988). The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. *Applied Spectroscopy, 42*(7), 1273–1284.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., François, C., & Ustin, S. L. (2009). PROSPECT + SAIL models: A review of use for vegetation characterisation. *Remote Sensing of Environment, 113*, S56-S66.

Jianya, G., Haigang, S., Guorui, M., & Qiming, Z. (2008). A review of multi-temporal remote sensing data change detection algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37*(B7), 757-762.

Jong, S. de (1998). Regression coefficients in multilinear PLS. *Journal of Chemometrics, 12(*1), 77–81.

Jopia, A., Zambrano Bigiarini, F., Pérez-Martínez, W., Vidal, P., & Molina, J. (2020). Time-Series of Vegetation Indices (VNIR/SWIR) Derived from Sentinel-2 (A/B) to Assess Turgor Pressure in Kiwifruit. *International Journal of Geo-Information, 9*(11), 641.

Junges, A., Fontana, D., Anzanello, R., & Bremm, C. (2017). Normalized difference vegetation index obtained by ground-based remote sensing to characterise vine cycle in Rio Grande do Sul, Brazil. *Ciência e Agrotecnologia, 41*, 543-553.

Keller, H. R., Roettele, Juergen., & Bartels, Hermann. (1994). Assessment of the Quality of Latent Variable Calibrations Based on Monte Carlo Simulations. *Analytical Chemistry, 66*(7), 937–943.

Khadse, G. K. (2012). Spectral Reflectance Characteristics of the Soils on Basaltic Terrain of Central Indian Plateau. *Journal Indian Society of Remote Sensing, 40*(4), 717–724.

Khanal, S., Kc, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sensing, 12*(22), 3783.

Kiers, H., & Mechelen, I. (2001). Three-Way Component Analysis: Principles and Illustrative Application. *Psychological methods, 6*(1), 84–110.

Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment, 1*(3), 155–159.

Kroonenberg, P. M., Harshman, R. A., & Murakami, T. (2009). Analysing Three-way Profile Data Using the Parafac and Tucker3 Models Illustrated with Views on Parenting. *Applied Multivariate Research, 13*(1), 5.

Lagouarde, J. P., Bhattacharya, B., Crébassol, P., Gamet, P., Adlakha, D., Murthy, C.S., Singhn S.K., Mishra, M., Niagm, R., Raju, P.V., Babu, S.S., Shukla, L.V., Pandya, M.R., Boulet, G., Briottet, X., Dadou, I., Dedieu, G., Gouhier, M., Hagolle, O., Irvine, M, Jacob, F.,

Kumar, K.K., Laignel, B., Maisongrandre, P., Mallick, K., Olioso, A., Ottlé, C., Roujean, J.L., Sobrino, J., Ramakrishnan, R., Sekhar, M., & Sarkar, S. (2019). Indo-french high-resolution thermal infrared space mission for earth natural resources assessment and monitoring-concept and definition of TRISHNA. In *ISPRS-GEOGLAM-ISRS Joint International Workshop on "Earth Observations for Agricultural Monitoring", 42*, 403.

Laroche-Pinel, E., Albughdadi, M., Duthoit, S., Chéret, V., Rousseau, J., & Clenet, H. (2021a). Understanding Vine Hyperspectral Signature through Different Irrigation Plans: A First Step to Monitor Vineyard Water Status. *Remote Sensing*, *13*(3), 536.

Laroche-Pinel, E., Duthoit, S., Albughdadi, M., Costard, A. D., Rousseau, J., Chéret, V., & Clenet, H.(2021b). Towards Vine Water Status Monitoring on a Large Scale Using Sentinel-2 Images. *Remote Sensing*, *13*(9), 1837.

Lawrence, R. L., & Ripple, W. J. (1998). Comparisons among Vegetation Indices and Bandwise Regression in a Highly Disturbed, Heterogeneous Landscape: Mount St. Helens, Washington. *Remote Sensing of Environment, 64*(1), 91-102.

Lee, H., Wang, J., & Leblon, B. (2020). Using Linear Regression, Random Forests, and Support Vector Machine with Unmanned Aerial Vehicle Multispectral Images to Predict Canopy Nitrogen Weight in Corn. *Remote Sensing*, *12*(3), 2071.

Leroux, C., Jones, H., Pichon, L., Guillaume, S., Lamour, J., Taylor, J., Naud, O., Crestey, T., Lablee, J-L., & Tisseyre, B. (2018). GeoFIS: An Open Source, Decision-Support Tool for Precision Agriculture Data. *Agriculture*, *8*(6), 73.

Li, L., Lin, D., Wang, J., Yang, L., & Wang, Y. (2020). Multivariate Analysis Models Based on Full Spectra Range and Effective Wavelengths Using Different Transformation Techniques for Rapid Estimation of Leaf Nitrogen Concentration in Winter Wheat. *Frontiers in Plant Science, 11.*

Li, S., Xu, L., Jing, Y., Yin, H., Li, X., & Guan, X. (2021). High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques. *International Journal of Applied Earth Observation and Geoinformation, 105*, 102640.

Liu, P. (2015). A survey of remote-sensing big data. *Frontiers in Environmental Science, 3.*

Liu, X., Deng, C., Zhao, B., & Chanussot, J. (2019). Multimodal-temporal fusion: Blending multimodal remote sensing images to generate image series with high temporal resolution. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 10083-10086). IEEE.

Liu, Y., Hiyama, T., Kimura, R., & Yamaguchi, Y. (2006). Temporal influences on Landsat5 Thematic Mapper image in visible band. *International Journal of Remote Sensing, 27*(15), 3183–3201.

Lobo, A., Legendre, P., Rebollar, J., Carreras, J., & Ninot, J. (2004). Land cover classification at a regional scale in Iberia: Separability in a multi-temporal and multi-spectral data set of satellite images. *International Journal of Remote Sensing, 25(*1), 205-213.

Loew, A., Stacke, T., Dorigo, W., de Jeu, R., & Hagemann, S. (2013). Potential and limitations of multidecadal satellite soil moisture observations for selected climate model evaluation studies. *Hydrology and Earth System Sciences, 17*(9), 3523-3542.

Lopez-Fornieles, E., Brunel, G., Devaux, N., Roger, J-M., & Tisseyre, B. (2022a). Is It Possible to Assess Heatwave Impact on Grapevines at the Regional Level with Time Series of Satellite Images? *Agronomy, 12*(3), 563.

Lopez-Fornieles, E., Brunel, G., Rancon, F., Gaci, B., Metz, M., Devaux, N., Taylor, J., Tisseyre, B., Roger, J-M. (2022). Potential of Multiway PLS (N-PLS) Regression Method to Analyse Time-Series of Multispectral Images: A Case Study in Agriculture. *Remote Sensing, 14*(1), 216.

Lorenzen, B., & Jensen, A. (1988). Reflectance of blue, green, red and near infrared radiation from wetland vegetation used in a model discriminating live and dead above ground biomass. *New Phytologist, 108*(3), 345–355.

Lovelle, B. R., Trambouze, W., & Jacquet, O. (2009). Évaluation de l'état de croissance végétative de la vigne par la 'méthode des apex'. *Le Progrès agricole et viticole, 126*(4), 77-88.

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems, 51*, 47-60.

Maimaitiyiming, M., Ghulam, A., Bozzolo, A., Wilkins, J. L., & Kwasniewski, M. T. (2017). Early Detection of Plant Physiological Responses to Different Levels of Water Stress Using Reflectance Spectroscopy. *Remote Sensing, 9*(7), 745.

Malegori, C., Nascimento Marques, E. J., de Freitas, S. T., Pimentel, M. F., Pasquini, C., & Casiraghi, E. (2017). Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. Talanta, 165, 112–116.

Martinez de Toda, F., Balda, P., & Oliveira, M. (2010). Estimation of vineyard water status (vitis vinifera l. CV. Tempranillo) from the developmental stage of the shoot tips. *Journal International des Sciences de la Vigne et du Vin, 44*(4), 201-206.

Martínez, A., & Gomez-Miguel, V. (2017). Vegetation index cartography as a methodology complement to the terroir zoning for its use in precision viticulture. *OENO One, 51*(3), 289.

Maynard, C. L., Lawrence, R. L., Nielsen, G. A., & Decker, G. (2007). Modeling Vegetation Amount Using Bandwise Regression and Ecological Site Descriptions as an Alternative to Vegetation Indices. *GIScience & Remote Sensing*, *44*(1), 68-81.

Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems, 118*, 62–69.

Messai, H., Farman, M., Sarraj-Laabidi, A., Hammami-Semmar, A., & Semmar, N. (2016). Chemometrics Methods for Specificity, Authenticity and Traceability Analysis of Olive Oils: Principles, Classifications and Applications. *Foods, 5*(4), 77.

Mishra, P., Roger, J.M., & Rutledge, D. N. (2021). A short note on achieving similar performance to deep learning with practical chemometrics. *Chemometrics and Intelligent Laboratory Systems, 214*, 104336.

Mishra, S., Mishra, D., & Santra, G. H. (2016). Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper. *Indian Journal of Science and Technology, 9*(38).

Montero, F. J., Meliá, J., Brasa, A., Segarra, D., Cuesta, A., & Lanjeri, S. (1999). Assessment of vine development according to available water resources by using remote sensing in La Mancha, Spain. *Agricultural Water Management, 40*(2-3), 363-375.

Navalgund, R. R., Jayaraman, V., & Roy, P. S. (2007). Remote sensing applications: An overview. *Current Science, 93*(12), 1747-1766.

Neethling, E., Barbeau, G., Coulon-Leroy, C., & Quénol, H. (2019). Spatial complexity and temporal dynamics in viticulture: A review of climate-driven scales. *Agricultural and Forest Meteorology*, *276*, 107618.

Nicholas, K. A., & Durham, W. H. (2012). Farm-scale adaptation and vulnerability to environmental stresses: insights from winegrowing in Northern California. Global Environmental *Change, 22*(2), 483–494.

Ojeda, H., Andary, C., Kraeva, E., Carbonneau, A., & Deloire, A. (2002). Influence of pre- and post-veraison water deficit on synthesis and concentration of skin phenolic compounds during berry growth of Vitis Vinifera cv. Shiraz. *American Journal of Enology and Viticulture, 53*(4), 261-267.

Ojeda, H., Carrillo, N., Deis, L., Tisseyre, B., Heywang, M., & Carbonneau, A. (2005). Precision viticulture and water status II: Quantitative and qualitative performance of different within field zones, defined from water potential mapping. In *XIV International GESCO Viticulture Congress, Geisenheim, Germany, 23-27 August, 2005* (pp. 741-748). Groupe d'Etude des Systemes de Conduite de la vigne (GESCO).

Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *CATENA, 113*, 56–69.

Ouertani, S. S. (2014). *N-way Partial Least Squares : nouvelles propriétés et apports à la discrimination des données métabolomiques* (Doctoral dissertation, University of Geneva).

Padoan, R., Steemers, T. A., Klein, M., Aalderink, B., & De Bruin, G. (2008). Quantitative hyperspectral imaging of historical documents: technique and applications. *Art Proceedings*, 25-30.

Pellegrino, A., Lebon, E., Voltz, M., & Wery, J. (2005). Relationships between plant and soil water status in vine (Vitis vinifera L.). Plant Soil, 266(1), 129-142.

Peng, M., Zhang, L., Sun, X., Cen, Y., & Zhao, X. (2020). A Fast Three-Dimensional Convolutional Neural Network-Based Spatiotemporal Fusion Method (STF3DCNN) Using a Spatial-Temporal-Spectral Dataset. *Remote Sensing 12*(23), 3888.

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H., & Turner, W. (2014). Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology, 51*(4), 839–848.

Phatak, A., & Jong, S. D. (1997). The geometry of partial least squares. *Journal of Chemometrics, 11(*4), 311–338.

Pichon, L., Leroux, C., Macombe, C., Taylor, J., & Tisseyre, B. (2019). What relevant information can be identified by experts on unmanned aerial vehicles' visible images for precision viticulture? *Precision Agriculture, 20(*2), 278–294.

Picoli, M. C., Simoes, R., Chaves, M., Santos, L. A., Sanchez, A., Soares, A., & Queiroz, G. R. (2020). CBERS data cube: a powerful technology for mapping and monitoring Brazilian biomes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 5*(3), 533-539.

Pinel, E. L., Duthoit, S., Costard, A. D., Rousseau, J., Hourdel, J., Vidal-Vigneron, Cheret, V., & Clenet, H. (2021). Monitoring vineyard water status using Sentinel-2 images: qualitative survey on five wine estates in the south of France. *OENO One, 55*(4), 115–127.

Plant, R. E., Munk, D. S., Roberts, B. R., Vargas, R. L., Rains, D. W., Travis, R. L., & Hutmacher, R.B. (2000). Relationships between remotely sensed reflectance data and cotton growth and yield. *Transactions of the ASAE, 43*(3), 535

Qiu, J., Crow, W. T., Wagner, W., & Zhao, T. (2019). Effect of vegetation index choice on soil moisture retrievals via the synergistic use of synthetic aperture radar and optical remote sensing. *International Journal of Applied Earth Observation and Geoinformation, 80,* 47-57.

Quintano, C., Fernandez-manso, A., & Fernández-Manso, O. (2018). Combination of Landsat and Sentinel-2 MSI data for initial assessing of burn severity. *International Journal of Applied Earth Observation and Geoinformation, 64,* 221–225.

Rácz, A., Bajusz, D., & Héberger, K. (2018). Chemometrics in Analytical Chemistry. In *Applied Chemoinformatics* (p. 471-499). John Wiley & Sons, Ltd.

Raddi, S., Giannetti, F., Martini, S., Farinella, F., Chirici, G., Tani, A., & Maltoni, A., Mariotti, B. (2021). Monitoring drought response and chlorophyll content in Quercus by consumer-grade, near-infrared (NIR) camera: a comparison with reflectance spectroscopy. *New Forests, 53*(2), 241-265

Ray, S. S. (2019). Exploring Machine Learning Classification Algorithms for Crop Classification Using Sentinel 2 Data. International Archives of Photogrammetry, *Remote Sensing and Spatial Information Sciences, XLII-3/W6*, 573-578.

Raza, A., Razzaq, A., Mehmood, S. S., Zou, X., Zhang, X., Lv, Y., & Jinsong, X. (2019). Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review. *Plants, 8*(2), 34.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, M. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature, 566,* 195.

Rembold, F., Meroni, M., Urbano, F., Royer, A., Atzberger, C., Lemoine, G., Eerens, H., & Haesen, D. (2015). Remote sensing time series analysis for crop monitoring with the SPIRITS software: New functionalities and use examples. *Frontiers in Environmental Science, 3*.

Richter, K., Hank, T. B., Vuolo, F., Mauser, W., & D'Urso, G. (2012). Optimal Exploitation of the Sentinel-2 Spectral Capabilities for Crop Leaf Area Index Mapping. *Remote Sensing, 4*(3), 561–582.

Rienth, M., & Scholasch, T. (2019). State-of-the-art of tools and methods to assess vine water status. *OENO One, 53*(4), 619–637.

Roger, J. M., Palagos, B., Bertrand, D., & Fernandez-Ahumada, E. (2011). CovSel: Variable selection for highly multivariate and multi-response calibration. Application to IR spectroscopy. *Chemometrics and Intelligent Laboratory Systems, 106*(2), 216-223.

Roman, A., and Ursu, T. (2016). Roman, A., & Ursu, T. (2016). Multispectral satellite imagery and airborne laser scanning techniques for the detection of archaeological vegetation marks. Landscape archaeology on the northern frontier of the roman empire at porolissum: an interdisciplinary research project. *Cluj-Napoca: Mega Publishing House*, 141-152.

Rouse, J., Haas, R., Deering, D., Schell, J. A., & Harlan, J. (1973). *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation.* [NASA/GSFCT Type II Report]. Greenbelt, MD: NASA/Goddard Space Flight Center.

Roznik, M., Boyd, M., & Porth, L. (2022). Improving crop yield estimation by applying higher resolution satellite NDVI imagery and high-resolution cropland masks. *Remote Sensing Applications: Society and Environment, 25*, 100693.

Sagan, V., Maimaitijiang, M., Sidike, P., Maimaitiyiming, M., Erkbol, H., Hartling, S., Peterson, K., Peterson, J., Burken, J., & Fritschi, F. (2019). UAV/Satellite Multiscale Data Fusion for Crop Monitoring and Early Stress Detection. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives.*

Salvatore, E., Bevilacqua, M., Bro, R., Marini, F., & Cocchi, M. (2013). Classification Methods of Multiway Arrays as a Basic Tool for Food PDO Authentication. In *Comprehensive Analytical Chemistry* (Vol. 60, pp. 339-382). Elsevier.

Santesteban, L. G., Guillaume, S., Royo, J. B., & Tisseyre, B. (2013). Are precision agriculture tools and methods relevant at the whole-vineyard scale? *Precision Agriculture, 14*(1), 2-17.

Sarker, S. D., & Nahar, L. (2015). Chapter 19—Applications of High Performance Liquid Chromatography in the Analysis of Herbal Products. In P. K. Mukherjee (Ed.), *Evidence-Based Validation of Herbal Medicine* (p. 405-425). Elsevier.

Schaffer, B., & Andersen, P. (2018). *Handbook of Environmental Physiology of Fruit Crops. Volume I: Temperate Crops*. CRC Press, Boca Raton, Florida

Schneider, P., & Xhafa, F. (2022). Chapter 2 - Data stream processing: Models and methods: The complexity of data stream processing. In P. Schneider & F. Xhafa (Eds.), *Anomaly Detection and Complex Event Processing over IoT Data Streams* (p. 29-47). Academic Press.

Schymanski, S. J., Or, D., & Zwieniecki, M. (2013). Stomatal Control and Leaf Thermal and Hydraulic Capacitances under Rapid Environmental Fluctuations. *PloS one, 8*(1), e54231.

Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons Inc: Hoboken, NJ, USA.ISBN 978-0-471-69755-8.

Seelig, H. D., Hoehn, A., Stodieck, L. S., Klaus, D. M., Adams, W. W., & Emery, W. J. (2009). Plant water parameters and the remote sensing R1300/R1450 leaf water index: controlled condition dynamics during the development of water deficit stress. *Irrigation Science, 27*(5), 357–365.

Segarra, J., Buchaillot, M. L., Araus, J. L., & Kefauver, S. C. (2020). Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. *Agronomy, 10*(5), 641.

Sena, M. M., & Poppi, R. J. (2004). N-way PLS applied to simultaneous spectrophotometric determination of acetylsalicylic acid, paracetamol and caffeine. *Journal of Pharmaceutical and Biomedical Analysis, 34*(1), 27–34.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. *Social Service Review, 76*(3), 510-514.

Shanmugapriya, P., Rathika, S., Ramesh, T., & Janaki, P. (2019). Applications of Remote Sensing in Agriculture - A Review. *International Journal of Current Microbiology and Applied Sciences, 8*(1), 2270-2283.

Shaw, G., & Manolakis, D. (2002). Signal processing for hyperspectral image exploitation. *IEEE Signal Processing Magazine, 19(*1), 12–16.

Silleos, G., Alexandridis, T., Gitas, I., & Perakis, K. (2006). Vegetation indices: Advances made in biomass estimation and vegetation monitoring in the last 30 years. *Geocarto International, 21*(4), 21-28.

Sims, D. A., & Gamon, J. A. (2003). Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: a comparison of indices based on liquid water and chlorophyll absorption features. *Remote Sensing of Environment, 84*(4), 526–537.

Singh, A. (1989). Review Article Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing, 10*(6), 989-1003.

Sisheber, B., Marshall, M., Mengistu, D., & Nelson, A. (2022). Tracking crop phenology in a highly dynamic landscape with knowledge-based Landsat–MODIS data fusion. *International Journal of Applied Earth Observation and Geoinformation, 106,* 102670.

Slaton, M. R., Hunt, E. R., & Smith, W. K. (2001). Estimating Near-Infrared Leaf Reflectance from Leaf Structural Characteristics. *American Journal of Botany, 88*(2), 278–284.

Smilde, A. K. (1997). Comments on multilinear PLS. *Journal of Chemometrics, 11*(5), 367–377.

Solano, F., Di Fazio, S., & Modica, G. (2019). A methodology based on GEOBIA and WorldView-3 imagery to derive vegetation indices at tree crown detail in olive orchards. *International Journal of Applied Earth Observation and Geoinformation, 83*, 101912.

Soudani, K., Hmimina, G., Delpierre, N., Pontailler, J-Y., Aubinet, M., Bonal, D., Caquet, B., de Grandcourt, A., Burban, B., Flechard, C., Guyon, D., Granier, A., Gross, P., Heinesh, B., Longdoz, B., Loustau, D., Moureaux, C., Ourcival, J-M., Rambal, S., Saint André, L., & Dufrêne, E. (2012). Ground-based Network of NDVI measurements for tracking temporal dynamics of canopy structure and vegetation phenology in different biomes. *Remote Sensing of Environment, 123*, 234-245.

Southworth, J., & Muir, C. (2021). Specialty Grand Challenge: Remote Sensing Time Series Analysis. *Frontiers in Remote Sensing, 34.*

Sozzi, M., Kayad, A., Marinello, F., Taylor, J., & Tisseyre, B. (2020). Comparing vineyard imagery acquired from Sentinel-2 and Unmanned Aerial Vehicle (UAV) platform. *OENO One, 54*(2), 189–197.

Srivastava, P. K., Malhi, R. K. M., Pandey, P. C., Anand, A., Singh, P., Pandey, M. K., & Gupta, A. (2020). 1 - Revisiting hyperspectral remote sensing: Origin, processing, applications and way forward. In P. C. Pandey, P. K. Srivastava, H. Balzter, B. Bhattacharya, & G. P. Petropoulos (Eds.), *Hyperspectral Remote Sensing* (p. 3-21). Elsevier.

Stefan, V. G., Merlin, O., Er-Raki, S., Escorihuela, M-J., & Khabba, S. (2015). Consistency between in Situ, Model-Derived and High-Resolution-Image-Based Soil Temperature Endmembers: Towards a Robust Data-Based Model for Multi-Resolution Monitoring of Crop Evapotranspiration. *Remote Sensing, 7*(8), 10444-10479.

Tauler, R., Maeder, M., & de Juan, A. (2020). Multiset Data Analysis: Extended Multivariate Curve Resolution. S. Brown, R. Tauler, and B. Walczak (Eds.), *Comprehensive Chemometrics* (pp. 305–336). Elsevier, Oxford.

Terliksiz, A. S., & Altýlar, D. T. (2019). Use of deep neural networks for crop yield prediction: A case study of soybean yield in Lauderdale county, Alabama, USA. In *8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)* (pp.1-4). IEEE.

Teskey, R., Wertin, T., Bauweraerts, I., Ameye, M., Mcguire, M. A., & Steppe, K. (2015). Responses of tree species to heat waves and extreme heat events. *Plant, Cell & Environment, 38*(9), 1699–1712.

Thenkabail, P.S., Lyon, J.G., & Huete, A. (2011). *Hyperspectral remote sensing of vegetation*. CRC Press: Boca Raton, FL, USA.

Tian, F., Fensholt, R., Verbesselt, J., Grogan, K., Horion, S., & Wang, Y. (2015). Evaluating temporal consistency of long-term global NDVI datasets for trend analysis. *Remote Sensing of Environment, 163*, 326-340.

Tisseyre, B. (2012). Peut-on appliquer le concept d'agriculture de précision à la viticulture. *CNECA. Montpellier.*

Tisseyre, B., Leroux, C., Pichon, L., Geraudie, V., & Sari, T. (2018). How to define the optimal grid size to map high resolution spatial data? *Precision agriculture*, *19*(5), 957-971.

Trevino, V., & Falciani, F. (2006). GALGO: An R package for multivariate variable selection using genetic algorithms. *Bioinformatics, 22*(9), 1154–1156.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment, 8*(2), 127-150.

Valcarce-Diñeiro, R., Arias-Pérez, B., Lopez-Sanchez, J., & Sanchez, N. (2019). Multi-Temporal Dual- and Quad-Polarimetric Synthetic Aperture Radar Data for Crop-Type Mapping. *Remote Sensing, 11*, 1518.

Venios, X., Korkas, E., Nisiotou, A., & Banilas, G. (2020). Grapevine Responses to Heat Stress and Global Warming. *Plants, 9*(12), 1754.

Verbeeck, N., Caprioli, R. M., & Van de Plas, R. (2020). Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry. *Mass Spectrometry Reviews*, *39*(3), 245–291.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 108*, 273-290.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J. P., Camps-Valls, & G., Moreno, J. (2012). Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment, 118*, 127–139.

Wacheux, F. (1996). Qualitative methods and research management. *Economica, Paris, France.*

Wang, B., Jia, K., Liang, S., Xie, X., Wei, X., Zhao, X., Yao, Y., & Zhang, X. (2018). Assessment of Sentinel-2 MSI Spectral Band Reflectances for Estimating Fractional Vegetation Cover. *Remote Sensing*, *10*(12), 1927.

Wang, Q., & Atkinson, P. M. (2018). Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sensing of Environment, 204,* 31-42.

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment, 108*(3), 290-310.

Webb, L., Whiting, J., Watt, A., Hill, T., Wigg, F., Dunn, G., Needs, S., & Barlow, E.W.R. (2010). Managing Grapevines through Severe Heat: A Survey of Growers after the 2009 Summer Heatwave in South-eastern Australia. *Journal of Wine Research, 21*(2-3), 147–165.

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment, 236,* 111402.

Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression: A Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58(*2), 109–130.

Woodcock, C. E., Loveland, T. R., Herold, M., & Bauer, M. E. (2020). Transitioning from change detection to monitoring with remote sensing: A paradigm shift. *Remote Sensing of Environment, 238,* 111558.

Woolley, J. T. (1971). Reflectance and Transmittance of Light by Leaves. *Plant Physiology, 47*(5), 656-662.

Wu, F., Wang, C., Zhang, H., Zhang, B., & Tang, Y. (2011). Rice Crop Monitoring in South China with RADARSAT-2 Quad-Polarization SAR Data. *IEEE Geoscience and Remote Sensing Letters, 8*(2), 196-200.

Wu, M., Yang, C., Song, X., Hoffmann, W. C., Huang, W., Niu, Z., Wang, C., Li, W., & Yu, B. (2018). Monitoring cotton root rot by synthetic Sentinel-2 NDVI time series using improved spatial and temporal data fusion. *Scientific Reports, 8*(1), 2016.

Xu, X., Kang, J., Shen, J., Zhao, S., Wang, B., Zhang, X., & Chen, Z. (2021). EEM–PARAFAC characterisation of dissolved organic matter and its relationship with disinfection by-products formation potential in drinking water sources of northeastern China. *Science of The Total Environment, 774,* 145297.

Xue, J., & Su, B. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors, 2017.*

Yang, J., Xu, J., Zhang, X., Wu, C., Lin, T., & Ying, Y. (2019). Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Analytica Chimica Acta, 1081*, 6-17.

Yang, X., Yu, X., Cheng, J., Zheng, R., Wang, K., Dai, Y., Tong. N., & Chow, A.T. (2018). Impacts of land-use on surface waters at the watershed scale in southeastern China: Insight from fluorescence excitation-emission matrix and PARAFAC. *Science of The Total Environment, 627*, 654-657.

Yao, X., Wang, N., Liu, Y., Cheng, T., Tian, Y., Chen, Q., & Zhu, Y. (2017). Estimation of Wheat LAI at Middle to High Levels Using Unmanned Aerial Vehicle Narrowband Multispectral Imagery. *Remote Sensing, 9*(12), 1304.

Yengoh, G. T., Dent, D., Olsson, L., Tengberg, A. E., & Tucker, C. J. (2015). Limits to the use of NDVI in land degradation assessment. In *Use of the normalized difference vegetation index (NDVI) to assess land degradation at multiple scales* (pp. 27-30). Springer, Cham.

Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., & Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment, 237*, 111511.

Zhang, J. (2010). Multi-source remote sensing data fusion: Status and trends. *International Journal of Image and Data Fusion, 1*(1), 5-24.

Zhang, X. (2015). *Application of chemometrics to hyperspectral imaging analysis of environmental and agricultural samples*. (Doctoral dissertation, University of Barcelona).

Zhang, X., & Tauler, R. (2013). Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging. *Analytica Chimica Acta, 762*, 25–38.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment, 84(*3), 471-475.

Zhao, Q., Caiafa, C., Mandic, D., Zhang, L., Ball, T., Schulze-Bonhage, A., & Cichocki, A. (2011). Multilinear Subspace Regression: An Orthogonal Tensor Decomposition Approach. *Advances in Neural Information Processing Systems, 24*, 1269–1277.

Zheng, Y., Wu, B., Zhang, M., & Zeng, H. (2016). Crop Phenology Detection Using High Spatio-Temporal Resolution Data Fused from SPOT5 and MODIS Products. *Sensors, 16*(12), 2099.

Zhou, W., Newsam, S., Li, C., & Shao, Z. (2017). Learning Low Dimensional Convolutional Neural Networks for High-Resolution Remote Sensing Image Retrieval. *Remote Sensing, 9*, 489.

Zhou, X., Wang, P., Tansey, K., Zhang, S., Li, H., & Tian, H. (2020). Reconstruction of time series leaf area index for improving wheat yield estimates at field scales by fusion of Sentinel-2, -3 and MODIS imagery. *Computers and Electronics in Agriculture, 177*, 105692.

Zhou, Y., Pei, F., Xia, Y., Wu, C., Zhong, R., & Wang, K. (2019). Assessing the Impacts of Extreme Climate Events on Vegetation Activity in the North South Transect of Eastern China (NSTEC). *Water*, *11*(11), 2291.

Zhu, L., Su, X., & Tai, X. (2021). A High-Dimensional Indexing Model for Multi-Source Remote Sensing Big Data. *Remote Sensing, 13(*7), 1314.

Zhu, Z. (2017). Change detection using Landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, *130*, 370-384.

Zielinski, A. A. F., Haminiuk, C. W. I., Nunes, C. A., Schnitzler, E., van Ruth, S. M., & Granato, D. (2014). Chemical Composition, Sensory Properties, Provenance, and Bioactivity of Fruit Juices as Assessed by Chemometrics: A Critical Review and Guideline. *Comprehensive Reviews in Food Science and Food Safety*, *13*(3), 300-316.

Publications and international conferences

Lopez, E. F., Brunel, G., Devaux, N., Rancon, F., Pichon, L., & Tisseyre, B. (2021). Potential of temporal series of Sentinel-2 images to define zones of vine water restriction. In *Precision agriculture'21* (pp. 77-88). Wageningen Academic Publishers.

Laurent, C., Rançon, F., **Fornieles, E. L.**, Scholasch, T., Metay, A., Taylor, J., & Tisseyre, B. (2021). Is it relevant to account for grapevine phenology in time series of satellite images? In *Precision agriculture'21* (pp. 60-72). Wageningen Academic Publishers.

Lopez-Fornieles, E., Brunel, G., Rancon, F., Gaci, B., Metz, M., Devaux, N., ... & Roger, J. M. (2022). Potential of Multiway PLS (N-PLS) regression method to analyse time-series of multispectral images: A case study in agriculture. *Remote Sensing*, *14*(1), 216.

Lopez-Fornieles, E., Brunel, G., Devaux, N., Roger, J. M., & Tisseyre, B. (2022). Is It Possible to Assess Heatwave Impact on Grapevines at the Regional Level with Time Series of Satellite Images? *Agronomy*, *12*(3), 563.

Lopez-Fornieles, E., Tisseyre, B., Cheraiet, A., Gaci, B., & Roger, J. M. (2022). Potential of N-CovSel for Variable Selection: A Case Study on Time-Series of Multispectral Images. *Frontiers in Analytical Science*, 6.