Potential of functional analysis applied to Sentinel-2 time-series to assess relevant agronomic parameters at the within-field level in viticulture

Sergio Vélez[1], Florian Rançon[2], Enrique Barajas[1], Guilhem Brunel[2], José Antonio Rubio[1], Bruno Tisseyre[2]

[1]Instituto Tecnológico Agrario de Castilla y León (ITACyL), Unidad de Cultivos Leñosos y Hortícolas. Valladolid. Spain.

[2] ITAP, Univ. Montpellier, Institut Agro, INRAE, 34060 Montpellier, France

*corresponding author: velmarse@itacyl.es

Abstract

Sentinel-2 satellite imagery offers a wealth of spectral information combined with a weekly temporal resolution. It is seen as a promising tool to extract spatial information about vineyards and link them to agronomic parameters. Usually, only one or a few images are commonly employed at specific stages like veraison in viticulture. Extracting further information from time-series images may be of interest; however, this remains an issue due to the noisy and complex nature of extracted time-series. The functional analysis proposes a robust continuous representation of these time-series, which can then be used with adapted statistical tools. This paper focuses on extracting relevant information at the within-field level on two vineyards in Spain, which can be jointly interpreted with field observations and measurements. More precisely, it discusses the use of popular linear dimensionality reduction techniques, namely Principal Component Analysis (PCA) and Partial Least Square (PLS), adapted to functional data in order to decompose NDVI time-series into a weighted sum of several functional components. The unsupervised methods, like PCA, decomposed the spatial structure within the vineyards using a few components, resulting in a better and more manageable dataset than the one obtained using simple non-constrained methods. The results show significant correlations with ground-truth data showing the added value of considering the whole NDVI temporal series compared to a single NDVI map at veraison. The proposed approach provided helpful information about each component's yearly trend. Moreover, the results are linked to grapevines' seasonal phenology and management practices, highlighting phenomena affecting

the vineyard's development. This method is particularly suited for interactions with field experts, who may derive relevant agronomic information from the decomposition maps.

Keywords: PCA, Vineyard, Dimensionality reduction, functional analysis, clustering.

Introduction

Satellite imagery is widely used in vineyard studies because it is related to vegetation (Vélez *et al.*, 2020). It can be applied to the analysis of within-field variability of the water status of the soil and plants (Borgogno-Mondino *et al.*, 2018), harvest prediction (Sun *et al.*, 2017), the analysis of the spatial heterogeneity of evapotranspiration (Knipper *et al.*, 2019) or the classification of vine fields according to their vigor level (Di Gennaro *et al.*, 2019). Compared to satellites, UAVs are also an interesting source of information because they offer higher resolution and more acquisition flexibility; however, UAVs have several disadvantages, such as having a smaller swath. In any case, UAVs offer a different and complementary profile to satellites (Emilien *et al.*, 2021). In this respect, the Sentinel-2 satellites (European Space Agency's Copernicus project) are balanced in their spatial resolutions (10-60m), providing high temporal resolution data and can be particularly helpful due to the free nature of the images and the relative ease of access (https://scihub.copernicus.eu/). The consequence is that although Sentinel-2 data have lower spatial resolution than typical UAV imagery, it is known to be able to detect relevant spatial variability within the vineyard in the absence of inter-row grass and if individual vine management is not the goal (Sozzi *et al.*, 2020; Di Gennaro *et al.*, 2019). Moreover, according to ESA (2015), these satellites provide spectral information in multiple bands that allow computing many vegetation indices, which are of great interest for agricultural applications and, in particular, the six bands covering the red/red-edge spectral interval from 650nm to 900nm that allow computation of the Normalized Difference Vegetation Index (NDVI; Rouse *et al.*, 1973). It has been shown that NDVI images at the veraison stage were the most relevant to assess parameters related to the vine vigor and resulting harvest quality (Anastasiou *et al.*, 2018; Vélez *et al.*, 2019).

The availability of time-series makes it possible to consider all the images acquired during the vegetative period and not to be limited to an image at a single stage. Devaux *et al.* (2019) have shown, using Sentinel-2 images, that the whole temporal NDVI series over the growing season may bring relevant information on the field and its spatial variability (for instance, blocks with different management units and vegetative developments). However, it is difficult to analyze the potential relationship with relevant agronomic parameters considering all the available Sentinel-2 dates due to the temporal data's high dimension. Moreover, it is usually highly correlated and non-continuous (defined by satellite acquisition dates).

Functional analysis (Ramsay *et al.*, 2005) has been proposed to address these issues by summarizing the whole curve shape of non-continuous measured data. It aims at proposing a continuous and smooth representation of the data that small local curves or functions can model. While that representation may make the data theoretically infinitely dimensional, it can effectively consider different temporal dynamics at different scales while being relatively robust to noise (Febrero-Bande *et al.*, 2008). Functional representation can arise naturally in research domains such as spectroscopy, in which near-continuous data is already available. Recent advances have increased the use of functional data for outlier removal (Febrero-Bande *et al.*, 2008), regression (Goldsmith *et al.*, 2011), or classification (Leng *et al.*, 2005). Functional data can also be easily derived to describe slopes and locate inflection points. This paper will focus on its use to extract relevant information from NDVI time-series at the within-field level, which can be jointly interpreted with field observations and measurements. More precisely, it focuses on using popular linear dimensionality reduction techniques, Principal Component Analysis (PCA) and Partial Least Square (PLS), applied to functional data to decompose NDVI curves into a weighted sum of several components. That decomposition can then be used to create several complementary maps describing different spatial dynamics in the vineyard.

Based on a real case study in viticulture, the objectives of the paper are, therefore: i) to test the interest of considering the whole NDVI time-series to estimate agronomic parameters of interest at the within-field level, ii) to show how classical multivariate analysis methods (dimension reduction) based on functional analysis are relevant to account for inter-image noise management and agronomic variables estimation simultaneously, and iii) finally to discuss the applicability of the approach by considering a real application.

Materials and Methods

1. Vineyards

The experiment was carried out during the 2018 campaign, using two datasets taken from two commercial vineyards, one in 'Villanueva de Duero, Valladolid' (Vineyard A), planted in 2003, and the other one in 'Aranda de Duero, Burgos' (Vineyard B), both in Spain (Table 1). They belong to different Appellations of Origin (AOP), located 100 km away from each other. Both vineyards were grafted on 110 Richter rootstock and trained in vertical shoot positioning (VSP).

Figure 1 shows irrigation sectors: 2 in vineyard A and 16 in vineyard B; and reveals vineyard B presents some heterogeneities: the western part was managed under an organic regime and was planted in 2000 with a clone obtained from local (massal) selection, while the eastern part was planted in 1997 using CL770 clone and managed under conventional/standard management. The veraison phenological stage was observed between July 27 and August 7 for both vineyards, showing similar phenology dynamics. Vineyard A and B were harvested on September 27 and October 9, respectively, according to the average Total Soluble Solids (TSS).

Table 1 - Characteristics of each vineyard. Rainfall A: Annual rainfall in 2018. Rainfall B: Accumulated rainfall between April 1 and harvest. Rainfall measured at the 'VA06 - Tordesillas station, Valladolid' (Vineyard A) and 'BU05 - Vadocondes station, Burgos' (Vineyard B). Rainfall A: Annual rainfall in 2018. Rainfall B: Accumulated rainfall between April 1 and harvest.

Vineyard	**Cultivar**	**AOP**	**Coordinates (ETRS89 UTM30N)**	**Area (ha)**	**Rainfall A/B (L/m^2)**	**Orientation**
A	Verdejo	Rueda	X: 344711.1 Y: 4594058.8	8.5	515 / 201	E-W
B	Tempranillo	Ribera del Duero	X: 444063 Y: 4611977	43	620 / 276	N-S

Both vineyards were subject to mechanical weeding, keeping the soil free of weeds and any elements affecting spectral information (Fountas *et al.*, 2014). Management operations were very similar for vineyards A and B (Table 2). Regarding image time-series, the period between the end of June and early July corresponded to canopy management (summer pruning) operation, which may impact NDVI values (biomass decrease).

Table 2 - Management calendar in vineyards A and B.

		April				May				June				July				August				September			
	week	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
A	Chisel plowing				■			■					■				■				■				
A	Spraying							■				■		■											
A	Summer Pruning													■			■								
A	Sulfur																■		■						
B	Chisel plowing				■			■			■			■			■				■				
B	Spraying							■		■		■		■											
B	Summer Pruning													■			■								
B	Sulfur																■		■						

2. Ground-truth data

A grid of equally distributed sampling sites (SS) was established in each vineyard: 16 SS in vineyard A and 98 SS in vineyard B (Figure 1). For vineyard A, a SS corresponded to 10 vine plants (five in one row and the other five in the adjacent one), resulting in 160 sampled vines.

In vineyard B, a SS corresponded to 3 vines (along the row), resulting in 294 sampled vines. Each SS included fewer plants in vineyard B since it is a larger vineyard, and the priority of sampling was to cover the whole area during a single sampling campaign.

Each SS was georeferenced with a Triumph-2 (Javad, USA) RTK receiver connected via Bluetooth to an Android smartphone to access local RTK correction signals.

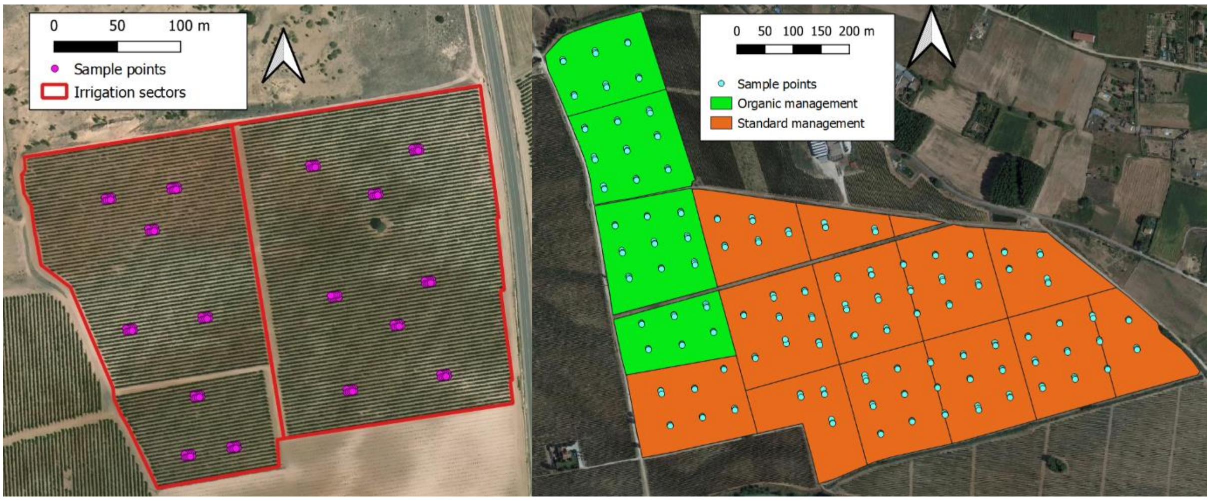


Figure 1 - Sample sites and irrigation sectors of (left) vineyard A and (right) vineyard B.

The vineyards belong to a more significant project in which several parameters related to agronomic, abiotic, orographic, and harvest quality variables were monitored. In order to ease the comparison, the variables were interpolated using kriging (equation 1) after the semi-variogram was fitted to the observed semi-variance (Figure 2):

$$\hat{z}\,(x_0) = \sum_{i=1}^{N} \lambda_i z(x_i) \quad \text{Eq. 1}$$

where $\hat{z}$ is the estimated value at point x_0, λ_i are the weights, and z are the known values at points x_i. The measurement of these parameters was carried out following standard protocols in viticulture which will not be described in this paper. Instead, the reader can refer to Hidalgo (2006), Reynolds (2010), and Jackson (2020) for a detailed description. Note that all measurements performed at the plant level were assigned to the geographic coordinate of the Sampling Site. Regarding agronomic variables, five parameters were evaluated: pruning wood weight (PWW), yield at harvest (Yield), and average cluster weight (CW). Moreover, the grapes' typical quality parameters were measured: Brix, Sugar, pH, total acidity, malic acid, tartaric acid, and potassium. Additionally, several abiotic and orographic parameters were measured at each SS, such as soil texture (sand, clay, and silt), soil electrical apparent conductivity (σ), pH, elevation, organic matter (OM), and Cation Exchange Capacity (CEC). In order to simplify the analysis, only a few variables were selected: Pruning wood weight (PWW), Yield at harvest, Cluster weight (CW), Tartaric Acid (TA), soil fraction in Clay, Sand and Silt, soil Organic Matter (OM), soil Cation Exchange Capacity (CEC), and Soil electrical apparent conductivity (σ), aiming to include the most relevant agronomic and abiotic parameters for vineyard managers in viticulture (Keller, 2015).

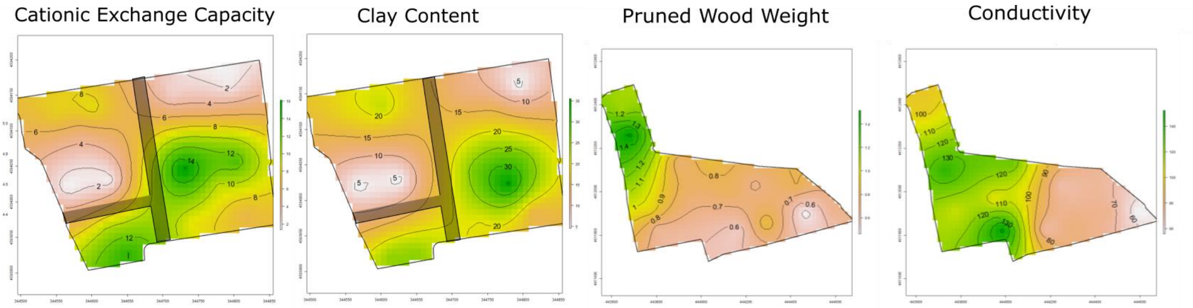


Figure 2 – Example of kriged maps on vineyard A (first two figures) and vineyard B (last two figures).

An additional specific variable was considered for vineyard B, which presented specific within-field spatial patterns (Figure 3), probably explained by a higher organic matter level (Vodyanitskii and Savichev, 2017) or ephemeral streams, causing sediment transport or even erosion (White, 2015). The ephemeral streams hypothesis is more likely since it fits with the vineyard's elevation map, which is lower on the southwest side, causing streams to flow towards the river.

This peculiarity was present only in vineyard B and was considered a driver that may explain plant vigor differences due to changes in soil properties and water content (Yang *et al.*, 2019).

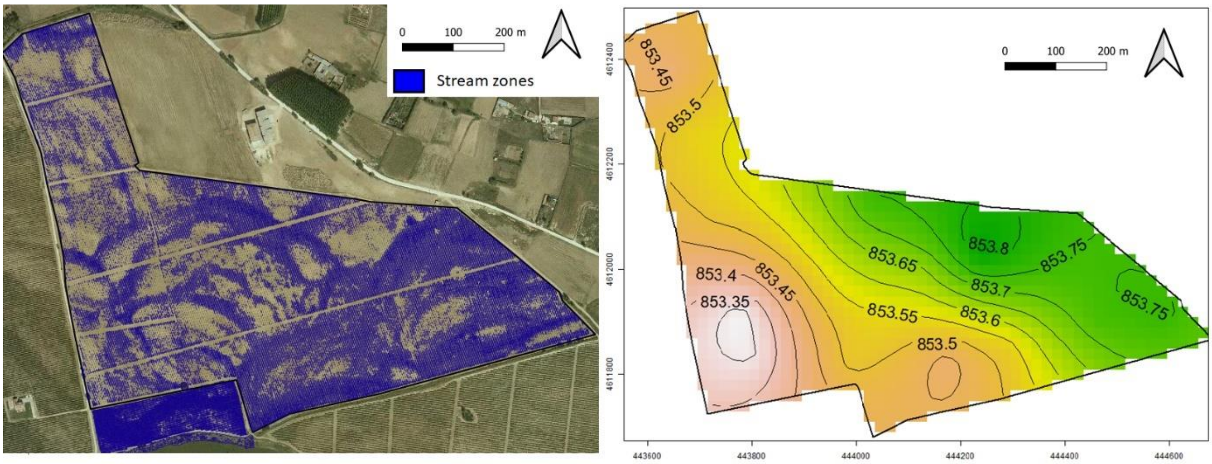


Figure 3 – Soil specificities of Vineyard B, map of "ephemeral streams and possible higher concentration in organic matter" corresponding to darker patterns highlighted in dark blue from a visible image in winter (left) and elevation map (right)

3. Satellite data acquisition and processing

Sentinel 2 Level-2A imagery was downloaded, and cloudy dates were manually removed. Images were used to build time-series, the vineyard boundaries were cropped, and the NDVI was computed for each pixel over time. B8 and B4 Sentinel-2 bands were used to compute the NDVI following equation 2. Obvious outliers were removed from the analysis; they included paths in vineyard A. No pixels were removed in the vineyard B.

$$NDVI = \frac{(NIR - red)}{(NIR + red)} = \frac{(B8 - B4)}{(B8 + B4)} \quad \text{Eq. 2}$$

For the analysis, the 2018 NDVI time-series were limited from April to September. Figure 4 shows raw NDVI curves on each vineyard for each pixel.

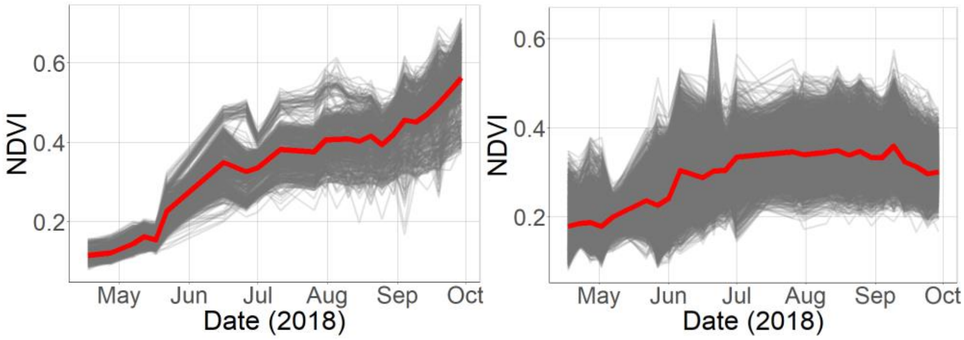


Figure 4 - NDVI evolution during the year 2018 for pixels in vineyard A (left) and vineyard B (right). The red curve indicates the mean NDVI evolution in each vineyard.

4. Functional representation and dimensionality reduction

The raw discrete time series were then transformed using functional analysis (Jolliffe, 2002). Given the original discrete data X (each point is an NDVI value derived from Sentinel-2 images), a functional representation Y of the data aims to create a smoothed continuous representation of that data, meaning the functional data does not exactly fit the original discrete data points. Residuals can be computed in a similar way to classical regression problems (Equation 3)

$$Y_i(t) = X_i(t) + \varepsilon_i(t) \quad \text{Eq. 3}$$

where $\varepsilon_i(t)$ are the residuals not explained by the functional representation, i is a sample (time-series for a given pixel), and t is the time. Residuals may be related to noise or punctual phenomena not accounted by the functional data.

A non-parametric approach (Ferraty *et al.*, 2006) was chosen to obtain the data's functional representation. It uses a Gaussian smoothing kernel and a bandwidth parameter w (Equation 4).

$$M_{ab} = \frac{1}{w} K\left(\frac{t_a - t_b}{w}\right) \quad \text{Eq. 4}$$

The matrix M, whose elements are the weights between points (a and b are the matrix rows and columns) in the time series, can then be employed to compute the smoothed data (Equation 5).

$$Y = M.X \quad \text{Eq. 5}$$

This approach was preferred over the parametric approach, as proposed by Ramsay *et al.*, (2005) because it is easier to settle and interpret as it involves only a kernel choice with a bandwidth parameter. The higher the bandwidth, the higher the smoothing and the higher residuals are, with a possible loss of information. Figure 5 exemplifies the effect of three different kernels, corresponding to three different bandwidths, on the NDVI time-series of vineyard A. In a first approach, a W = 5 was chosen to mitigate the smoothing vs. a potential loss of information.

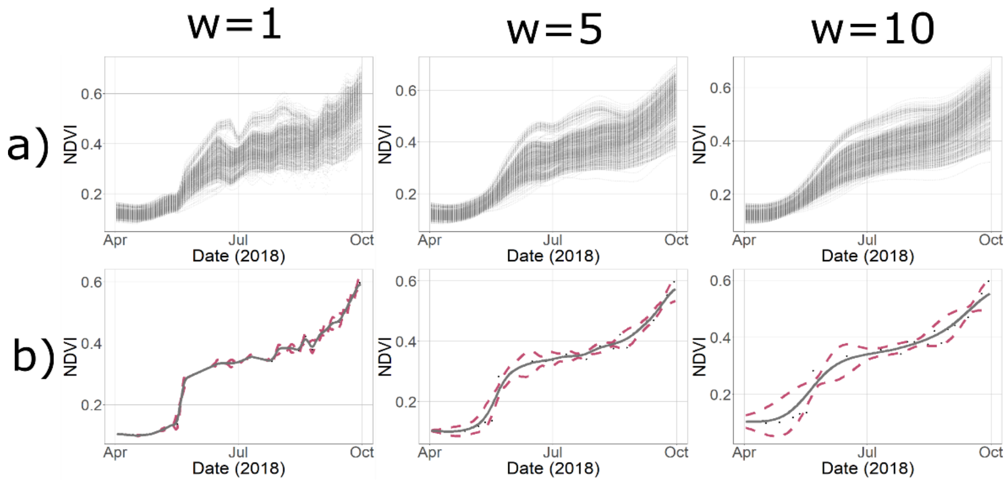


Figure 5 - Effect of the bandwidth w on non-parametric functional analysis representation of the vineyard A data. (a) NDVI curves for all the pixels (grey lines) within the vineyard (b) Example of NDVI curve for one pixel (blue line) with original Sentinel-2 measurements (black dots) and confidence envelope (red dotted line) overlaid.

The next step involves reducing the dimensionality of the time series, going from high dimensions to a handful of informative components related to spatial phenomena in the vineyard. Dimensionality reduction aims at doing this in an unsupervised (no training samples available) or a supervised way (training samples, e.g., yield measurements available).

Several methods exist for seasonality decomposition and summarizing temporal information in time series, such as ARIMA (Box *et al.*, 2015) or exponential smoothing models (ETS, Hyndman *et al.*, 2008). In this paper, a Functional Principal Component Analysis (f-PCA, Cardot *et al.*, 1999), which is the direct adaptation of PCA to functional data, was chosen. The main goal of f-PCA is bringing classical component decomposition (PCA) to semi-continuous time series. It was preferred over other methods for several reasons: i) the study was conducted on a single year, meaning no cyclical patterns need to be extracted, only a main trend during the summer season ii) scores and components generated by the PCA decomposition have a natural interpretation and can be used to produced visual maps iii) the functional paradigm combined with non-parametric dimensionality reduction allows us to handle Sentinel-2 related noise efficiently. With the f-PCA, each sample (NDVI evolution throughout the year for a pixel) can be expressed as a set of scores for each principal component following Equation 6.

$$C_i(t) = \sum_{k=1}^{K} S_{ik}\xi_k \text{ Eq. 6}$$

where S_{ik} is the score of sample i for the principal component k and $\xi_k$is the principal component function which can be itself expressed as a set of functional basis. The choice of the number of components depends on the explained variance (like for classical PCA).

Functional Partial Least Square (f-PLS) (Preda *et al.*, 2007) was chosen as a complementary supervised method for the study to take advantage of the availability of a scalar variable (such as field measurements). f-PLS was performed to consider a subspace maximizing the covariance between the exploratory variables (functional data) and available ground truth data. In other words, f-PLS projects the data in a way that classical regression can be easily performed. The need for a scalar explanation variable means not every pixel can be used to train the model, which can be a problem for vineyard A with few measurement sites.

Figure 6 presents a decomposition example of a set of NDVI curves into 3 Principal Components (PC) curves (interpretation would be similar for f-PLS Latent Variables). The individual curves for each sample can themselves be decomposed as the PC weighted by the sample score, indicating how a trend modeled by a component is modulated for each pixel in the vineyard. Time-series can then be summarized by a set of scores (equal to the number of principal components) that best describe the whole original time series. Since the samples are pixels, it is possible to create a score map for each PC. In Figure 6, the bottom row is a representation of the reconstructed curves using the principal components and the scores. For example, the PC1+2+3 reconstruction is a sum of the first three components weighted by the pixel scores. These reconstructions can be directly compared with the original signal in order to obtain residuals time-series. Unlike the global scores, which summarize the whole time series, residuals can be mapped at any date, being helpful to check specific dates that are poorly modeled by the decomposition. Further results were analyzed using the component curves, the global score maps, and the residuals maps.

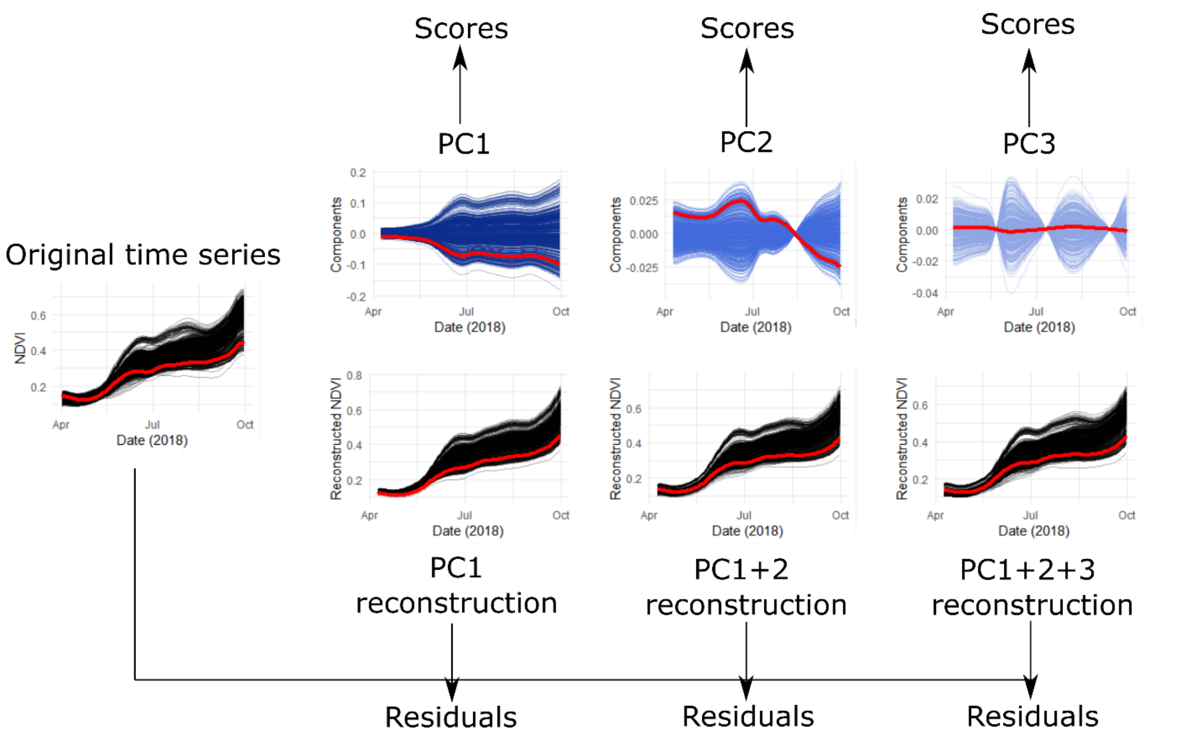


Figure 6 - Visual decomposition of NDVI curves as a weighted sum of the 3 functional principal components (columns). Top row represents the components modulated by the scores. Bottom row represents the time-series reconstruction using different combinations of components. The red curves detail the decomposition of an example time-series (pixel within the vineyard).

5. Validation

Spearman correlation analysis was used to check whether the components obtained from the dimensionality reduction are linked to measured variables in the vineyard. Since measurements are punctual, only the data at these Sample Sites was considered. Additionally, a Mann-Whitney test was used to test for significance on categorical variables in vineyard B (stream zones/non-stream zones and organic/non-organic management zones). This statistic considers the number of samples used and thus may be adapted to low sample sizes.

Subsequently, functional representation and associated dimensionality reduction algorithms were explored on vineyards A and B in order to assess if they can explain known spatial variabilities within the vineyard (due to various factors such as soil, water status, or management practices) better than conventional veraison maps. The study first focused on applying f-PCA and later on a comparison of the results using f-PCA and f-PLS. Field data presented beforehand were used to validate these observations.

All image, processing, and data analyses were performed using QGIS (version 3.14.X, QGIS developer team 2020) and customized codes written in R statistical program (version 4.0.3, R Foundation for Statistical Computing (R Core Team 2019), https://www.R-project.org/, Vienna, Austria) using 'fda.usc' package as the basis for functional analysis representation (Febrero-Bande *et al.*, 2012), among others such as 'raster', 'sf', 'sp', 'rgdal' and 'rgeos'.

Results

4.1. Previous exploratory analysis

A first exploratory data analysis on yield and Pruning Wood Weight (PWW) confirms that vineyard A has much less variability than vineyard B; it also has fewer measurement sites since it is a smaller vineyard. Figure 7 highlights the differences in the magnitude of variation in yield and PWW for both vineyards. Upon conducting simple regression, it appears PWW has the best correlation with the Sentinel-2 NDVI at veraison, confirming many literature studies on the ability of NDVI at veraison to estimate vegetative plant biomass and plant vigor.

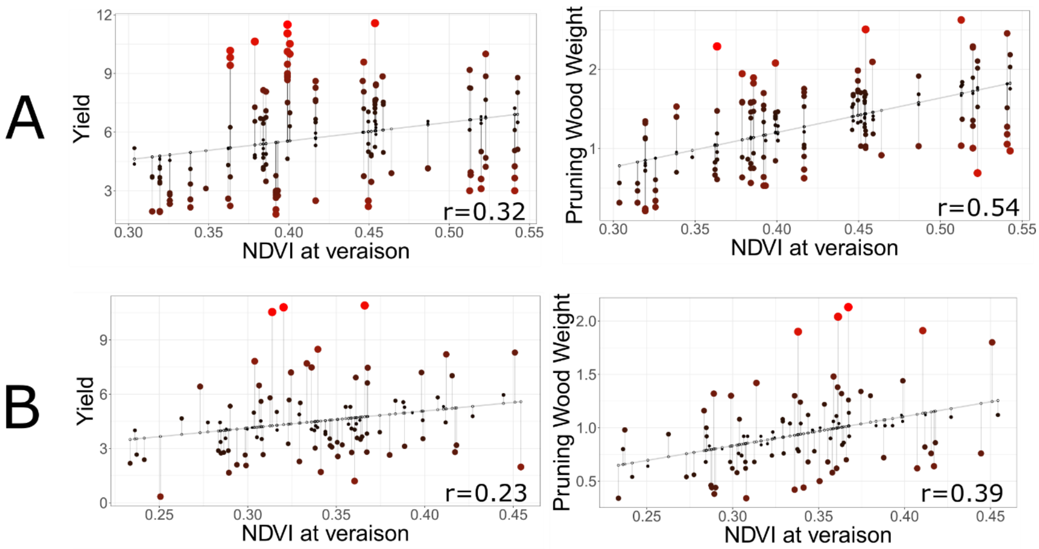
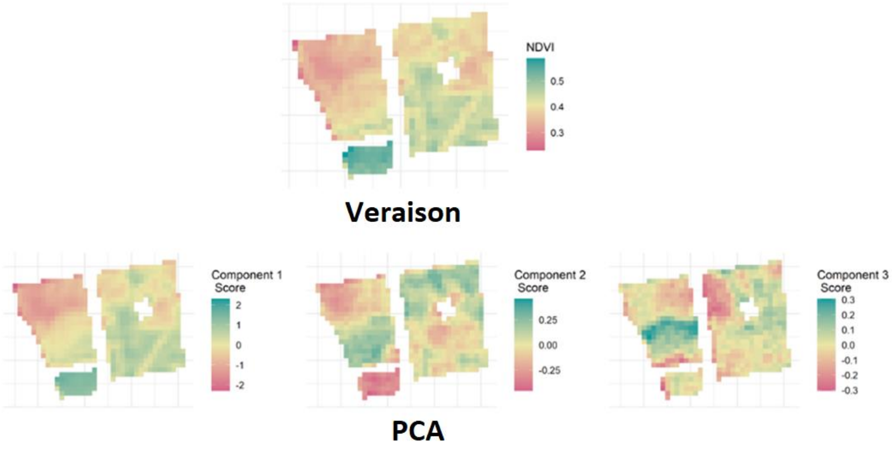


Figure 7 - Classical maximum likelihood regression between NDVI at veraison (July 31 Sentinel-2 image was used) and yield/Pruned weight on vineyards A and B. Vertical lines, dot color, and size indicate fit residuals. Associated r values (Spearman correlation) are indicated in the bottom right corner of each plot.

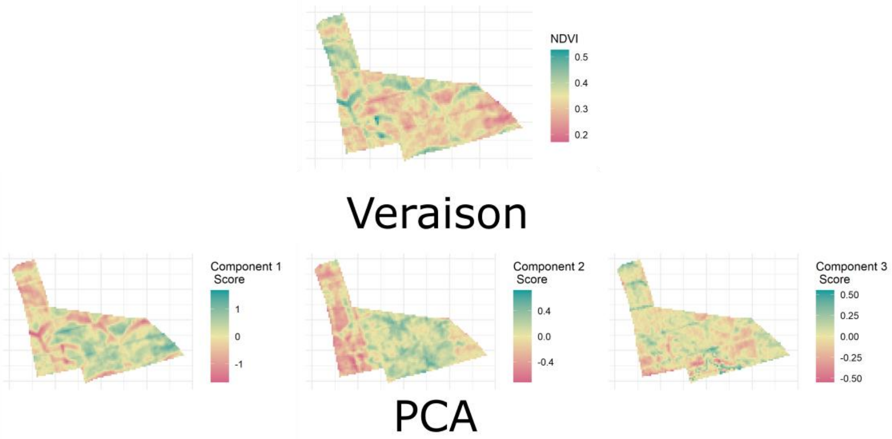
However, these studies only focus on NDVI information at one date and do not use the available temporal information. The functional analysis combined with dimensionality reduction aims to use this information to perform a general decomposition of seasonal time-series and could potentially be a relevant link to field values.

4.2. Spatial structure of f-PCA components

Only the first three principal components, accounting for more than 95% of the variance on both vineyards, are considered in the study. Figure 8 presents the score maps for each component. These scores, computed on the whole time-series, are compared to NDVI maps at veraison for both vineyards A and B. While NDVI and component scores do not share the same scale, these values' global spatial organization within the vineyard can be compared.



(A)



(B)

Figure 8 - Comparison between classical veraison NDVI map and the first three components of PCA in both vineyards. (A) Vineyard A and (B) vineyard B.

Results show that the first principal component map tends to be very similar to the NDVI map obtained at veraison; both maps highlight the same spatial patterns. According to the correlations between the first component and field values shown in table 3, the first component is linked to the final pruning wood weight. Moreover, the correlation is higher between the PWW and the PC1 (r = -0.62 and -0.51, respectively for vineyards A and B) than just the PWW and the veraison map (r = 0.54 and 0.39, respectively for vineyard A and B. Figure 7).

The first principal component is also linked to other quantitative parameters such as the yield, cluster weight, and tartaric acid in vineyard A, but only with tartaric acid in vineyard B. Additionally, strong correlations were found significant with soil variables such as sand and silt content, Cation Exchange Capacity, organic matter content, and soil apparent electrical conductivity. Detailed agronomic analysis of the relationships between the variables is beyond the scope of this paper; however, the results highlighted here are consistent with other literature results, showing that vine vigor is often related to soil characteristics. These characteristics directly determine other parameters such as access to water or soil fertility and indirectly the components of yield and grape quality, i.e., the significant correlation with tartaric acid may be related to the total vegetation (Hidalgo, 2006). It seems that the first principal component of the f-PCA integrates the changes in NDVI over the entire vegetative period with sites that systematically have high NDVI values and sites that systematically have low NDVI values. The NDVI image at veraison after plant vines stopped growing is somewhat the final result of the plant growth, explaining the remarkable similarity between the first component and the map at veraison. This result confirms the relevance of the f-PCA. However, the other components must be analyzed in order to investigate their additional potential information.

*Table 3. Vineyard correlations (Spearman). Significance level: * p-value<0.05; ** p-value<0.01.*

	PWW (kg)		Yield (kg)		CW (g)		TA (g/l)		Clay (%)	
Vineyard	A	B	A	B	A	B	A	B	A	B
PC1	-0.62**	-0.51**	-0.36**	-0.16	-0.48**	-0.19	-0.74**	-0.36**	-0.46	-0.24
PC2	-0.09	-0.08	0.15	0.39**	0.12	0.53**	-0.44	0.05	-0.75**	0.06
PC3	0.12	-0.14	0.35**	-0.15	0.36**	-0.02	0.22	-0.02	-0.05	0.12

	Sand (%)		Silt (%)		OM (%)		CEC (meqNa/100g)		σ (uS/cm)	
Vineyard	A	B	A	B	A	B	A	B	A	B
PC1	0.68**	0.54*	-0.63**	-0.54*	-0.73**	-0.48	-0.66**	-0.47	-0.48	-0.54*
PC2	0.69**	0	0.14	-0.01	-0.36	-0.15	-0.81**	-0.22	-0.7**	-0.41
PC3	-0.15	0.06	0.38	-0.09	0.28	-0.12	0.13	-0.17	-0.24	0.02

The correlations between the second component of each vineyard and the vineyard variables differ. In vineyard A, it is consistent with soil variables such as clay and sand content, soil apparent conductivity, and Cation Exchange Capacity, showing clear zones that deviate from the main vigor (growth) trend highlighted by the first component and related to soil variations. However, in vineyard B, the second component is more correlated with the plant variables than the veraison map showing that the correlation between the yield and the PC2 is higher (r = 0.39) than the correlation between the yield and the veraison map (r = 0.23, Figure 7). Furthermore, the second component of vineyard B divides the vineyard into two sections: the western part with the lowest scores and the eastern part with the highest scores. These two sections were not visible either in the veraison map or the first component map. They fit with different clones and different management practices (Figure 1).

The interpretation of the third component should be handled with care since it explains less than 5% of the total variance. It may indicate some remaining border effects and subtle heterogeneity within the zones that were not explained by the first and the second component. It is not strongly related to any soil components but to cluster weight and yield variables. One hypothesis would be that the third component accounts for residual variation in vigor explained by the different source-sink ratios between yield and canopy development (Urretavizcaya *et al.*, 2017).

The three components exhibit clear spatial patterns, with different zones cohesive with field observations. Unsurprisingly, additional components led to more noisy maps, supporting the decision only to use a handful of components.

4.3. Assessment of principal components functions

Plotting the functions defining the principal components can be viewed as a complementary analysis (Figure 9). The results are based on the example of vineyard A, knowing that Vineyard B presents the same characteristics. Although this visualization of principal component functions loses relevant spatial information (pixel scores within vineyards), it provides a general temporal description of the phenomena modeled.

The first gained insight is that June and the resulting spatial heterogeneity cannot be modeled with a single component since each component exhibits different variation patterns. The first component models the vineyard's primary trend (also valid for vineyard B) over the whole vegetative growth, confirming the assumption of integrating the vigor's changes over the entire vegetative period. As a result, the first component models the general trend of the vineyard (Figure 9 left), and the scores, as explained in figure 6 and shown in figure 9 right, represent the deviation of each within-field site from this general trend. Figure 9 right also confirms the similarity between NDVI values at veraison and scores of the first component. In terms of temporal variation, it should be highlighted that the first component experiences a slight depression at the beginning of July, which modifies the upward trend and matches with the summer pruning operation.

The adjustments applied by the other components have different amplitudes depending on the date. The second component firmly adjusts the representation at the beginning of June, while the third component adjusts the representation in June and August (firstly in a positive and secondly in a negative trend). The importance of June may be linked to the complex dynamics within the vineyard at this time. Indeed, the vine growth is at its maximum and starts to decrease before entering the plateau zone. Some factors may have a significant impact on NDVI values during this period. The correlation with soil parameters indicates that the specific soil conditions (such as water availability or soil fertility) may affect NDVI changes (and therefore vine growth) and may explain precocity or delay in growth stop. Finally, note that all components exhibit a decrease in absolute value at the beginning of July. This result may be related to the previously mentioned summer pruning period, which homogenizes the size of the canopy and the resulting observed NDVI over the whole field. The f-PCA shows here that this period has a low informative value due to canopy management operations.

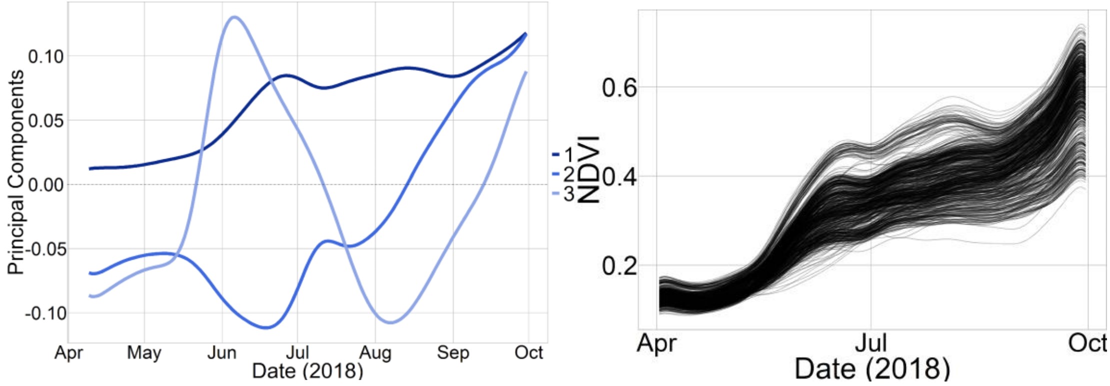


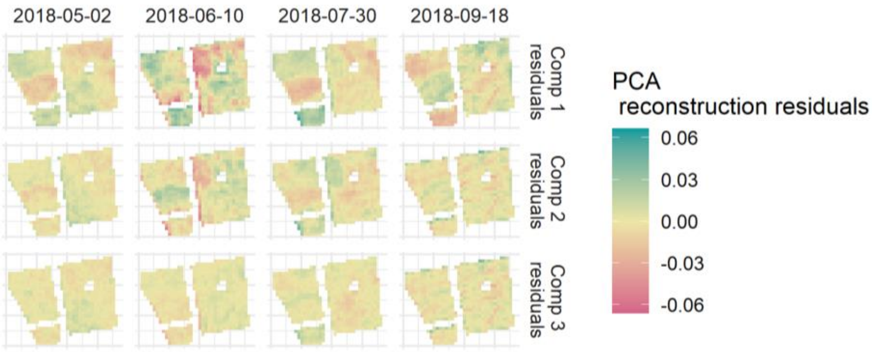
Figure 9 - (left) "Centered principal functions" obtained for the first three components of PCA on vineyard A. (right) NDVI functional data

4.4. Mapping of unexplained variance

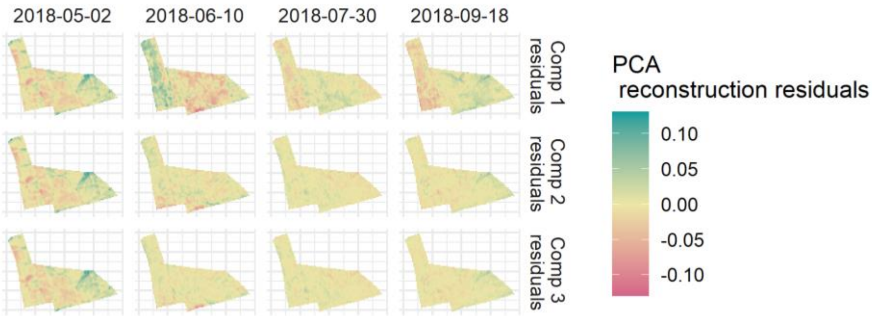
Figure 10 presents maps of residuals at four critical dates during the season for both vineyards. It was computed by considering the difference between the reconstructed and the original NDVI values at a given date. Since residuals can be computed at any date, the main advantage is getting a more precise temporal interpretation of the f-PCA modeling ability over time. At first glance, residual maps tend to be similar to score maps for some dates, which is not surprising because zones with low scores tend to be underestimated while high scores tend to be overestimated by the model. Since the goal is not to perfectly fit the data but to understand the underlying phenomena behind the NDVI, underestimation and overestimation can be seen as complementary information.

Compared to the measurements mapped using kriging (Figure 2), the first component's residuals are consistent along the season with CEC and soil variables, showing depression in the middle-west and east of the vineyard. The first component residuals are also consistent with other variables such as soil apparent conductivity and soil organic matter.

In vineyard B, the first component's residuals are clearly consistent with vineyard management and clones and, perhaps, with ephemeral water streams, highlighting differences at this period of high vegetative growth and showing differences between the west-east parts of the vineyard.



(A)



(B)

Figure 10 - Local residuals on both vineyards obtained at different dates during the reconstruction of the NDVI using combinations of PCA components: top: PCA 1, middle: PCA 1 + 2, bottom: PCA 1 + 2 + 3.

Regarding the temporal variability, the June image shows high heterogeneity in the first component's residuals; however, this heterogeneity is lower in the July image. This peculiarity was already observed in figure 9 and may explain the substantial adjustment of component 2 in June. As hypothesized previously, it can be explained due to the top pruning performed at the beginning of July, balancing the vegetation cover. Moreover, June being a high vegetative growth period, limiting factors related to soil and its variability may explain observed results.

4.5. Comparison between PCA and PLS results

The PLS analysis using Pruning Wood Weight as training samples shows that the first component is similar to the one obtained with PCA (results not shown). In vineyard A, the correlation between PCA and PLS first components was $R^2 = 0.99$. However, several differences were observed for the PLS second component, finding that it was highly similar to the first one ($R^2 = 0.85$). This high correlation can be explained by the fact that PLS tries to map the data in a subspace in which latent variables are correlated with explanatory variables, in that case Pruning Wood Weight. Here, a single component is enough to achieve that goal, just like the veraison map was directly linked to Pruning Wood Weight measurements, and any other information is redundant

Discussion

The f-PCA/f-PLS methods for dimensionality reduction were applied to two vineyard time series to highlight different agronomic spatial phenomena. The f-PCA explained different occurring phenomena such as the stream zones related to supposed ephemeral water streams, soil differences, or differential management of the vineyard.

In both vineyards, the first component was quite similar to the information obtained from a classical veraison NDVI map. There is a strong chance that this result will be observed on the majority of vineyards since this first component reflects the primary vigor trend of the vineyard over the whole season. As a result, regarding the first component, the f-PCA does not provide newer information than a NDVI map at veraison (the latter being the final result of growth at the end of the season). However, in both vineyards, the relationship between the pruning wood weight and the f-PCA first component values was higher than with the veraison NDVI maps, suggesting that the first component map computed from the f-PCA methodology accounts better for vine vigor than a single veraison map. One hypothesis could be that vegetation development is the most important source of NDVI variability over the year, and the first component focuses on the main variations within the year, leaving out other and more secondary sources of variation to the other components. By considering the whole time series, f-PCA guarantees that the mapped phenomenon results from the plant vigor dynamic over the whole season.

Additional information could be derived from the other components of the f-PCA. Indeed, temporal changes of NDVI cannot be modeled using a single component since each component exhibits variation patterns, especially in high heterogeneity periods such as the vegetation development period. Other components can be considered adjustments of the general trend; however, the information provided by the second and the third components requires agronomic expertise and a particularly good knowledge of field management. It was expected that the first component underfits the original time series, but it is not wished that subsequent components overfit it. Since this phenomenon is difficult to check, keeping only the first three components or even sometimes only the first two is crucial to avoid overfitting data and to create misleading interpretations. The explained variance could be used as an objective metric (e.g., stopping when 95% of the original variance is explained), but visually checking the components and trying to link them with field knowledge should be advised. In addition, each component exhibits variations in different periods of the year. The most noticeable variations were observed in June, related to the complex dynamics within the vineyard.

Finally, it is also interesting to note that, while the f-PCA method does not consider spatial autocorrelation (dependencies between neighboring Sentinel-2 pixels), the obtained components naturally show a strong spatial structure, meaning spatial information is not needed in the process.

The key findings of this paper can be seen in two ways. The first one is the application of the method: to our knowledge, there is no study using Sentinel-2 temporal data to decompose the vineyard behaviors into several key components. Historically, MODIS satellite images are used to describe seasonal trends on field crops, which were then used for crop classification or forecasting (Carreño-Conde *et al.*, 2021). The coarse resolution of the images made within-field applications in the vineyards impractical, thus they focused only on the temporal variation within several years. The recent Sentinel-2 launches blurred the limits between high spatial resolution imagery and high temporal resolution imagery. Methods like the presented f-PCA can be seen as an appropriate solution to tackle these challenges in a novel way, even though the method is not new. The second one is the knowledge of within-field variability and the increasing importance of satellite imagery in the process (Precision Viticulture). The results support the common observations about within-vineyard variability (Tisseyre *et al.*, 2008) and Sentinel-2 observed variability (Devaux *et al.*, 2019). Furthermore, it supports the observation that simplified indicators like NDVI at a single date may struggle to describe several sources of spatial and temporal variability in the vineyard (Hall *et al.*, 2010), thus supporting the need for decomposition algorithms applied to the complete set of Sentinel-2 bands during the season.

Conclusions

Functional analysis, combined with dimensionality reduction methods (f-PCA, f-PLS), is a promising tool to extract information from NDVI Sentinel-2 time-series of the vine field describing vine growth and vigor. It builds upon the free and open data policy adopted for Sentinel-2 products combined with the rich temporal information of the images. In this work, several key parameters that may support vineyard managers in understanding field variability were linked to a principal component decomposition of the temporal data. The components of the f-PCA were able to explain the vineyard spatial variability better than a single NDVI veraison map since the principal component values had a higher correlation with within-field parameters such as pruning wood weight or yield.

Moreover, this paper stressed the need to analyze each component separately, which adds complementary and valuable information, identifying new phenomena that cannot be explained in a veraison map. One key advantage of the method is that only a few components are needed to explain most of the vineyard variability contained in Sentinel-2 time series.

The potential for a more granular examination of the f-PCA decomposition at individual dates was also explored. For instance, noticeable component variations were linked to the vineyard's complex dynamics in the vegetation growth period associated with management practices. To our knowledge, this type of approach is quite new in precision agriculture. The methodology proposed here remains a first step to identify the potential of image time series at the within-field scale. It is likely that other methods will be proposed in the near future. It will then be necessary to be able to compare their relevance by proposing, for example, shared reference data to conduct this type of comparison.

Further studies may also include the joint use of all Sentinel-2 bands and applying the described methods to larger scales, ranging from groups of vineyards to a whole region. In particular, the spatial structures and patterns emerging from that type of data may be of great interest to better understand regional-scale dynamics.

Acknowledgments

This work has been possible thanks to the economic support of Junta de Castilla y León (Spain), Instituto Tecnológico Agrario de Castilla y León (ITACyL), the project INIA RTA2014-00077-C02, FPI-INIA2016-017, FEDER funds and the cooperation of ‘Bodega Martín Berdugo’ and ‘Bodega Cuatro Rayas”.

References

Anastasiou, E., Balafoutis, A., Darra, N., Psiroukis, V., Biniari, A., Xanthopoulos, G. & Fountas, S. (2018). Satellite and Proximal Sensing to Estimate the Yield and Quality of Table Grapes. *Agriculture*, 8, 94, doi:10.3390/agriculture8070094.

Borgogno-Mondino, E., Novello, V., Lessio, A., Tarricone, L. & de Palma, L. (2018). Intra-vineyard variability description through satellite-derived spectral indices as related to soil and vine water status. *Acta Horticulturae*, 1197, 59–68, doi:10.17660/ActaHortic.2018.1197.8.

Box G.E.P., Jenkins G.M., Reinsel G.C. & Ljung G.M. (2015). *Time Series Analysis: Forecasting and Control*, 5th Edition. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1

Cardot H., Ferraty F. & Sarda P. (1999). Functional Linear Model. *Statistics & Probability Letters*, 45(1), 11–22.

Carreño-Conde, F., Sipols, A. E., de Blas, C. S., & Mostaza-Colado, D. (2021). A Forecast Model Applied to Monitor Crops Dynamics Using Vegetation Indices (NDVI). *Applied Sciences*, 11(4), 1859. https://doi.org/10.3390/app11041859

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. https://doi.org/10.20870/oeno-one.2019.53.1.2293

Di Gennaro, S.F., Dainelli, R., Palliotti, A., Toscano, P. & Matese, A. (2019). Sentinel-2 Validation for Spatial Variability Assessment in Overhead Trellis System Viticulture Versus UAV and Agronomic Data. *Remote Sensing*, 11, 2573. https://doi.org/10.3390/rs11212573

Emilien, A.V., Thomas, C., & Thomas, H. (2021). UAV & satellite synergies for optical remote sensing applications: A literature review. *Science of Remote Sensing*, 3, 100019. https://doi.org/10.1016/j.srs.2021.100019

European Space Agency (ESA) (2015). SENTINEL-2 User Handbook, *ESA Standard Document.* Available Online: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook (accessed on March 30, 2021).

Febrero-Bande M., Galeano P. & González-Manteiga W. (2008). Outlier Detection in Functional Data by Depth Measures, with Application to Identify Abnormal NOx Levels. *Environmetrics*, 19(4), 331–345. https://doi.org/10.1002/env.878

Febrero-Bande M. & Oviedo de la Fuente M. (2012). Statistical Computing in Functional Data Analysis: The R Package fda.usc. *Journal of Statistical Software,* 51(4), 1–28. https://doi.org/10.18637/jss.v051.i04

Ferraty, F. & Vieu, P. (2006). *Non-parametric functional data analysis*. Springer Series in Statistics, New York.

Fountas, S., Anastasiou, E., Balafoutis, A., Koundouras, S., Theoharis, S. & Theodorou, N. (2014). The influence of vine variety and vineyard management on the effectiveness of canopy sensors to predict winegrape yield and quality. *In Proceedings of the International Conference of Agricultural Engineering*, Zurich, Switzerland, 6–10 July 2014.

Goldsmith J., Bobb J., Crainiceanu C., Caffo B. & Reich D (2011). *Penalized Functional Regression. Journal of Computational and Graphical Statistics*. https://doi.org/10.1198/jcgs.2010.10007

Hall, A., Lamb, D. W., Holzapfel, B. P., & Louis, J. P. (2010). Within-season temporal variation in correlations between vineyard canopy and winegrape composition and yield. *Precision Agriculture*, 12(1), 103–117. https://doi.org/10.1007/s11119-010-9159-4

Hidalgo, J. (2006), *La calidad del vino desde el viñedo*, Mundi-Prensa, Spain.

Hyndman, R., Koehler, A., Ord, K., & Snyder, R. (2008). Forecasting with Exponential Smoothing. *Springer Series in Statistics*. https://doi.org/10.1007/978-3-540-71918-2

Jackson, R. (2020). *Wine Science: Principles and Applications*, 5th ed., Elsevier: Cambridge.

Jolliffe I.T. (2002). *Principal component analysis*, 2nd edn New York, NY: Springer-Verlag.

Keller, M. (2015). *The Science of Grapevines: Anatomy and Physiology*. Second edition. Academic Press, Elsevier, Amsterdam.

Knipper, K.R., Kustas, W.P., Anderson, M.C., Alsina, M.M., Hain, C.R., Alfieri, J.G., Prueger, J.H., Gao, F., McKee, L.G. & Sanchez, L.A. (2019). Using High-Spatiotemporal Thermal Satellite ET Retrievals for Operational Water Use and Stress Monitoring in a California Vineyard. *Remote Sensing*, 11, 2124. https://doi.org/10.3390/rs11182124

Leng, X., & Müller, H. G. (2005). Classification using functional data analysis for temporal gene expression data. *Bioinformatics*, 22(1), 68–76. https://doi.org/10.1093/bioinformatics/bti742

Preda C., Saporta G. & Lévéder C.L. (2007). PLS classification of functional data. Comput. Stat, 22(2), 223–235. https://doi.org/10.1007/s00180-007-0041-4

Ramsay, J. & Silverman, B. (2005). Functional Data Analysis, 2nd Edition, Springer, New York. https://doi.org/10.1007/b98888

Reynolds, A.G. (2010). *Managing Wine Quality. Volume 1: Viticulture and wine quality*. Woodhead Publishing.

Rouse, J.W., Jr., R.H. Haas, J.A. Schell, & D.W. Deering. (1973). Monitoring vegetation systems in the Great Plains with ERTS. *ERTS Symp*. 3rd, Washington, DC. 10–14 Dec. 1972. NASA SP-351. Vol. I:309–317. NASA, Washington, DC

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. https://doi.org/10.20870/oeno-one.2020.54.1.2557

Sun, L., Gao, F., Anderson, M.C., Kustas, W.P., Alsina, MM, Sanchez, L., Sams, B., McKee, L., Dulaney, W. & White, W.A. (2017). Daily Mapping of 30 m LAI and NDVI for Grape Yield Prediction in California Vineyards. *Remote Sensing*, 9, 317, doi:10.3390/rs9040317.

Tisseyre, B., Mazzoni, C., & Fonta, H. (2008). Whithin-field temporal stability of some parameters in Viticulture: potential toward a site specific management. *OENO One*, 42(1), 27–39. https://doi.org/10.20870/oeno-one.2008.42.1.834

Urretavizcaya, I., Royo, J. B., Miranda, C., Tisseyre, B., Guillaume, S., & Santesteban, L. G. (2017). Relevance of sink-size estimation for within-field zone delineation in vineyards. *Precision Agriculture*, 18(2), 133-144. https://doi.org/10.1007/s11119-016-9450-0

Vélez, S., Rubio, J.A., Andrés, M.I. & Barajas, E. (2019). Agronomic classification between vineyards ('Verdejo') using NDVI and Sentinel-2 and evaluation of their wines. *Vitis Journal of Grapevine Research*, 58, 33–38, doi:10.5073/vitis.2019.58.special-issue.33-38.

Vélez, S., Barajas, E., Rubio, J.A., Vacas, R. & Poblete-Echeverría, C. (2020). Effect of Missing Vines on Total Leaf Area Determined by NDVI Calculated from Sentinel Satellite Data: Progressive Vine Removal Experiments. *Applied Sciences*, 10, 3612. https://doi.org/10.3390/app10103612

Vodyanitskii, Y. N., & Savichev, A. T. (2017). The influence of organic matter on soil color using the regression equations of optical parameters in the system CIE- L*a*b*. *Annals of Agrarian Science*, 15(3), 380–385. https://doi.org/10.1016/j.aasci.2017.05.023

White, R.E. (2015). *Understanding Vineyard Soils*, Second edition., Oxford University Press: Oxford.

Yang, X. D., Wang, J., Xu, M. S., Ali, A., Xu, Y., Lamb, D., Duan, L. C., Yan, K. H., & Yang, S. T. (2019). Effects of the ephemeral stream on plant species diversity and distribution in an alluvial fan of arid desert region: An application of a low altitude UAV. *PLOS ONE*, 14(2), e0212057. https://doi.org/10.1371/journal.pone.0212057