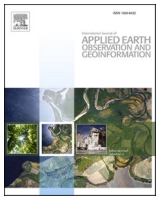
International Journal of Applied Earth Observations and Geoinformation 114 (2022) 103069

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Using NDVI, climate data and machine learning to estimate yield in the Douro wine region

André Barriguinha [a,*], Bruno Jardim [a], Miguel de Castro Neto [a], Artur Gil [b,c]

[a] Nova Information Management School, UNL, Campus de Campolide, 1070-312 Lisboa, Portugal
[b] IVAR - Research Institute for Volcanology and Risks Assessment, University of the Azores, 9500-321 Ponta Delgada, Portugal
[c] cE3c- Centre for Ecology, Evolution and Environmental Changes, Azorean Biodiversity Group, CHANGE – Global Change & Sustainability Institute, Faculty of Sciences and Technology (FCT), University of the Azores, 9500-321 Ponta Delgada, Portugal

ARTICLE INFO

Keywords:
Remote sensing
Vineyard
Yield
Estimation
Prediction
NDVI
Climate
Machine learning

ABSTRACT

Estimating vineyard yield in advance is essential for planning and regulatory purposes at the regional level, with growing importance in a long-term scenario of perceived climate change. With few tools available, the current study aimed to develop a yield estimation model based on remote sensing and climate data with a machine-learning approach. Using a satellite-based time-series of Normalized Difference Vegetation Index (NDVI) calculated from Sentinel 2 images and climate data acquired by local automatic weather stations, a system for yield prediction based on a Long Short-Term Memory (LSTM) neural network was implemented. The study was conducted in the Douro Demarcated Region in Portugal over the period 2016–2021 using yield data from 169 administrative areas that cover 250,000 ha, in which 43,000 ha of the vineyard are in production. The optimal combination of input features, with an Mean Absolute Error (MAE) of 672.55 kg/ha and an Mean Squared Error (MSE) of 81.30 kg/ha, included the NDVI, Temperature, Relative Humidity, Precipitation, and Wind Intensity. The model was tested for each year, using it as the test set, while all other years were used as input to train the model. Two different moments in time, corresponding to FLO (flowering) and VER (veraison), were considered to estimate in advance wine grape yield. The best prediction was made for 2020 at VER, with the model over-estimating the yield per hectare by 8 %, with the average absolute error for the entire period being 17 %. The results show that with this approach, it is possible to estimate wine grape yield accurately in advance at different scales.

1. Introduction

Because yield is a quality grape and wine indicator (De la Fuente Lloreda, 2014; Diago et al., 2015; Santesteban and Royo, 2006; Sun et al., 2017; Zabawa et al., 2019) an early estimation allows growers to find the best balance between vegetative and reproductive growth and make better management and planning decisions (Fernandez-Gonzalez et al., 2011; Fernández-González et al., 2011; Nuske et al., 2014) that can directly impact the business model.

Estimating yield is complex and requires knowing driving factors related to climate, plant, and crop management (Weiss et al., 2020) that directly influence the number of clusters per vine, berries per cluster, and berry weight. These three yield components (Nuske et al., 2014) explain 60 %, 30 %, and 10 % of the yield, respectively (Cunha et al., 2015; Guilpart et al., 2014).

The different approaches for vineyard yield estimation depend on the scale of implementation, and from there, direct (based on manual sampling) or indirect methods (statistical and regression models, proximal/remote sensing, and dynamic or crop simulation models) are used (Bindi et al., 1996; Sirsat et al., 2019; Taylor et al., 2019; Ubalde et al., 2007; Weiss et al., 2020). The first represent the traditional method (De La Fuente et al., 2015) susceptible to spatial and temporal variability and dependent on historical data (Victorino et al., 2022), costly and time consuming (Diago et al., 2015), with low accuracy (Tardaguila et al., 2013) and limited to small-scale application. On the other hand, indirect methods can cope with the limitations off the traditional manual sampling methods and with better results on accuracy, despite the low adoption in real commercial vineyards (Barriguinha et al., 2021).

At a regional level, the vineyard yield estimation goals are more related to regulation and monitoring activities (Barriguinha et al.,

* Corresponding author.
E-mail address: abarriguinha@novaims.unl.pt (A. Barriguinha).

https://doi.org/10.1016/j.jag.2022.103069
Received 2 August 2022; Received in revised form 12 October 2022; Accepted 17 October 2022
Available online 28 October 2022

2021), with yield estimation becoming more and more relevant due to inter-annual variability attributed to climate change's impact on quality, sustainability, efficiency, commercial strategies, regulations, and management of insurance, stock, and quotas (Cunha et al., 2015; Newlands, 2022). The decisions made on this scale can have a large impact, especially in terms of vineyard area and the number of producers involved. A clear example are the Wine Protected Designation of Origin (PDO) label as an European quality scheme that protects high quality wines by linking them to legally defined geographic areas and a set of specific production practices that covers 21 countries (Candiago et al., 2022) with >2,1 million hectares of PDO vineyards (Eurostat. Restructuring of vineyards in the EU continues, 2022).

From previous works, the authors found few examples of yield estimation for regional scales (Barriguinha et al., 2021), divided mainly into climate-based models (Fraga and Santos, 2017; Gouveia et al., 2011; Santos et al., 2020; Sirsat et al., 2019); pollen-based models (Besselat, 1987; Cunha et al., 2015; González-Fernández et al., 2020; Cristofolini and Gottardini, 2000); a combination of one or both adding phenological and phytopathological variables (Fernandez-Gonzalez et al., 2011; Fernández-González et al., 2011); Simulateur mulTIdisciplinaire pour les Cultures Standard, or multidisciplinary simulator for standard crops (STICS) models (Fraga et al., 2015); and models based on correlation with Vegetation Indices (VI) (Arab et al., 2021; Cunha et al., 2010). Only a few are referenced for real environment, producing estimation for decision-making (Barriguinha et al., 2021).

The more commonly used for regional yield estimation are the ones based on the relationship between airborne pollen and yield, relying on the principle that more flowers per area unit in more productive years relates to higher airborne pollen concentrations (Besselat, 1987; Cunha et al., 2015; Fernandez-Gonzalez et al., 2011; Fernández-González et al., 2011, 2020; González-Fernández et al., 2020; Cristofolini and Gottardini, 2000). The main disadvantages/difficulties of using pollen-based models (Barriguinha et al., 2021) are: choosing the best placement for sampling devices to represent effectively spatial variability; the number of observations for model calibration (historical data not commonly available); costly and complex laboratory processes; plant dynamics (high variations of the area with vineyards around the pollen traps); temperature and precipitation variations; vineyard management activities (fertilization impact); and identification of the beginning and final of the pollen season.

Another relevant approach for large areas is the combination of meteorological data and Remote Sensing (RS), based on satellite imagery products such as VI to effectively estimate in advance vineyard yield (Cunha et al., 2010; Gouveia et al., 2011; Sun et al., 2017), with VI explaining crop characteristics and climatic conditions directly influencing crop yield prediction (Muruganantham et al., 2022).

Regarding climate, wine grapes are susceptible and dependent on a region's climatic environment and weather dynamics, with climatic variables impacting vine and grape growth and development (Anderson et al., 2012; Badr et al., 2018; Fraga et al., 2013). Precipitation, humidity, temperature, radiation, and wind have the more influence on grapevine phenology, yield, and wine quality (Badr et al., 2018; Parker et al., 2022; Santos et al., 2012).

VI, as mathematical expressions corresponding to values of growth, vigor, and other vegetation properties, can be derived from satellite time-series images (Matese and Di Gennaro, 2021; Di Gennaro et al., 2019) and are related to vineyard productive and vegetative parameters including yield (Matese and Di Gennaro, 2021; Stamatiadis et al., 2010; Xue and Su, 2017). These indices have been widely implemented within remote sensing (RS) applications (Murali et al., 2021; Snevajs et al., 2022) using multiple satellite platforms. Giovos et al. (Giovos et al., 2021) traced their origin to 1968 with RVI (Birth and McVey, 1968) and in 1973 with NDVI (Rouse et al., 1974).

There are unlimited combinations for creating different VI, but regarding viticulture, NDVI is the most used (Giovos et al., 2021). The Index Database (Henrich et al., 2009) has over 500 different indices extensively used in applications of RS for precision agriculture (Sishodia et al., 2020), with the Normalized Difference Vegetation Index (NDVI) considered a critical parameter (Carrillo et al., 2016; Pelta et al., 2022) capable of reliable yield prediction models (Arab et al., 2021). For discontinuous crops such as a vineyard, proximal data acquisition with spectroradiometers (Maimaitiyiming et al., 2019) or with multispectral cameras mounted on Unmanned Aerial Vehicles (UAV) (Matese and Di Gennaro, 2021) can overcome the limitations attributed to satellite data, namely the soil effects (low for the vineyard as the critical growing stage - were indices/yield correlations tend to increase - occurs when cover crops are in most cases, senescent (Sun et al., 2017), cloud coverage, or the fact that spectral measurement only describes the top part of the canopy, being nevertheless of limited use in large areas due to sparse sampling and high acquisition costs. Gouveia et al. (Gouveia et al., 2011) developed multi-linear regression models of wine production, using NDVI and meteorological variables as predictors to estimate yield (e.g., monthly averages of maximum, minimum, and daily mean temperature and precipitation). A similar approach was made by Cunha et al. (Cunha et al., 2010) with *Satellite Pour l'Observation de la Terre* (SPOT) ten-day synthesis vegetation product (S10) for three different regions in Portugal with significant interannual variability, based on a correlation matrix between the wine yield of a current year and the full set of 10-day synthesis NDVI.

In recent years, Deep Learning (DL) has been considered a breakthrough technology in Machine Learning (ML) and Data Mining (DM), including in the RS research field (Zhong et al., 2019). ML methods are increasingly being used as a tool for crop yield prediction (Arab et al., 2021; van Klompenburg et al., 2020), with Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) being the most widely used DL approaches, with better results when compared to traditional ML approaches for crop yield prediction, taking advantage of the ability to extract features from available data (Muruganantham et al., 2022). This data science approach based on Artificial Neural Networks (ANN), despite recent, is not new to vineyard yield estimation and is leading the alternative methods as one of the most utilized techniques for attempting an early yield estimation. However, it has been limited to small-scale experiments, mostly in controlled environments associated with models based on computer vision and image processing (Barriguinha et al., 2021).

The purpose of the present paper is two-fold: first, to evaluate a new methodology for estimating vineyard yield at the regional level, using the Douro Demarcated Wine Region as a study area with readily available data, and allowing transferability to other regions, to give decision-makers, as far in advance as possible, a good estimation, not only for the total regional and sub-regional wine grape production areas but also at a more detailed scale, considering three sub-regions and 169 sub-administrative regions where there are vineyards in production. Secondly, to cope with the limitations identified in the current model in use in the study area, based on the work of Cunha et al. (Cunha et al., 1999, 2003), which relies on the relationship between airborne pollen and wine production, namely: predicting only for the entire region; predicting wine production instead of wine grape production; the need to maintain representative pollen sampling devices with high maintenance and operational costs, and complex laboratory process to treat the data; and a wide prediction interval.

The proposed model using NDVI and climate data with a DL approach based on a Long Short Term Memory (LSTM) Neural Network can produce an adequate estimation of wine grape yield up to 1–2 months before harvest. To the best of the authors' knowledge, it is the first application of DL to regional vineyard yield estimation.

2. Materials and methods

2.1. Study area

The study was carried out using the different datasets described in

the next points, covering six years (2016–2021).

The study area is covered by the Douro Demarcated Wine Region (DDR), which is the oldest wine-demarcated region in the world. It is located in the northeast of Portugal (Fig. 1) in the Douro watershed, surrounded by complex terrain with unique orographic, mesological and climatic characteristics. The region extends over a total area of about 250,000 ha and is divided into three naturally distinct sub-regions ("Baixo Corgo", "Cima Corgo" and "Douro Superior"), not only due to climatic factors but also socio-economic ones. Regarding regulatory purposes, the DDR has some specificities. From the total area planted with vines (about 43,000 ha), only 26,000 ha are authorized to produce Port Wine. In fact, the vineyards suitable for production are selected according to qualitative criteria (classified through a scale) that consider soil, climate, and cultural parameters with decisive importance in the qualitative potential of the plots. Only vineyards with more than five-year-old can be considered for producing Port Wine. According to the cadastral elements, each plot is entitled to a certain benefit coefficient that needs to be determined every year and indexed to the classification scale. The vineyard areas are divided into 104,000 individual plots (47 % on "Cima Corgo"; 39 % on "Baixo Corgo", 14 % on Douro Superior) spread into 169 administrative regions called "Freguesias" (Parishes). These were considered for the present study as the minimum scale areas for grape yield estimation, followed by the sub-regions and the entire DDR.

Each year the vineyard area in production varies since there are new areas, areas not yet in production, and areas considered unsuitable for producing wine with denomination of origin. This was considered for the present study due to the impact on determining the grape yield per area unit (kg/ha) for each year and each parish. Table 1 shows the aggregated data for the three sub-regions and the entire DDR.

2.2. Remote sensing data

For the present study, the initial dataset used to produce the temporal NDVI profiles was collected from Copernicus Sentinel-2A (launched on June 23rd, 2015) and 2B (launched on March 7th, 2017), with a Level-2A of processing level and 10 m of spatial resolution, for the period 2016–2021. A total of 686 usable Sentinel images were retrieved from the Copernicus Open Access Scientific Hub, corresponding to 343 different acquisition dates (two images per acquisition date due to the study area extension), from January 11th, 2016, to December 30th, 2021, from which the NDVI was calculated using Band 4 (RED) and Band 8 (NIR) as described in equation (1).

$$NDVI = (NIR - RED)/(NIR + RED) \quad (1)$$

Where: NIR is the reflectance in the near-infrared channel and RED is the reflectance in the red channel.

From the initial dataset, as explained in 2.4, only values between March (when on average, budburst occurs, marking the beginning of seasonal grapevine growth and resumed physiological activity) and October (when most of the harvest has already taken place) were considered (Table 2). Those were used to build a spatiotemporal cube by clipping the areas of each parish with the vineyard in production, resulting in the average NDVI values for each parish at each date used in the model described in 2.6.

With both satellite data, the best average temporal resolution for the study area is five days from 2018, 2020, and 2021 (71 images retrieved) followed by 2019 with six days (60 images retrieved). The lower temporal resolutions in 2016 (15 days – 25 images retrieved) and 2017 (8 days – 43 images retrieved) are related to the inexistence of the Sentinel-2B sensor until March 2017.

Regarding the expected negative effect of cloud coverage, we first considered all images for conducting the evaluation of the yield prediction model through a stepwise backward feature selection process, thus allowing us to assess the true impact and the limit to which we might consider the validity (or not) of each image.

2.3. Climate data

The DDR climate is the Mediterranean, with continental influence and marked annual thermal contrast and water stress, especially during summer with the vineyards located in some of the aridest regions in Europe, with strong and consistent post-flowering vine water and thermal stress (Cunha et al., 2010).

The climate data used in the present study resulted from observed daily values of the parameters described in Table 3 acquired by six IPMA (*Instituto Português do Mar e da Atmosfera* - https://www.ipma.pt/pt/) automatic weather stations between 2016 and 2021. The considered areas of influence of every station (closest distance to the plot's polygons) are shown in Fig. 2. For the present study, the climate data computed in the prediction model described in 2.5 follows the same date range (March-October), similar to the approach made regarding remote sensing data.

2.4. Phenology data

Phenology data was used to define the different timeframes

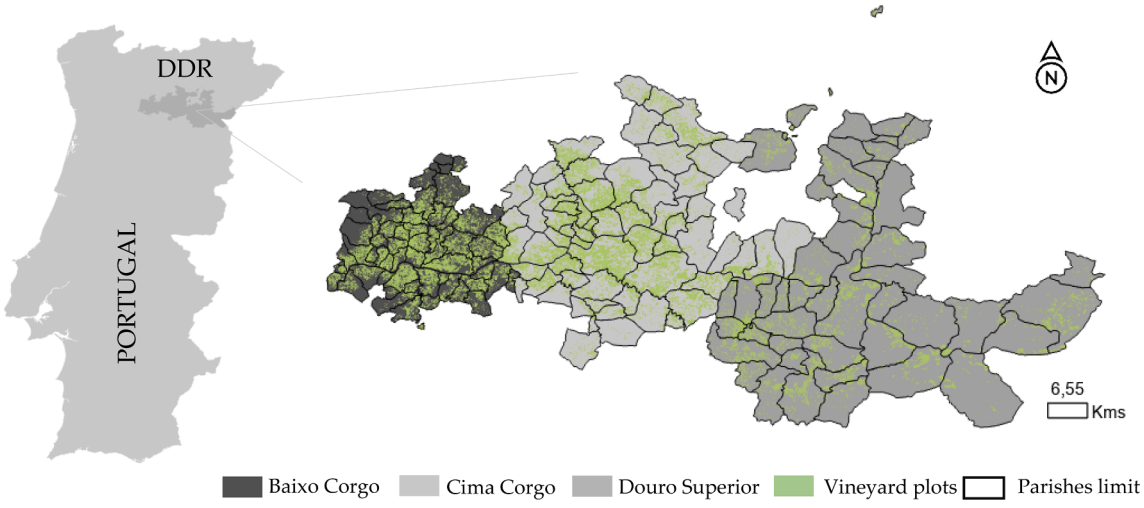


Fig. 1. Study area overview with the three sub-regions, vineyard plots locations (provided by the IVV - Instituto da Vinha e do Vinho, IP (Portuguese Institute of Vine and Wine), and 169 administrative regions considered for the present study as the minimum scale areas for grape yield estimation.

Table 1
Vineyard area distribution in the DDR sub-regions (2016–2021).

Year	BC (57 parishes)			CC (64 parishes)			DS (48 parishes)			DDR (169 parishes)		
	sum	avg	sd	sum	avg	sd	sum	avg	sd	sum	avg	sd
2016	12,808	224.7	165.7	19,700	307.8	318.5	9598	200.0	171.0	42,106	249.1	240.0
2017	12,842	225.3	166.2	19,778	309.0	319.9	9600	200.0	170.4	42,220	249.8	240.8
2018	12,794	224.5	165.0	19,899	310.9	322.9	9661	201.3	174.0	42,354	250.6	242.8
2019	12,740	223.5	164.9	19,958	311.8	325.8	9684	201.7	175.3	42,382	250.8	244.6
2020	13,202	231.6	171.1	20,429	319.2	334.2	10,078	210.0	181.4	43,709	258.6	251.3
2021	12,966	227.5	166.5	20,510	320.5	335.7	10,207	212.6	182.1	43,683	258.5	251.3

BC (Baixo Corgo sub-region); CC (Cima Corgo sub-region); DS (Douro Superior sub-region); DDR (Douro Demarcated Region); sum (productive vineyard area - hectare); avg (average productive vineyard area/parish – hectare); sd (standard deviation).

Table 2
Descriptive statistics of NDVI data from the entire DDR for the areas with vineyards (2016–2021) – from March to October.

Year	CC		MIN		MAX		MEAN		PCT90	
	avg	sd	avg	sd	avg	sd	avg	sd	avg	sd
2016	34.8	39.0	0.026	0.062	0.648	0.224	0.303	0.126	0.420	0.165
2017	23.7	32.1	0.009	0.070	0.603	0.218	0.250	0.117	0.348	0.153
2018	25.3	33.7	−0.002	0.086	0.628	0.222	0.275	0.136	0.393	0.176
2019	30.3	34.0	−0.002	0.076	0.569	0.242	0.233	0.129	0.334	0.172
2020	42.5	37.5	−0.006	0.096	0.540	0.283	0.234	0.154	0.332	0.202
2021	34.0	35.6	0.001	0.077	0.606	0.241	0.272	0.139	0.385	0.182

CC (Average Cloud Coverage - %); MIN (Average Minimum NDVI value); MAX (Average Maximum NDVI value); MEAN (Average NDVI value); PCT90 (Average NDVI value 90 percentile); sd (standard deviation).

Table 3
Descriptive statistics of climate data for the six areas of influence, considering the data from the six automatic weather stations used for the present study (2016–2021) – from March to October.

	P (mm)		T (°C)		H (%)		W (m/s)		R (KJ/m^2)	
	avg	sd	avg	sd	avg	sd	avg	sd	avg	sd
Area 1	429	152	16.7	4.5	66.0	8.6	1.6	0.3	19,813	5878
Area 2	256	110	19.8	4.8	55.3	9.6	1.7	0.3	20,251	5045
Area 3	438	187	16.1	4.8	64.3	9.3	2.3	0.4	20,062	4283
Area 4	364	181	15.5	4.6	67.1	8.8	1.5	0.3	20,025	5489
Area 5	258	110	18.4	5.0	63.1	10.2	2.1	0.4	20,255	5213
Area 6	298	108	17.8	5.0	57.5	9.9	2.2	0.4	21,125	5371

P (Average annual total precipitation amount - mm); T (Average daily air temperature at 1.5 m – °C); H (Average daily relative humidity - %); W (Average daily wind speed – m/s); R (Average daily global radiation – KJ/m^2); sd (standard deviation).

necessary for the model to predict yield as far in advance as possible effectively. The three main grapevine phenological stages are (1) budburst (BUD), which marks the beginning of seasonal grapevine growth and resumed physiological activity after a long period of winter dormancy; (2) flowering (FLO), which is crucial for the reproductive cycle and closely followed by the fruit set stage; and (3) veraison (VER), which initiates the ripening stage, correlated to wine grape quality attributes (Fraga et al., 2016; Jones and Davis, 2000). The dates for the beginning of each stage were collected from the harvest report generated by ADVID (Association for the Development of Viticulture in the Douro Region - https://www.advid.pt/en) each year (ADVID, 2016; ADVID, 2017; ADVID, 2018; ADVID, 2019; ADVID, 2020; ADVID, 2021). The harvest (HAR) start and end dates were collected from the IVDP dataset (Instituto dos Vinhos do Douro e do Porto, I.P. - https://www.ivdp.pt/en) according to the registration of grape entry in the wine-producing facilities (Table 4).

2.5. Yield data

Yield data was provided by IVDP for each parish from 2016 to 2021. The data is collected yearly in grape receptions units scattered along the entire DDR, with the grapes' amount (kg) and origin (parish) recorded for each delivery. The evolution through the different years is aggregated by sub-region and for the entire DDR in Table 5. This same table also shows the average production in kg/ha.

2.6. Yield prediction model

The system implemented for yield prediction is a Long Short Term Memory (LSTM) Neural Network (Hochreiter and Schmidhuber, 1997) implemented using the Keras framework (https://keras.io/), an open-source software library that provides a Python interface for artificial neural networks, part of TensorFlow library (https://www.tensorflow.org/).

This model was chosen since its architecture is designed to learn long-term dependencies in sequences like time series. The LSTM can process sequences of variables by holding a cell state c_t that carries information across the different time steps of the sequence, receiving minimal updates based on three different gates, namely the forget gate (Equation (2)), the input gate (Equation (3)), and the output gate (Equation (4)).

$$f_t = \sigma\left(W_f \bullet [h_{t-1}; x_t] + b_f\right) \quad (2)$$

$$i_t = \sigma(W_i \bullet [h_{t-1}; x_t] + b_i) \quad (3)$$

$$o_t = \sigma(W_o \bullet [h_{t-1}; x_t] + b_o) \quad (4)$$

Fig. 3 displays the system's architecture. The network receives an

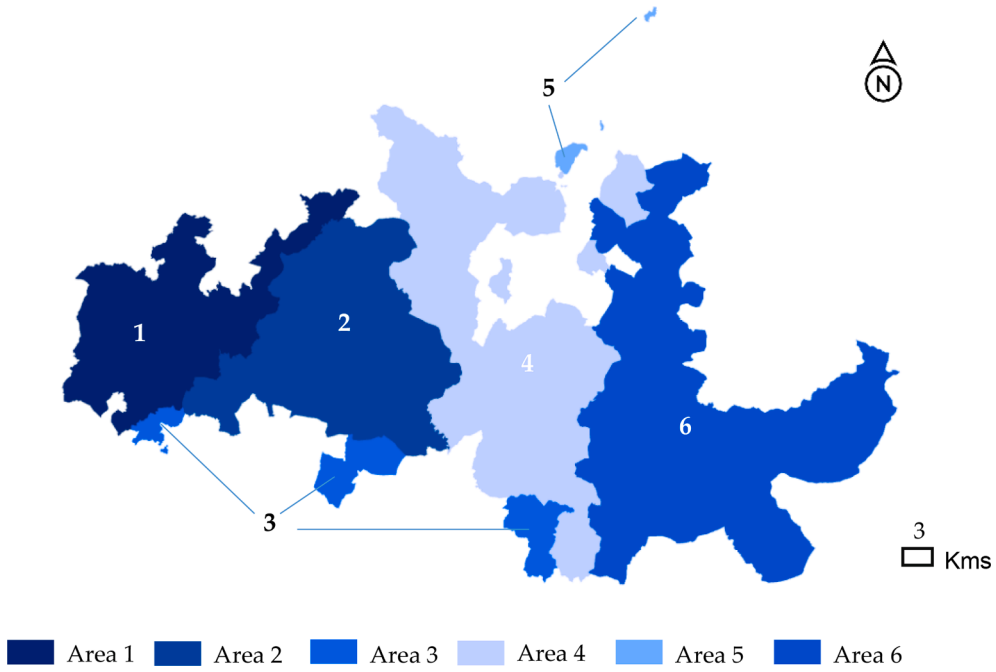


Fig. 2. Considered areas of influence for the six weather stations used in the present study.

Table 4
Average start date for the main phenological stages and harvest in the DDR (2016–2021).

Year	BUD	FLO	VER	HAR (start)	HAR (end)
2016	>15 Feb	>15 May	>15 Jul	>18 Aug	<17 Nov
2017	>15 Mar	>15 Apr	>15 Jun	>07 Aug	<21 Nov
2018	>22 Mar	>19 May	>26 Jul	>14 Aug	<15 Nov
2019	>12 Mar	>06 May	>13 Jul	>12 Aug	<15 Nov
2020	>04 Mar	>08 May	>07 Jul	>05 Aug	<18 Nov
2021	>06 Mar	>07 May	>08 Jul	>26 Jul	<16 Nov

BUD (budburst); FLO (flowering); VER (veraison); HAR (harvest).

input $x_{yp} = [x_t, x_{t+1}, \cdots, x_{t+n}]$, a sequence of 49 vectors (sequence length corresponding to the number of observations between March and October), each containing the observed values of every input (NDVI, Rad., Temp., CC, Hum., Prec., Wind – see Table 9) at a point in time t, for a given year y and parish p.

This input, together with a cell state c_{t-1} and the hidden state h_{t-1}, of the previous time step t, are passed through the network. The forget gate f_t, a sigmoid layer, takes h_{t-1} and x_t and computes what information should be erased from the previous steps at the current one. Similarly, the input gate i_t, another sigmoid layer, decides what information from the input x_t should be kept. Next, x_t passes a tanh layer that computes new candidate values $\widetilde{c}_t$ for the cell state (Equation (5)). The cell state c_t is updated by multiplying the old one with the output of the forget gate f_t, and adding the resulting value with the product of the input gate i_t result and the candidate values $\widetilde{c}_t$ (Equation (6)). This process enables the network to store the information from the current time step and pass it to future steps.

$$\widetilde{c}_t = tanh(W_c \bullet [h_{t-1}; x_t] + b_c) \quad (5)$$

$$c_t = f_t{*}c_{t-1} + i_t{*}\widetilde{c}_t \quad (6)$$

Lastly, x_t goes through the output gate o_t. The resulting value is multiplied with the cell state value squashed by a tanh layer (Equation (7)).

$$h_t = o_t{*}tanh(c_t) \quad (7)$$

Through this calculation, we obtain the output value of the network at the current time step h_t. The hidden state of the last time step, h_{yp}, goes through a linear activation layer that computes the Yield Production in Kg/ha for that year y and parish p.

3. Results and discussion

LSTM is one of the most widely used deep learning algorithms in crop yield prediction, along with CNN and Deep Neural Networks (DNN), with temperature, precipitation, and humidity among the most used independent variables (van Klompenburg et al., 2020) to predict yield (dependent variable). This is consistent with the developed model as all three variables are part of the model with the best metrics, with NDVI

Table 5
Wine grapes yield in the sub-regions of the DDR (2016–2021).

Year	BC (57 parishes)			CC (64 parishes)			DS (48 parishes)			DDR (169 parishes)		
	Sum	avg	sd	sum	avg	sd	sum	avg	sd	sum	avg	sd
2016	43,747	3223	539	79,329	3427	806	32,430	2969	819	155,506	3228	750
2017	57,671	4258	1017	76,524	3565	1161	34,551	3342	816	168,746	3736	1088
2018	41,828	3197	769	67,487	3047	863	33,297	3104	803	142,611	3114	813
2019	64,514	4747	1124	95,523	4370	1132	44,187	4212	1214	204,024	4452	1167
2020	44,824	3171	737	73,819	3200	931	35,129	3194	968	153,771	3189	877
2021	55,723	4083	957	95,896	4233	1148	44,167	3846	1149	195,787	4073	1092

BC (Baixo Corgo sub-region); CC (Cima Corgo sub-region); DS (Douro Superior sub-region); DDR (Douro Demarcated Region); sum (total annual wine grape production in tons); avg (average annual wine grape production in kg/ha); sd (standard deviation).

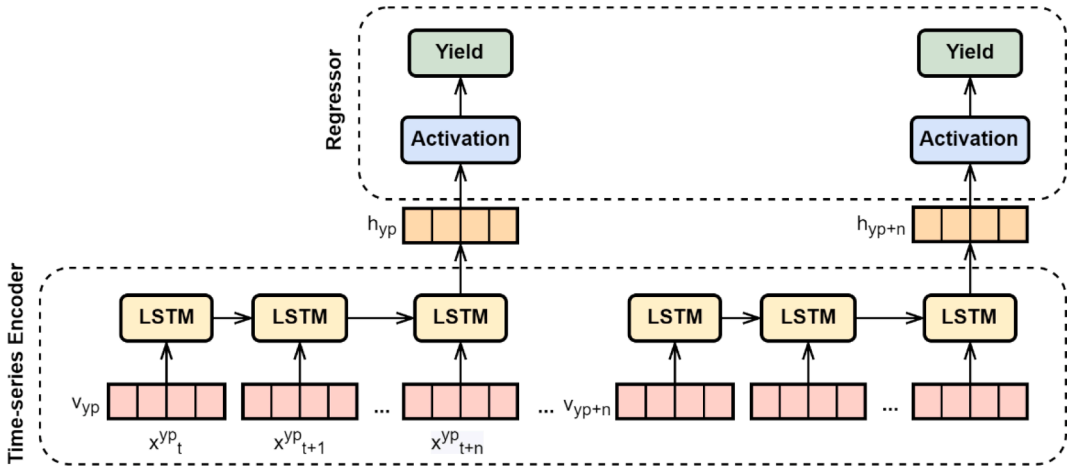


Fig. 3. Yield prediction model overview. The model is divided into two parts: 1) a time-series encoder module that uses an LSTM to generate dense representations; and 2) a regressor module that receives the last hidden state of the LSTM and calculates the yield volume for a year and location (parish).

and wind also as explanatory features (also referenced by T. van Klompenburg, et al. (van Klompenburg et al., 2020) as commonly used for the same purpose).

The average annual wine grape production inter-annual variability can be observed in Table 5 and Fig. 4 and is transversal to the different scales of observation, namely DDR, the three subregions (BC, CC and DS), and the 169 parishes. The lowest total production occurred in 2018, with 142,611 tons of grapes for the entire DDR, reaching its peak in the following year (2019) with a value of 204,024 tons of grapes. The behavior of each sub-region reveals the same tendency registered for the

Fig. 4. Spatial distribution of the average Yield Production in kg/ha by year and parish.

entire DDR either in total aggregated value or production per area unit (kg/ha).

Fig. 4 displays the spatial distribution of the average wine grape yield production in kg/ha for each year and each parish. The variability between years and parishes is visible (the values are related to the areas with vineyards and are represented by parish administrative boundaries for easier visualization).

This variability was identified by Cunha et al. (Cunha and Richter, 2011) for the DDR and other regions (Cunha et al., 2010). It can be explained by the spatio-temporal distribution that characterizes agricultural systems as a whole and vineyards in particular, with vulnerability to inter-annual climate variability, especially in the case of our study area, where the vineyards grow under marginal conditions for production with distinctive climatic, topographic and soil characteristics (Gouveia et al., 2011) with temperature and precipitation having a deep connection to yield variability (Camps and Ramos, 2012).

The range between minimum and maximum production per area unit also shows high inter-annual variability, reaching higher values in years when total productivity was higher, namely in 2017, 2019, and 2021.

The spatial autocorrelation was evaluated using Global Moran's I, showing the randomness of the yield data for all years (Table 6).

Calculated NDVI values show a great inter-annual variability throughout the six years between March and October of each year (Table 2), with a strong standard deviation, in the vineyard areas of each parish. Being a region with a low precipitation level, low NDVI values were expected.

The spatial distribution considering the average NDVI values for the period between March and October from vineyard plots represented at the parish level is shown in Fig. 5. The identifiable clusters in the different sub-regions, with BC showing higher average values throughout the vegetative cycle, followed by CC and DS with the lower scores, were evaluated using Global Moran's Index, indicating a clustered pattern of the average NDVI values for all years (Table 7). As already mentioned, these values represent the average pixel values inside the areas with vineyards in production in each parish.

The NDVI profile over the crop vegetative cycle tends to increment after BUD, reaching its highest values between FLO and VER and decreasing after harvest. This is consistent with the work of several authors (Boulton et al., 1996; Cunha et al., 2010; Gouveia et al., 2011) and associated with the growing period of the vineyards until flowering in May.

Vineyard growth could be restricted in the early stages of the growing season due to the soil water content frequently low at BUD and the lack of winter rainfall (Cunha et al., 2010). The higher average NDVI values demonstrate this in 2016 and 2018, where winter rainfall was more elevated. The highest average NDVI value was recorded on July 9th of 2018, with a 2 % cloud coverage value. The effect of cloud coverage is very noticeable, as expected, although the cloud coverage percentage is related to the entire image (to cover the whole DDR spatially, two Sentinel-2 scenes are required). It is usually not noticeable if it concerns areas effectively occupied by vineyards.

Climate plays a fundamental role in the productivity of the vineyard (Fraga et al., 2013, 2016) as phenological events and composition are significantly influenced by the climate of preceding months, especially during the growing season (Bock et al., 2011). Weather variables can explain 57.3 %, 64.3 %, and 57.8 % of the variance in yield, sanitary status, and grape composition (Ferrer et al., 2017). According to Gouveia et al. (Gouveia et al., 2011), low rainfall in March positively affects vegetative growth, and high temperatures in late spring are beneficial. This can be seen in the year 2018, where abnormal high precipitation in March accompanied by below-average mean temperatures (Fig. 6) was reflected in the production, being the year with the lowest production according to the series considered in the present study (Table 5). Although the precipitation throughout the cycle allowed the recovery from water stress that occurred in 2017 (the driest year of the analyzed time range, with lower NDVI values, especially in CC and DS subregions, as visible in Fig. 5), its volume and timing had a negative impact through an increase in the phytosanitary pressure, scorching and dehydration at a later stage (ADVID, 2018). Precipitation stands out in the years of the present study not only by the high variability between the different areas, months, and years but also from the difference in the available 30-year climatological series (1931–1960 (ADVID, 2016) and 1970–2000 (ADVID, 2021).

The year 2016 was characterized by a warm and rainy winter, a cold and extremely rainy spring, and very hot and dry summer, contributing to an earlier BUD and a later delay in the previous phenology stages. Intense precipitation in a sensitive phase of the vegetative cycle gave rise to the strong pressure of mildew (ADVID, 2016). Regarding yield, it was the third lowest year of the studied period and below the average of 170,074 tons (−9 %) for the entire DDR.

The year 2017, as already stated, was an arid and hot year where the climatic conditions contributed to a significant advance in the vegetative cycle. The prolonged scarcity of precipitation and very high temperatures led to intense hydric and thermal stress at an early stage of the cycle, conditioning the evolution of the vegetation wall and impacting production (ADVID, 2017). Despite that, according to the data provided by IVDP, 2017 had a grape production higher than the one recorded in 2016, but still under the average for the six years (−1 %), which can be justified by the quasi-absence of pressure on the phytosanitary aspect.

As already stated, the year with the lowest total production was 2018 (−16 % from the average), with a cold and dry winter, cold and extremely rainy spring, and, in its first phase, a cold and rainy summer, and in its second phase, a hot and arid one. Despite the perceived high production potential, climate instability significantly reduced it due to the abnormal harmfulness of the downy mildew (ADVID, 2018).

The years with the highest production were 2019 (+17 % from the average) and 2021 (+15 % from the average). Both years are characterized by standard dry years with low disease impact (ADVID, 2019; ADVID, 2021) and are the closest to the 30-year Climatological Normals series. This is also true for the year 2020, although in this case, the spring precipitation led to high pressure regarding the phytosanitary aspect (namely mildew and powdery mildew (ADVID, 2020), which could explain the lower total production (−11 % from the average).

Testing for normality through the D'Agostino's (D'agostino, 1970) and Shapiro-Wilk Test (Shapiro and Wilk, 1965), we concluded that none of the variables was normally distributed, despite some of them, namely Yield Production, Temperature, and Relative Humidity, showing a Gaussian pattern. We calculated the Pearson correlation to assess how the different explanatory variables are related (Table 8).

NDVI presents a negative correlation of −0.87 with Cloud Coverage, −0.61 with Humidity, −0.51 with Precipitation and −0.19 with Wind Intensity. On the other hand, NDVI has a positive correlation of 0.75 with radiation and 0.47 with Temperature. The polarity of the correlations provides a clear distinction between variables that exhibit a similar pattern to NDVI, namely Radiation, and variables that present almost an opposite behavior, namely Cloud Coverage and Humidity. It is also possible to conclude that radiation values vary inversely to Cloud Coverage and Humidity.

Table 6
Spatial autocorrelation assessment (with Global Moran's Index) regarding parish-based wine grape yield in kg/ha (2016–2021).

	Moran's Index	z-score	p-value
2016	−0.003765	0.171343	0.863954
2017	0.017237	1.042057	0.297385
2018	−0.024546	−1.010945	0.312043
2019	−0.022819	−0.833334	0.404656
2020	−0.002469	0.144244	0.885308
2021	−0.005352	0.026708	0.978693

(Spatial relationships: Inverse distance; Distance method: Euclidian; Standardization: Row; Distance threshold: 14187 m).

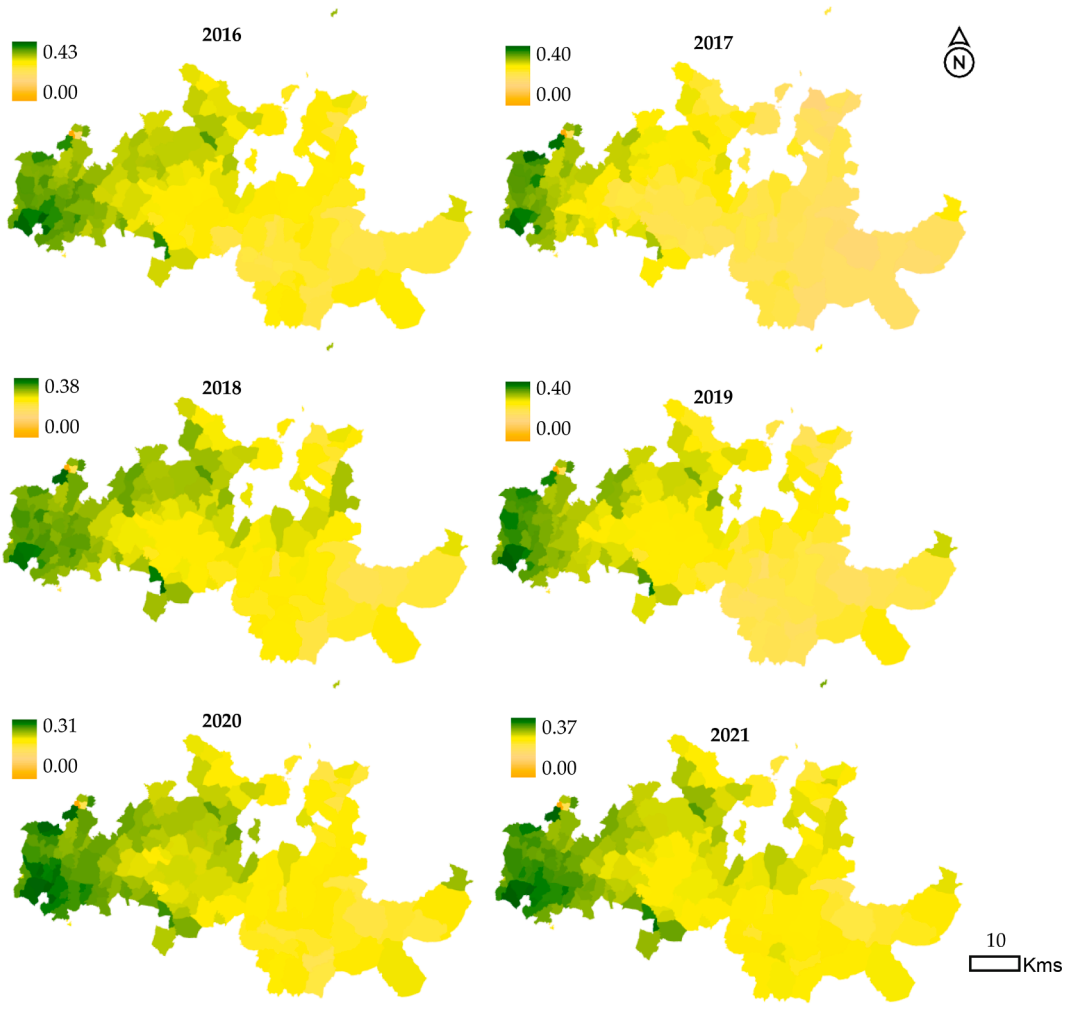


Fig. 5. Spatial distribution of the average NDVI for each year and each parish (March-October).

Table 7
Spatial autocorrelation assessment (with Global Moran's Index) regarding parish-based average NDVI (2016–2021).

	Moran's Index	z-score	p-value
2016	0.522983	21.769401	0.000000
2017	0.591388	24.319456	0.000000
2018	0.367304	15.532358	0.000000
2019	0.536924	22.273663	0.000000
2020	0.532426	22.293926	0.000000
2021	0.380097	16.151957	0.000000

(Spatial relationships: Inverse distance; Distance method: Euclidian; Standardization: Row; Distance threshold: 14187 m).

3.1. Yield prediction model optimization

Yield prediction was evaluated for the period between March and October by performing a random training/test split, leaving 80 % of the observations to train the model and 20 % to test it. The metrics used to evaluate the prediction performance were the MAE and the MSE. The different number of LSTM layers and 8, 16, 32, 64, and 128 hidden units were tested during the training setup. We also introduced a dropout layer with different values and experimented with different learning optimization methods and rates. Ultimately, a small model with only one layer and 16 hidden units, no dropout, using the Adam optimizer (Kingma, 2014) and a fixed learning rate of 0.001, yielded the best performances (execution environment: GPU; Loss Function: Cross-entropy). Moreover, to understand the impact of each input variable on the yield prediction performance and find the best combination of variables, we ran a stepwise backward feature selection process, in which we started by evaluating the model using all variables as input and gradually removing one at the time, based on their correlation with NDVI (higher absolute correlations were removed first). Table 9 summarizes these experiments.

In the scenario in which the model's performance increases or remains the same after removing a variable, we excluded the variable for the next tests. Alternatively, the variable would be added in the following experiment if the performance decreased. We also ran the model considering only NDVI as input (step 8). The optimal combination of input features, with an MAE of 672.55 and an MSE of 81.30, considered NDVI, Temperature, Relative Humidity, Precipitation, and Wind Intensity (step 3). The removed variables in the best model, Radiation, and Cloud Coverage, were the ones with the highest correlation with NDVI (see Table 8). This was expected since their explanatory power is already expressed in NDVI. On the other hand, the most significant drop in model performance seems to be when removing the feature Wind, the one with the lowest correlation with NDVI. Wind influence can be negative (e.g., physiological effects of photosynthesis disruption, breaking off new shoots, increasing evapotranspiration) and positive (e.g., reduced disease infestations, limiting the occurrence of radiation frosts) on vine health and yield. The data referring to this variable shows high interannual variability between areas of influence, where areas 3 and 6 stand out with consistently higher values than the other areas. Also, it is worth noting that the model using only NDVI as a

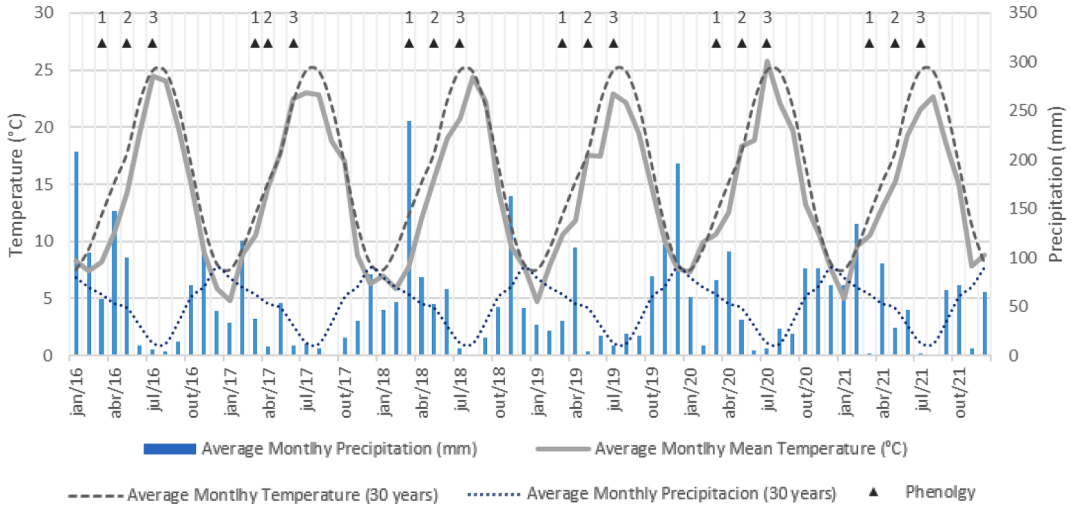


Fig. 6. Average monthly precipitation and mean temperature (2016–2021); Average monthly precipitation and mean temperature (30-year climatological series for 1931–1960 and 1970–2000); Phenology (BUD: 1; FLO: 2; VER:3) for DDR.

Table 8
Pearson correlation between the variables used for Yield Prediction.

Pearson Correlation	NDVI	R	T	W	P	H	CC
NDVI	1.00						
Radiation	0.75	1.00					
Temperature	0.47	0.61	1.00				
Wind	−0.19	−0.07	−0.28	1.00			
Precipitation	−0.51	−0.49	−0.36	0.36	1.00		
Humidity	−0.61	−0.65	−0.61	0.07	0.49	1.00	
Cloud Coverage	−0.87	−0.66	−0.43	0.11	0.38	0.65	1.00

NDVI (Normalized Difference Vegetation Index); R (Global radiation); T (Air temperature); W (Wind speed); P (Precipitation); H (Relative humidity); CC (Cloud Coverage).

feature for yield prediction achieves higher performance than the one using NDVI, Humidity, Precipitation, and Temperature.

3.2. Yield prediction model analysis

The model with the best metrics was run to analyze in more detail its prediction performance. The prediction was made for each year, using it as the test set, while all other years were used as input to train the model. We considered two different moments in time, corresponding to FLO (May) and VER (July) phenology stages, considering the main characteristics of vineyards at DDR (Table 4).

The model's performance at the FLO stage is considered very poor with an average absolute prediction error for the entire DDR between 2016 and 2021 of 38 % against the 17 % average error achieved when the same model is run at the VER stage. This error represents the deviation regarding kg/ha from the actual average of wine grapes collected.

Analyzing each year for the entire DDR and the different sub-regions at the FLO stage (Table 10), the best prediction was made in 2021 for the whole DDR, with the model underestimating the yield per hectare at 25 % and 19 % for the DS sub-region. In 2016, we can see the most significant difference between predictions in sub-regions, with DS showing almost twice the error as CC. The worst performances are for 2017, an arid year, and 2019, the year with the highest productivity per hectare compared to the other years.

The results mainly improved at the VER stage (Table 11) and didn't follow the same FLO estimation pattern. The best prediction was made in 2020 for the whole DDR, with the model overestimating the yield per hectare at 8 % and 6 % for CC and DS sub-regions, respectively, followed by the results for 2019. These years are less deviant from normal climate

Table 9
Evaluation of Yield prediction model through a stepwise backward feature selection process. Best metrics highlighted.

Step	Variables	MAE (kg/ha)	MSE (kg/ha)
1. All variables	NDVI, Rad., Temp., CC, Hum., Prec., Wind	688.29	83.77
2. Remove Cloud Coverage	NDVI, Rad., Temp., Hum., Prec., Wind	678.84	81.80
3. Remove Radiation	NDVI, Temp., Hum., Prec., Wind	**672.55**	**81.30**
4. Remove Relative Humidity	NDVI, Temp., Prec., Wind	685.28	82.80
5. Remove Precipitation	NDVI, Temp., Wind. (+) **Hum.**	680.05	82.64
6. Remove Temp	NDVI, Hum., Wind., (+) **Prec.**	698.93	83.73
7. Remove Wind	NDVI, Hum., Prec., (+) **Temp.**	823.49	106.85
8. Only NDVI	NDVI	766.59	99.24

variables, despite having the higher production (in 2019) and the second lower production (in 2020). This is also true for 2021, but with a worst prediction. The biggest difference between predictions in sub-regions is in 2017, with CC and DS having almost triple of error value as BC, in 2020 where the error in BC doubles the one in the other sub-regions. The worst performances are for 2016, with the model underestimating yield, characterized by climate conditions that favored phytosanitary problems, and in 2017, especially arid in CC and DS sub-regions.

Table 12 illustrates the descriptive statistics at parish level. Furthermore, Fig. 7 shows the spatial distribution by parish of the average absolute errors for the six-year analyzed period. The model's overall performance at the parish level is considered poor, not only regarding the error but also the high inconsistency when we look at individual parishes and the error behavior for each year.

Between the two stages, it is clear that also at the parish level, the model works better at VER stage, although in 36 % of the parishes the error is lower if the model is run in the FLO stage, with an average difference of 10 %, being the lowest 2 % and the highest 23 %.

3.3. Comparative study and discussion

To further evaluate our model, we conducted a comparative study regarding other alternatives for regional vineyard yield estimation (Table 13). A more in-depth comparison was made considering the

Table 10
Prediction for the DDR and sub-regions made at the FLO stage.

Year	Region	AVG	PRED	DIF	DIF_abs	DIF_%	DIF_abs_%
2016	DDR	3228	4196	968	968	30	30
	BC	3223	4157	934	934	29	29
	CC	3427	4194	767	767	22	22
	DS	2969	4245	1276	1276	43	43
2017	DDR	3736	1873	−1863	1863	−50	50
	BC	4258	1921	−2337	2337	−55	55
	CC	3565	1865	−1700	1700	−48	48
	DS	3342	1827	−1515	1515	−45	45
2018	DDR	3114	4124	1010	1010	32	32
	BC	3197	4286	1089	1089	34	34
	CC	3047	3983	937	937	31	31
	DS	3104	4118	1013	1013	33	33
2019	DDR	4452	1779	−2673	2673	−60	60
	BC	4747	1711	−3036	3036	−64	64
	CC	4370	1751	−2619	2619	−60	60
	DS	4212	1897	−2315	2315	−55	55
2020	DDR	3189	4127	939	939	29	29
	BC	3171	4110	939	939	30	30
	CC	3200	4121	920	920	29	29
	DS	3194	4156	962	962	30	30
2021	DDR	4073	3071	−1001	1001	−25	25
	BC	4083	2999	−1084	1084	−27	27
	CC	4233	3090	−1143	1143	−27	27
	DS	3846	3132	−714	714	−19	19

BC (Baixo Corgo sub-region); CC (Cima Corgo sub-region); DS (Douro Superior sub-region); DDR (Douro Demarcated Region); AVG (Real wine grape production average in kg/ha); PRED (Estimated wine grape production in kg/ha); DIF (PRED-AVG in kg/ha); DIF_abs (DIF in absolute value); DIF_% (PRED-AVG in %); DIF_abs (DIF_% in absolute value).

Table 11
Prediction for the DDR and sub-regions made at VER stage.

Year	Region	AVG	PRED	DIF	DIF_abs	DIF_%	DIF_abs_%
2016	DDR	3228	2459	−769	769	−24	24
	BC	3223	2454	−769	769	−24	24
	CC	3427	2460	−967	967	−28	28
	DS	2969	2463	−505	505	−17	17
2017	DDR	3736	4441	706	706	19	19
	BC	4258	4605	347	347	8	8
	CC	3565	4392	826	826	23	23
	DS	3342	4312	970	970	29	29
2018	DDR	3114	3621	507	507	16	16
	BC	3197	3730	533	533	17	17
	CC	3047	3552	505	505	17	17
	DS	3104	3583	479	479	15	15
2019	DDR	4452	3827	−625	625	−14	14
	BC	4747	4000	−747	747	−16	16
	CC	4370	3737	−633	633	−14	14
	DS	4212	3741	−470	470	−11	11
2020	DDR	3189	3446	257	257	8	8
	BC	3171	3549	378	378	12	12
	CC	3200	3399	198	198	6	6
	DS	3194	3385	191	191	6	6
2021	DDR	4073	3312	−761	761	−19	19
	BC	4083	3226	−857	857	−21	21
	CC	4233	3373	−860	860	−20	20
	DS	3846	3333	−513	513	−13	13

BC (Baixo Corgo sub-region); CC (Cima Corgo sub-region); DS (Douro Superior sub-region); DDR (Douro Demarcated Region); AVG (Real wine grape production average in kg/ha); PRED (Estimated wine grape production in kg/ha); DIF (PRED-AVG in kg/ha); DIF_abs (DIF in absolute value); DIF_% (PRED-AVG in %); DIF_abs (DIF_% in absolute value).

model currently used in the study area by the local authorities.

Comparing our results with the works of Cunha et al. (Cunha et al., 1999, 2003), that developed and used an estimation model in the same study area (DDR) and relied on the relationship between airborne pollen and yield (Table 14), we can state that our results are very satisfactory. The pollen-based model predicts wine production for the whole DDR with a minimum and maximum threshold, and ADVID has used it since 1992 with the predictions made yearly at the VER stage. To compare both errors, we considered the average absolute error and a conversion factor of 750 kg of grape for 550 L of wine (average based on the IVDP data set for the six years).

Table 12
Descriptive statistics for average absolute error in each parish (2016–2021).

	avg	sd	median	min	max
2016	0.26	0.14	0.27	0.00	0.80
2017	0.38	0.58	0.22	0.00	6.55
2018	0.32	0.36	0.22	0.00	2.21
2019	0.24	0.21	0.20	0.00	1.51
2020	0.29	0.38	0.18	0.00	3.11
2021	0.26	0.23	0.22	0.00	2.10

avg (Parish average absolute error); sd (standard deviation); median (Parish average median absolute error); min (Parish average minimum absolute error); max (Parish average maximum absolute error).

Since the pollen model only estimates the whole DDR, the comparison was only made considering that at the VER stage. According to the authors, both models can obtain very good results. In the study's time frame, the pollen model achieved a lower average error for the years 2016 and 2019 to 2021 and a worse result for the years 2017 and 2018. Considering the average differential for the pollen model, both models coincide in the years in terms of underestimation and overestimation yield. The worst performance could be attributed in 2017 due to early flowering and in 2018 due to significant variation in water stress as discussed by the authors, stating that additional parameters, such as disease occurrence, agronomic, and weather conditions after flowering are required (Cunha et al., 2003).

The model developed in the present study can deliver prediction at a sub-regional level. Implementing the pollen model would require a more comprehensive network of pollen traps with cost implications. Furthermore, developing a model estimating grape yield in kg/ha and not in wine production can be seen as an advantage for being more comprehensive for the different actors in the DDR and for other regions where the regulations are not so specific and focused on the Port wine.

The current model also performs well when referring to other pollen-based models (and without the limitations mentioned above) with estimations in line with the work of Cunha et al., (Cunha et al., 1999, 2003, 2015) for the DDR. The different studies for regional-scale applications identified in the authors' previous work (Barriguinha et al., 2021) have an overall average R^2 between 0,71 and 0,99. Cristofolini et al. (Cristofolini and Gottardini, 2000) determination of the pollen index between the days when 5 and 95 % of the season's total pollen concentration were found achieved very good results, similarly to the work of Besselat (Besselat, 1987). With a different approach, Gonzaléz et al. (González-Fernández et al., 2020) and Fernandez et al. (Fernandez-Gonzalez et al., 2011; Fernández-González et al., 2011, 2020) combined aerobiological, phenological, and meteorological data achieving equally accurate production estimations more than one or two months in advance.

Compared with other models based on vegetation indices applied to vineyard yield estimation at the regional level, the current model also performs well. Gouveia et al. (Gouveia et al., 2011) worked on multilinear regression models using Corine Land Cover, wine statistics, NDVI, and meteorological variables (monthly averages of maximum, minimum, and daily mean temperature and precipitation) to estimate yield with $0,62 < R^2 < 0,90$ in a simulated test environment. Using Satellite Pour l'Observation de la Terre (SPOT) ten-day synthesis vegetation product (S10) Cunha et al. (Cunha et al., 2010) based on a correlation matrix between the wine yield of a current year and the full set of 10-day synthesis NDVI also achieved good results ($0,73 < R^2 < 0,84$). Sun et al. (Sun et al., 2017) combined satellite-based NDVI from Landsat and MODIS with LAI obtained using a Li-Cor LAI-2000 instrument with good results, $0,66 < R < 0,83$ (for NDVI and Yield) and $0,66 < R < 0,83$

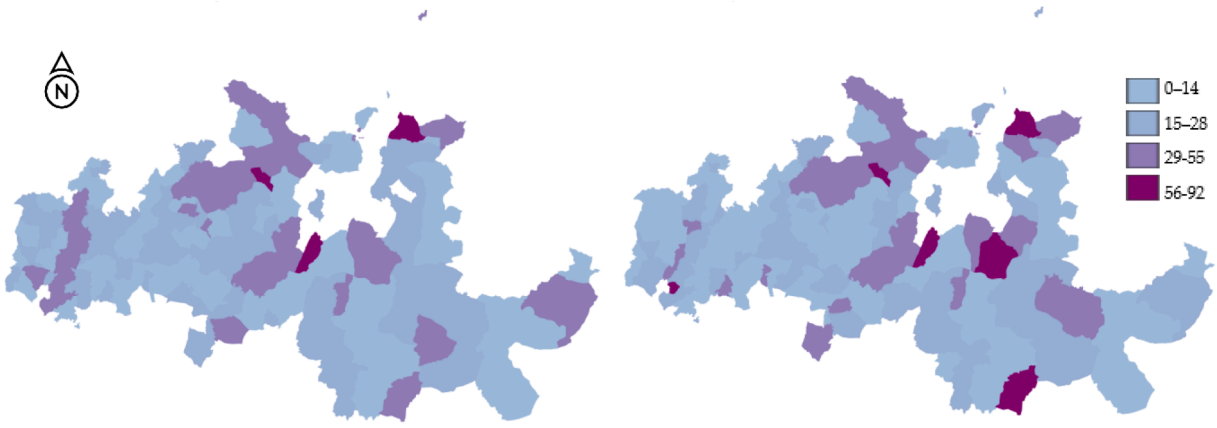


Fig. 7. Spatial distribution of the average absolute error (%) in each parish for the period 2016–2021 at FLO (left) and VER (right).

Table 13
Different methodological approaches for regional vineyard yield prediction (Barriguinha et al., 2021).

Reference	Methodological Approach	Data Sources	Test environment	Related Variables	Estimation
(Cristofolini and Gottardini, 2000)	Pollen Based	Hirst type sampler volumetric spore trap (Lanzoni VPPS-2000)	In-field	Airborne pollen concentration	$R^2 = 0,92$ (for grape production)
(Fernandez-Gonzalez et al., 2011)	Pollen Based	Aerobiological data (Lanzoni VPPS-2000 volumetric trap)	In-field	Meteorological and phytopathological variables	$R^2 = 0,98$ (for yield)
(Fernández-González et al., 2020)	Pollen Based	Pollen Hirst volumetric sampler and Cour passive trap	In-field	Airborne pollen concentration, weather data	$R^2 = 0,96$ (Cour); $R^2 = 0,99$ (Hirst)
(Besselat, 1987)	Pollen Based	Pollen concentration data	In-field	Airborne pollen concentration	$R^2 < 0,98$ (for yield)
(Cunha et al., 2015)	Pollen Based	Airborne pollen trap	Simulated	Airborne pollen concentration	$0,71 < R^2 < 0,86$ (for annual wine production)
(Cunha et al., 1999)	Pollen Based	Pollen concentration data	In-field	Airborne pollen concentration	$R^2 = 0,93$ (for yield)
(Cunha et al., 2003)	Pollen Based	One Cour Pollen Trap	In-field	Airborne pollen concentration	$0,66 < R^2 < 0,99$ (for wine production)
(González-Fernández et al., 2020)	Pollen Based	Aerobiological data (Lanzoni VPPS-2000 volumetric sampler), Meteorogical data	In-field	Airborne pollen concentration and Meteorologic data	$R^2 = 0,99$ (for yield)
(Gouveia et al., 2011)	Vegetation Indices	Corine Land Cover map, wine statistics, monthly means of climate variables and NDVI	Simulated	tmax, tmin, tavg, prec, NDVI	$0,62 < R < 0,90$ (for wine production)
(Sun et al., 2017)	Vegetation Indices	Satellite-based (NDVI) and (LAI)	In-field	NDVI, LAI	$0,66 < R < 0,83$ (for NDVI and Yield) and $0,66 < R < 0,83$ (for LAI and Yield)
(Cunha et al., 2010)	Vegetation Indices	Satellite data from vegetation (NDVI from SPOT)	In-field	NDVI	$0,73 < R^2 < 0,84$ (for yield)
(Cola et al., 2014)	Crop Simulation Model	Weather data and plant characteristics	Simulated/In-field Validation	Weather data and plant characteristics	$R^2 = 0,96$ (for yield in low-density canopies) $R^2 = 0,94$ (for yield in high-density canopies)
(Fraga et al., 2015)	Crop Simulation Model	Climate, soil, and management practices	Simulated/In-field Validation	Climate data, soil and terrain parameters, water stress indices, management practices	$R^2 = 0,86$ (for yield)
(Valdes-Gomez et al., 2009)	Crop Simulation Model	Phenology and harvest date, Soil water content, water stress, and grapevine growth and yield	Simulated/In-field Validation	Phenology and harvest date, soil water content, water stress, and grapevine growth and yield	$R^2 = 0,85$ (for yield)
(Sirsat et al., 2019)	Crop Simulation Model	Weather, yield, phenological dates, fertilizer information, soil analysis, and maturation index data	Simulated/In-field Validation	Weather, phenological dates, fertilizer information, soil analysis, and maturation index data	24,2 %<RRMSE < 28,6 %
(Fraga and Santos, 2017)	Other Models	Daily historic meteorological conditions, yield data	In-field	Temperature and Precipitation	$0,68 \leq r \leq 0,84$ (for grapevine production)
(Santos et al., 2020)	Other Models	Monthly mean air temperatures and monthly total precipitation data	In-field	Monthly mean air temperatures and monthly total precipitation	Wine production classes (1-low, 2-normal, 3-high): average estimation ratio of 79 % (calibration) 67 % (validation)

(for LAI and Yield) although the validation was made locally in a small area.

At the regional scale, crop simulation models are also an alternative for vineyard yield estimation (Barriguinha et al., 2021). This approach allows virtual experiments that can be made, for example, at specific phenological stages for testing hypotheses that could take years under real field conditions, with the added capability of integrating the findings in decision support systems (DSS). Cola et al. (Cola et al., 2014)

Table 14
Prediction for the DDR based on pollen model made at the VER stage (prediction data from ADVID reports (ADVID, 2016; ADVID, 2017; ADVID, 2018; ADVID, 2019; ADVID, 2020; ADVID, 2021).

Year	PROD	PRED (min)	DIF (min)	DIF (min) (%)	PRED (max)	DIF (max)	DIF (max) (%)	PRED (avg)	DIF (avg)	DIF (avg) (%)
2016	158,677	143,727	−14950	−9 %	158,318	−359	0 %	151,023	−7654	−5
2017	171,413	199,500	28,087	16 %	215,864	44,450	26 %	207,682	36,269	21
2018	145,282	190,636	45,355	31 %	204,955	59,673	41 %	197,795	52,514	36
2019	207,900	197,318	−10582	−5 %	216,273	8373	4 %	206,795	−1104	−1
2020	153,940	148,364	−5576	−4 %	168,136	14,197	9 %	158,250	4310	3
2021	195,802	176,455	−19347	−10 %	191,045	−4756	−2 %	183,750	−12052	−6

DDR (Douro Demarcated Region); PROD (wine grape production for the entire DDR in tons/year); PRED(min) (minimum estimated wine grape production in tons/year considering a conversion factor of 750 kg of grape for 550 L of wine); PRED(max) (maximum estimated wine grape production in tons/year considering a conversion factor of 750 kg of grape for 550 L of wine); PRED(avg) (average estimated wine grape production in tons/year considering a conversion factor of 750 kg of grape for 550 L of wine); DIF(min) (PRED(min)-PROD in kg/year); DIF(min)(%) (PRED(min)-PROD in %); DIF(max) (PRED(max)-PROD in kg/year); DIF(max)(%) (PRED(max)-PROD in %); DIF(avg) (PRED(avg)-PROD in kg/year); DIF(avg)(%) (PRED(avg)-PROD in %).

achieved good results simulating the fruit load based on light interception derived gross assimilation and thermal and water limitations with $R^2 = 0{,}96$ (for yield in low-density canopies) and $R^2 = 0{,}94$ (for yield in high-density canopies). Sirsat et al. (Sirsat et al., 2019) focused on grape yield predictive models for flowering, coloring, and harvest phenostages using ML techniques and climatic conditions, yield, phenological dates, fertilizer data, soil analysis, and maturation index data to construct the relational dataset. The authors identified dew point, relative humidity, and air temperature as the most favorable variables in building the model, with 24,2 %<RRMSE < 28,6 % for yield estimation. Fraga et al. (Fraga et al., 2015) and Valdes et a. (Valdes-Gomez et al., 2009) used a similar approach using STICS models with $R^2 = 0{,}86$ and $R^2 = 0{,}85$ respectively both with overestimation and underestimation, depending on the regions. In terms of performance, the current model can perform as well as the crop simulation alternatives. Those are much more complex as they are not limited to yield and simulate plant growth and development. They need to be calibrated and validated, requiring adaptability for new environments (distinct climate, soil, varieties, and management), making operationality and transferability difficult, complex, and costly in terms of time and biophysical data requirements (Sirsat et al., 2019).

The current model outperforms other models, such as the simple grape production model (PGP) based on favorable meteorological conditions, developed by Fraga et al. (Fraga and Santos, 2017), and the empirical model proposed by Santos et al. (Santos et al., 2020) where temperature and precipitation averaged over different periods, along with the anomalies of wine production in the previous five years, were used as predictors.

Models based on computer vision and image processing (by extraction of variables that can be related to the actual yield: number of berries, bunch/cluster area, leaf area, number of flowers, stems, and branches), trellis tension, laser, radar, and radio frequency data processing also constitute viable approaches for estimating vineyard yield. Nevertheless, those are not suitable for regional-scale implementation. Apart from the trellis tension approach, the real applicability under field conditions in commercial vineyards is not referenced for the most part (Barriguinha et al., 2021).

4. Conclusions

The use of LSTM neural network can be applied to vineyard yield prediction at the regional scale. It can perform as well as the other identified methodologies, outperforming some of them while dealing with some of the above-mentioned limitations. This and other ML-based methods can help study complex interactions between biotic and abiotic systems to understand and make predictions (Thessen, 2016), as in the current study.

The developed model allows for an early yield estimation with better results at the VER stage (one month before harvest start) when compared to the FLO stage (3 months before harvest start), with an absolute error for the whole study region between 8 % and 24 %, and between 6 % and 29 % for the sub-regions. The estimation range is much broader regarding estimations made at higher spatial resolution (parish level). Although 68 % of the parishes have an average error below 20 %, we consider that the model is not yet capable of predicting at that more detailed scale.

Despite the good results, the fact that there are no production data at the plot level (limiting the size of the dataset), a short time series of yield data, and a low number of weather stations (limiting the size and quality of the dataset considering the size and characteristics of the study area), are factors perceived as sources of error and limitations for the current model, and the reason for not being able to go further for larger scales. This is consistent with the limitations identified for this type of model applied to yield prediction (van Klompenburg et al., 2020).

Being a prediction model, this DL approach falls short of interpretability and, unlike more common inferential models, is a black-box model for making predictions (Emmert-Streib et al., 2020).

The variability and randomness of the yield and the different explanatory variables used between seasons, sub-regions, and parishes make the challenge of rapidly estimating yield very complex (Cunha et al., 2010). For the present study, we concluded that using NDVI alone is insufficient for a robust and accurate model developed with this methodology. As climatic variables have a strong correlation to yield (Badr et al., 2018; Ferrer et al., 2017), using satellite data and meteorological variables constitutes a better strategy for regional scale estimation of wine production (Gouveia et al., 2011). The performance is also very dependent on environmental conditions and management strategies (Sun et al., 2017), with yield correlated with an extensive list of climate, soil, and plant variables with high temporal and spatial heterogeneity. Also, the relation to quality is one of the biases that yield estimation needs to deal with, as the producer's management decision directly impacts quality and yield.

The integration of more specific multispectral based VI data, such as Leaf Area Index (LAI), or the use of Synthetic Aperture Radar (SAR) and Light Detection And Ranging (LIDAR) data can be tested as potential future developments in this field.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge: IVDP - Instituto dos Vinhos do Douro e do Porto, IP (Institute of Douro and Port Wines) (https://www.ivdp.pt/en), for providing historical data related to wine grape production for the entire DDR at the parish level; IPMA - Instituto Português do Mar e da Atmosfera, IP (Portuguese Institute for Sea and Atmosphere) (https://www.ipma.pt/en/), for providing historical climate data referring to the six climatic stations inserted in the study area; IVV - Instituto da Vinha e do Vinho, IP (Institute of Vine and Wine) (https://www.ivv.gov.pt/), for providing spatial data related to the vineyard areas located in the DDR.

References

ADVID, 2016, *Boletim - Ano Vitícola 2016 - Balanço Final*. 2016, Associação para o Desenvolvimento da Viticultura Duriense. p. 18.

ADVID, 2017, *Boletim 14-17 - Ano Vitícola 2017 - Balanço Final*. 2017, Associação para o Desenvolvimento da Viticultura Duriense. p. 22.

ADVID, *2018, Boletim Ano Vitícola 2018 - Balanço Final*. 2018, Associação para o Desenvolvimento da Viticultura Duriense. p. 33.

ADVID, *2019, Boletim Ano Vitícola 2019 - Balanço Final*. 2019, Associação para o Desenvolvimento da Viticultura Duriense. p. 25.

Anderson, J.D., Jones, Gregory V., Tait, Andrew, Hall, Andrew, Trought, Michael C.T., 2012. Analysis of viticulture region climate structure and suitability in New Zealand. OENO One 46 (3), 149.

Arab, Sara Tokhi, Noguchi, Ryozo, Matsushita, Shusuke, Ahamed, Tofael, 2021. Prediction of grape yields from time-series vegetation indices using satellite remote sensing and a machine-learning approach. Remote Sensing Applications: Society and Environment 22, 100485.

Badr, Golnaz, Hoogenboom, Mohammad, Moyer, Michelle, Keller, Markus, 2018. Analysis of several bioclimatic indices for viticultural zoning in the Pacific Northwest. Climate Research 76 (3), 203–223.

Barriguinha, André, de Castro Neto, Miguel, Gil, Artur, 2021. Vineyard Yield Estimation, Prediction, and Forecasting: A Systematic Literature Review. Agronomy 11 (9), 1789.

Besselat, B., 1987. Les prévisions de récolte en viticulture. OENO One 21 (1), 1.

Bindi, M., Fibbi, L., Gozzini, B., Orlandini, S., Miglietta, F., 1996. Modelling the impact of future climate scenarios on yield and yield variability of grapevine. Climate Research 7 (3), 213–224.

Birth, Gerald S., McVey, George R., 1968. Measuring the Color of Growing Turf with a Reflectance Spectrophotometer1. Agronomy Journal 60 (6), 640–643.

Bock, A., Sparks, T., Estrella, N., Menzel, A., 2011. Changes in the phenology and composition of wine from Franconia, Germany. Climate Research 50 (1), 69–81.

Boulton, Roger B., Singleton, V.L., Bisson, L.F., Kunkee, R.E., 1996. Principles and Practices of Winemaking. Springer, US.

Camps, Josep Odó, Ramos, María Concepción, 2012. Grape harvest and yield responses to inter-annual changes in temperature and precipitation in an area of north-east Spain with a Mediterranean climate. International Journal of Biometeorology 56 (5), 853–864.

Candiago, Sebastian, Tscholl, Simon, Bassani, Leonardo, Fraga, Helder, Egarter Vigl, Lukas, 2022. A geospatial inventory of regulatory information for wine protected designations of origin in Europe. Scientific Data 9 (1), 394.

Carrillo, E., Matese, A., Rousseau, J., Tisseyre, B., 2016. Use of multi-spectral airborne imagery to improve yield sampling in viticulture. Precision Agriculture 17 (1), 74–92.

ADVID, *2021, Boletim Ano Vitícola 2021 - Balanço Final do Ano Vitícola*. 2021, Associação para o Desenvolvimento da Viticultura Duriense. p. 28.

ADVID, *2020, Boletim Ano Vitícola. 2020 - Balanço Final*. Associação para o Desenvolvimento da Viticultura Duriense. p. 27.

Cristofolini, F. and E. Gottardini, *Concentration of airborne pollen of Vitis vinifera L. and yield forecast: A case study at S. Michele all'Adige, Trento, Italy*. Aerobiologia, 2000. **16** (1): p. 125-129.

Cola, Gabriele, Mariani, Luigi, Salinari, Francesca, Civardi, Silvia, Bernizzoni, Fabio, Gatti, Matteo, Poni, Stefano, 2014. Description and testing of a weather-based model for predicting phenology, canopy development and source-sink balance in Vitis vinifera L. cv. Barbera. Agricultural and Forest Meteorology 184, 117–136.

Cunha, Mário, Abreu, I., Pinto, Pedro, Castro, R., 2003. Airborne Pollen Samples for Early-Season Estimates of Wine Production in a Mediterranean Climate Area of Northern Portugal. American Journal of Enology and Viticulture 54, 189–194.

Cunha, Mário, Costa, Paulo, Abreu, Ilva, Pinto, Pedro, Castro, Ravena, 1999. Early estimate of wine prodction by means of airborne pollen: demarcated region of Douro. Ciência e Técnica Vitivinicola 14, 45–54.

Cunha, Mário, Marcal, Andre R.S., Silva, Lisa, 2010. Very early prediction of wine yield based on satellite data from VEGETATION. International Journal of Remote Sensing 31 (12), 3125–3142.

Cunha, Mário, Ribeiro, H., Abreu, I., 2015. Pollen-based predictive modelling of wine production: application to an arid region. European Journal of Agronomy 73, 42–54.

Cunha, Mário, Richter, Christian, 2011. Measuring the impact of temperature changes on the wine production in the Douro Region using the short time fourier transform. International Journal of Biometeorology 56, 357–370.

D'agostino, Ralph B., 1970. Transformation to normality of the null distribution of g1. Biometrika 57 (3), 679–681.

De La Fuente, Mario, Linares, Rubén, Baeza, Pilar, Miranda, Carlos, Lissarrague, José Ramón, 2015. Comparison of different methods of grapevine yield prediction in the time window between fruitset and veraison. OENO One 49 (1), 27.

De la Fuente Lloreda, M., 2014. The relevance of the yield prediction methods in vineyard management. Le Bulletin de l'OIV ISSN 0029–7121 (87), 387–394.

Diago, Maria-Paz, Tardaguila, Javier, Aleixos, N., Millan, Borja, Prats Montalbán, José, Cubero, Sergio, Blasco, Jose, 2015. Assessment Of Cluster Yield Components By Image Analysis. Journal of the Science of Food and Agriculture 95.

Di Gennaro, S.F., Toscano, P., Cinat, P., Berton, A., Matese, A. A precision viticulture UAV-based approach for early yield prediction in vineyard. in Precision Agriculture 2019 - Papers Presented at the 12th European Conference on Precision Agriculture, ECPA 2019. 2019. Montpellier, France.

Eurostat. *Restructuring of vineyards in the EU continues* 2022 [cited 2022 9/10/2022]; Available from: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20220520-1.

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., Dehmer, M., 2020. An Introductory Review of Deep Learning for Prediction Models With Big Data. Front Artif Intell 3, 4.

Fernández-González, M., Escuredo, O., Rodríguez-Rajo, F.J., Aira, M.J., Jato, V., 2011. Prediction of grape production by grapevine cultivar Godello in north-west Spain. Journal of Agricultural Science 149 (6), 725–736.

Fernández-González, M., Ribeiro, Helena, Piña-Rey, Alba, Abreu, Ilda, Rodríguez-Rajo, F. Javier, 2020. Integrating Phenological, Aerobiological and Weather Data to Study the Local and Regional Flowering Dynamics of Four Grapevine Cultivars. Agronomy 10 (2), 185.

Fernandez-Gonzalez, M., Rodriguez-Rajo, F.J., Jato, V., Escuredo, O., Aira, M.J., 2011. Estimation of yield 'Loureira' variety with an aerobiological and phenological model. Grana 50 (1), 63–72.

Ferrer, Milka, Echeverria, Gerardo, Miras-Avalos, Jose, 2017. Meteorological Conditions: Influence on Yield, Sanitary Status and Grape Composition. International Journal of Environmental & Agriculture Research (IJOEAR) 3.

Fraga, H., Costa, R., Moutinho-Pereira, J., Correia, C.M., Dinis, L.T., Gonçalves, I., Silvestre, J., Eiras-Dias, J., Malheiro, A.C., Santos, J.A., 2015. Modeling phenology, water status, and yield components of three Portuguese grapevines using the STICS crop model. American Journal of Enology and Viticulture 66 (4), 482–491.

Fraga, H., Malheiro, A., Moutinho Pereira, J., Pinto, J., Santos, J., 2013. *Future scenarios for viticultural zoning in Europe: ensemble projections and uncertainties*. International Journal of Biometeorology p.–2067.

Fraga, Helder, Santos, João, 2017. Daily prediction of seasonal grapevine production in the Douro wine region based on favourable meteorological conditions. Australian Journal of Grape and Wine Research 23 (2), 296–304.

Fraga, H., Santos, J.A., Moutinho-Pereira, J., Carlos, C., Silvestre, J., Eiras-Dias, J., Mota, T., Malheiro, A.C., 2016. Statistical modelling of grapevine phenology in Portuguese wine regions: observed trends and climate change projections. The Journal of Agricultural Science 154 (5), 795–811.

Giovos, Rigas, Tassopoulos, Dimitrios, Kalivas, Dionissios, Lougkos, Nestor, Priovolou, Anastasia, 2021. Remote Sensing Vegetation Indices in Viticulture: A Critical Review. Agriculture 11 (5), 457.

González-Fernández, Estefanía, Piña-Rey, Alba, Fernández-González, María, Aira, María J., Rodríguez-Rajo, F. Javier, 2020. Prediction of Grapevine Yield Based on Reproductive Variables and the Influence of Meteorological Conditions. Agronomy 10 (5), 714.

Gouveia, C., Liberato, M.L.R., DaCamara, C.C., Trigo, R.M., A. Ramos, M., 2011. Modelling past and future wine production in the Portuguese Douro Valley. Climate Research 48 (2–3), 349–362.

Guilpart, Nicolas, Metay, Aurélie, Gary, Christian, 2014. Grapevine bud fertility and number of berries per bunch are determined by water and nitrogen stress around flowering in the previous year. European Journal of Agronomy 54, 9–20.

Hochreiter, Sepp, Schmidhuber, Jürgen, 1997. Long Short-Term Memory. Neural Computation 9 (8), 1735–1780.

Jones, Gregory, Davis, Robert, 2000. Climate Influences on Grapevine Phenology, Grape Composition, and Wine Production and Quality for Bordeaux, France. American Journal of Enology and Viticulture 51, 249–261.

Henrich, V., Götze, Christian, Jung, András, Sandow, Christopher, Thürkow, Detlef, Cornelia, Glaesser, Development of an online indices database: Motivation, concept and implementation, in 6th EARSeL Imaging Spectroscopy SIG Workshop Innovative Tool for Scientific and Commercial Environment Applications. 2009: Tel Aviv, Israel.

Maimaitiyiming, Matthew, Sagan, Vasit, Sidike, Paheding, Kwasniewski, Misha T., 2019. Dual Activation Function-Based Extreme Learning Machine (ELM) for Estimating Grapevine Berry Yield and Quality. Remote Sensing 11 (7), 740.

Matese, V., Di Gennaro, Salvatore Filippo, 2021. Beyond the traditional NDVI index as a key factor to mainstream the use of UAV in precision viticulture. Scientific Reports 11 (1).

Murali, Krishna, Gumma, Murali, Kadiyala, Dakshina Murthy, Panjala, Pranay, Ray, Shibendu, Venkata, Akuraju, Radha, Dubey, Sunil, Smith, Andrew, Das, Rajesh, Whitbread, Anthony, 2021. Assimilation of Remote Sensing Data into Crop Growth Model for Yield Estimation: A Case Study from India. Journal of the Indian Society of Remote Sensing.

Kingma, Diederik, Ba, Jimmy, 2014. *Adam: A Method for Stochastic Optimization.* International Conference on Learning Representations.

K. Newlands, N., *Artificial Intelligence and Big Data Analytics in Vineyards: A Review*. 2022, IntechOpen.

Muruganantham, Priyanga, Wibowo, Santoso, Grandhi, Srimannarayana, Samrat, Nahidul Hoque, Islam, Nahina, 2022. A Systematic Literature Review on

Crop Yield Prediction with Deep Learning and Remote Sensing. Remote Sensing 14 (9), 1990.
Nuske, S., Gupta, K., Narasimhan, S., Singh, S., 2014. Automated Visual Yield Estimation in Vineyards. Journal of Field Robotics 31 (6), 996.
Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., Singh, S., 2014. In: *Modeling and calibrating visual yield estimates in vineyards*, in *Springer Tracts in Advanced Robotics*. Springer:, Berlin/Heidelberg, Germany, pp. 343–356.
Parker, Amber K., Fourie, Jaco, Trought, Mike C.T., Phalawatta, Kapila, Meenken, Esther, Eyharts, Anne, Moltchanova, Elena, 2022. Evaluating sources of variability in inflorescence number, flower number and the progression of flowering in Sauvignon blanc using a Bayesian modelling framework. OENO One 56 (1), 1–15.
Pelta, Ran, Beeri, Ofer, Tarshish, Rom, Shilo, Tal, 2022. Sentinel-1 to NDVI for Agricultural Fields Using Hyperlocal Dynamic Machine Learning Approach. Remote Sensing 14 (11), 2600.
Santesteban, Luis G., Royo, J. Bernardo, 2006. Water status, leaf area and fruit load influence on berry weight and sugar accumulation of cv. 'Tempranillo' under semiarid conditions. Scientia Horticulturae 109 (1), 60–65.
Santos, João A., Ceglar, Andrej, Toreti, Andrea, Prodhomme, Chloe, 2020. Performance of seasonal forecasts of Douro and Port wine production. Agricultural and Forest Meteorology 291.
Santos, J., Malheiro, A., Pinto, J., Jones, G., 2012. Macroclimate and viticultural zoning in Europe: observed trends and atmospheric forcing. Climate Research 51 (1), 89–103.
Shapiro, S.S., Wilk, M.B., 1965. An Analysis of Variance Test for Normality (Complete Samples). Biometrika 52 (3/4), 591–611.
Sirsat, Manisha, Moreira, João, Ferreira, Carlos, Cunha, Mário, 2019. Machine Learning predictive model of grapevine yield based on agroclimatic patterns. Engineering in Agriculture, Environment and Food 12.
Sishodia, Rajendra, Ray, Ram, Singh, Sudhir, 2020. Applications of Remote Sensing in Precision Agriculture: A Review. Remote Sensing 12, 3136.
Snevajs, Herman, Charvat, Karel, Onckelet, Vincent, Kvapil, Jiri, Zadrazil, Frantisek, Kubickova, Hana, Seidlova, Jana, Batrlova, Iva, 2022. Crop Detection Using Time Series of Sentinel-2 and Sentinel-1 and Existing Land Parcel Information Systems. Remote Sensing 14 (5), 1095.
Stamatiadis, Stamatis, Taskos, Dimitris, Tsadila, Eleftheria, Christofides, Calliopi, Tsadilas, Christos, Schepers, James S., 2010. Comparison of passive and active canopy sensors for the estimation of vine biomass production. Precision Agriculture 11 (3), 306–315.
Rouse, J.W., Jr., Haas, R.H., Schell, J.A., Deering, D.W., *Monitoring Vegetation Systems in the Great Plains with Erts*, in *NASA Special Publication*. 1974. p. 309.
Sun, L., Gao, F., Anderson, M.C., Kustas, W.P., Alsina, M.M., Sanchez, L., Sams, B., McKee, L., Dulaney, W., White, W.A., Alfieri, J.G., Prueger, J.H., Melton, F., Post, K., 2017. Daily mapping of 30 m LAI and NDVI for grape yield prediction in California vineyards. Remote Sensing 9 (4).
Tardaguila, J., Diago, M.P., Millan, B., Blasco, J., Cubero, S., Aleixos, N., 2013. *Applications of Computer Vision Techniques in Viticulture to Assess Canopy Features, Cluster Morphology and Berry Size*, in *I International Workshop on Vineyard Mechanization and Grape and Wine Quality*. Int Soc Horticultural Science, Leuven, Belgium.
Taylor, J.A., Dresser, J.L., Hickey, C.C., Nuske, S.T., Bates, T.R., 2019. Considerations on spatial crop load mapping. Australian Journal of Grape and Wine Research 25 (2), 144–155.
Thessen, A., 2016. Adoption of Machine Learning Techniques in Ecology and Earth Science. One Ecosystem 1, e8621.
Ubalde, J.M., Sort, X., Poch, R.M., Porta, M., 2007. Influence of edapho-climatic factors on grape quality in Conca de Barbera vineyards (Catalonia, Spain). J. Int. Des Sci. De La Vigne Et Du Vin 41 (1), 33–41.
Valdes-Gomez, H., Celette, F., de Cortazar-Atauri, I.G., Jara-Rojas, F., Ortega-Farias, S., Gary, C., 2009. Modelling soil water content and grapevine growth and development with the stics crop-soil model under two different water management strategies. Journal International Des Sciences De La Vigne Et Du Vin 43 (1), 13–28.
van Klompenburg, Thomas, Kassahun, Ayalew, Catal, Cagatay, 2020. Crop yield prediction using machine learning: A systematic literature review. Comput. Electron. Agric. 177, 105709.
Victorino, Gonçalo, Braga, Ricardo, Santos-Victor, José, Lopes, Carlos, 2022. Comparing a New Non-Invasive Vineyard Yield Estimation Approach Based on Image Analysis with Manual Sample-Based Methods. Agronomy 12, 1464.
Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sensing of Environment 236, 111402.
Xue, Jinru, Su, Baofeng, 2017. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. J. Sensors 2017, 1–17.
Zabawa, L., Kicherer, A., Klingbeil, L., Milioto, A., Topfer, R., Kuhlmann, H., Roscher, R., 2019. Detection of single grapevine berries in images using fully convolutional neural networks. in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2019. Long Beach, CA, USA.
Zhong, Liheng, Hu, Lina, Zhou, Hang, 2019. Deep learning based multi-temporal crop classification. Remote Sensing of Environment 221, 430–443.