

## CLIMATIC EFFECTS ON ARGENTINA'S TRADITIONAL DRINK – A MACHINE LEARNING BASED YERBA MATE PRODUCTIVITY PREDICTION

Hannah Kemper<sup>a\*</sup>, Sebastian Schmidt<sup>b</sup>, Theresa Keller<sup>c</sup>

a: GGS Solutions, Speyer, Germany – hannah.kemper@ggs-speyer.de;

b: Department of Geoinformatics – Z\_GIS, University of Salzburg, Austria – sebastian.schmidt@plus.ac.at;

c: Federal Institute for Hydrology, Koblenz, Germany – theresa.keller@bafg.de

**KEY WORDS:** agricultural production, machine learning, regression analysis, climate indices, climate change, El Niño Southern Oscillation

### ABSTRACT:

Yerba mate belongs to the most important agricultural products of Argentina. Due to climate change, together with El Niño and droughts, yields are negatively affected. Drought propagation from precipitation deficits to plant water stress, its impacts and especially its predictability are becoming an emerging field of research. This paper explores how the future yerba mate production in the states of Misiones and Corrientes, north-eastern Argentina, can be projected by using climate variables and satellite data. The applied methodology focuses on a machine learning approach based on multiple linear regression and random forest regression. The results indicate a significant relationship between yerba mate and the NDVI as well as the SOI. The highest yerba mate productivity is expected during weak La Niña events. The methodology of this analysis can successfully predict mate productivity based on satellite and climate data and can also easily be used for further research areas.

### 1. INTRODUCTION

Climate change and enforced effects of global teleconnections like El Niño are severely impacting traditional agriculture. In particular, the increasingly frequent occurrence of droughts has a major impact on agricultural productivity. In general, a drought describes a condition in which a precipitation deficit over a certain period of time causes moisture and water availability below average (Mishra & Singh 2010). The term is characterised by diverse definitions, which in turn are fundamental for the overall understanding of the context.

A *meteorological drought* occurs along with a precipitation deficit over a period of time. As it is based on the long-term mean relationship between evapotranspiration and precipitation, meteorological droughts can occur at any time of the year (Lloyd-Hughes & Saunders 2002). If the lack of precipitation persists for several weeks or months, soil moisture and infiltration decrease and consequently limit plant growth (Tallaksen & van Lanen 2004). In this case, an *agricultural drought* occurs (Mishra & Singh 2010). As soon as hydrological systems experience water shortages (e.g. unusually low groundwater recharge or river streamflow), the agricultural drought turns into a *hydrological drought* (Tallaksen & van Lanen 2004). If a drought has economic and social consequences, such as health, it is referred to as a *socio-economic drought* (Apurv et al. 2017). The whole process is also known as drought propagation (van Loon et al. 2012). When this paper discusses drought, we only refer to meteorological and agricultural drought.

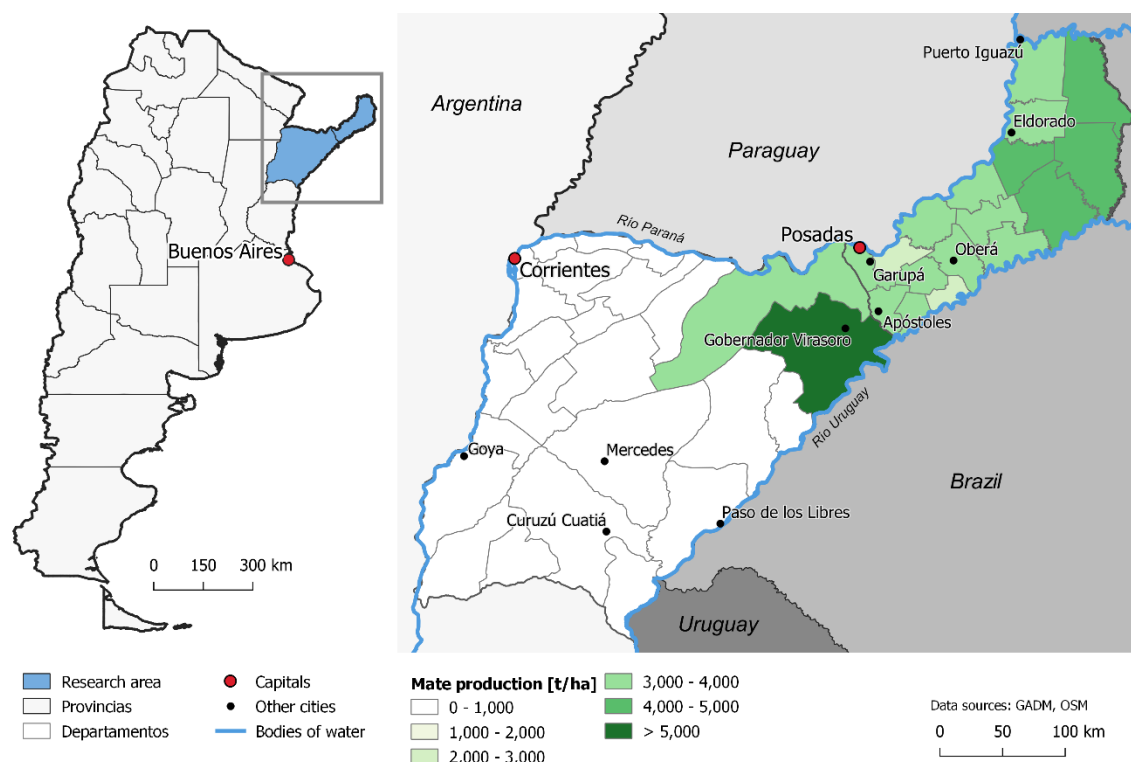
As a quantitative assessment of a meteorological drought occurrence, its characteristics and propagation, drought indices like the Standard Precipitation Index (SPI) are a common tool (Salimi et al. 2021). In addition, analysing a drought

through indicators is increasingly linked with research on drought impacts to deduce useful threshold values for drought prediction. As agricultural droughts are mostly expressed through crop production failures, forest fires or livestock sales, such variables are often correlated with drought indices (Stagge et al. 2015). In particular, the prediction of future crop production is one of the central components of agricultural research. As in many other areas of research (e.g. remote sensing), newer approaches use a diverse range of machine learning (ML) algorithms to predict future developments of drought indices (e.g. Byakatonda et al. 2019). Furthermore, the ML-based prediction of drought impacts on crop production is developed by several scholars, including Potop et al. (2012) or Kogan et al. (2019).

Climate change and global warming are accompanied by a multitude of global and regional impacts. Accelerated hydrological processes, caused by rising temperatures and an increased moisture demand of the atmosphere, and the absence of precipitation will contribute to a higher probability of severe future drought events (Li et al. 2009; Mukherjee et al. 2018). Predicting future climatic conditions is very difficult, as current climatic phenomena like El Niño are not completely understood yet.

The El Niño Southern Oscillation (ENSO) is a large-scale ocean-atmosphere system associated with ocean current and temperature changes (Morid et al. 2007). To assess the intensity of ENSO, the Southern Oscillation Index (SOI) is used, which measures the difference in air pressure between Tahiti and Darwin (Australia). Negative SOI values indicate an El Niño event, while positive SOI values are indicative of a La Niña event. These air pressure differences can lead to far-reaching effects that affect large parts of the Earth, so-called teleconnections. For Argentina, an El Niño event is usually synonymous with high precipitation (Ulla 2016). However, the effects of ENSO differ greatly between regions, as do the impacts of climate change.

\* corresponding author



**Figure 1.** Research area. The departments where yerba mate is cultivated are colour-coded based on production in t/ha.

For north-eastern Argentina, the study area of this paper, climate change models mostly predict reducing rainfall amounts in winter and spring. In summer and fall, however, increasing precipitation is more likely. This increase can probably be attributed to more frequent extreme events and is overall characterised through a higher interannual precipitation variability (Cabr   et al. 2016; Marengo et al. 2010). Studies such as Thomasz et al. (2018) show different impacts of increasing drought occurrence in Argentina in response to climate change. Portela et al. (2015) examine historic drought events in the border zone of Brazil, Paraguay and Argentina and show drought phases between 1960 and the 1970s as well as since the mid-1980s. Historic droughts occurred in the research area in the years 1977/78, 1981, 1988, 1999/00, 2006, 2012/13 and 2020/21 (Vargas et al. 2011; Global Drought Observatory 2022).

Yerba mate (*Ilex paraguariensis*) is a tree that originates from southern South America (between 21  S and 30  S and 48  30'W and 56  10'W). The plant needs regular precipitation, a minimum annual rainfall of 1200mm and high soil humidity. It is less sensitive to temperatures but grows best at an average temperature of 21  C (Croge et al. 2021; Heck & Mejia 2007). Its leaves were already consumed in pre-Colombian times by indigenous peoples (e.g. Guaran  ) as an infusion, which is nowadays very popular in Argentina, Uruguay (*mate*), Brazil (*chimarr  o*) and Paraguay (*terer  *). In addition, the plant is also used for e.g. energy drinks, cosmetics and textiles (Croge et al. 2021). In Argentina, it is mainly cultivated on plantations in the two states of Misiones and Corrientes since the early 20th century (Heck & Mejia 2007). In 2019, approximately 945,962 tonnes of yerba mate were produced worldwide, about 31.9% of them in Argentina (FAO 2021). The harvesting season in Misiones lasts roughly from April to September (Butiuk et al. 2016). Per

capita consumption of yerba mate in Argentina is approximately 5 kg per year, only surpassed by Uruguay. Similar to some Asian beverages, mate tea has a strong social component, as it is usually shared (Bracesco et al. 2011). Furthermore, mate tea has numerous positive health properties. For instance, it might prevent some cancer types and is anti-inflammatory (Souza et al. 2015).

Most studies on drought impacts focus on staple crops such as grains. Less often, however, plants that have social rather than nutritional benefits are considered. For this reason, we are trying to find out with our research whether it is possible to make statistically significant statements about mate productivity in north-eastern Argentina on the basis of climate data and drought indices. Consequently, this paper tries to answer the following research question:

*How can yerba mate productivity in Argentina be predicted reliably based on satellite and climate data?*

To answer this question, we use multiple linear regression and random forest regression. We are not aware of any study that investigates mate productivity based on satellite data or drought indices. We published our data and script on GitHub (<https://github.com/HannahKemper/YerbaMate.git>).

## 2. METHODOLOGY

### 2.1 Research area

The states (*provincias*) of Misiones and Corrientes are located in the extreme northeast of Argentina, enclosed by the neighbouring countries Paraguay, Brazil and Uruguay. Fig-

ure 1 shows a map of the research area. In the national context, both Misiones and Corrientes are comparatively small and sparsely populated states (INDEC 2022). The landscape consists mainly of forests, savannas and grasslands (Velazco et al. 2018). Some important bodies of water include the Iberá wetlands, the Iguazú waterfalls and the Río Paraná, Argentina's longest river. In 2021, the latter was featured internationally in the press due to its low water level (BBC 2021).

The climates of Misiones and Corrientes are characterised by subtropical, humid conditions, which belong to the category Cfa according to Köppen and Geiger. While the average annual rainfall is around 1800mm, the mean annual temperature is at 21°C. In the summertime, the maximum daily air temperature can rise up to 39°C, whereas the coldest days in July can reach -6°C (Eibl et al. 2000). ENSO can influence weather patterns and agricultural yields in the region due to its ocean-atmospheric interactions. As ENSO events strongly correlate with precipitation, wet summer conditions in north-eastern Argentina are associated with El Niño (warm events) and dry summer phases with La Niña (cold events) (Portela et al. 2015, Podestá et al. 2002; Pol & Binyamin 2014). The latter has a higher influence in the research area, with a greater likelihood of remaining stable over a longer time (Rusticucci & Vargas 2002). Along with this, yields tend to be reduced during La Niña (Gutierrez 2017; Podestá et al. 2002).

## 2.2 Data

### 2.2.1 Data sources

In order to create a suitable database for the analysis, meteorological and remote sensing-based datasets were used. The research period extended from 1990 to 2020, limited by the agricultural data availability. Information on the yerba mate production was obtained from the open data portal of the Argentinian Ministry of Agriculture, Livestock and Fisheries (Ministerio de Agricultura, Ganadería y Pesca 2022). Information on temperature was downloaded via Google Earth Engine, using a dataset from Rodell et al. (2004).

The Standardized Precipitation Index (SPI) provides information on the development of the precipitation deficit through the water cycle. In combination with data on agricultural production, it can be used to make statements about the occurrence of agricultural droughts and emerging vulnerabilities (Barker et al. 2016; Hayes et al. 2011). Negative values indicate dry conditions, with moderately dry conditions classified at values from -1.0 to -1.49, severely dry conditions at values from -1.5 to -1.99, and extremely dry conditions at values smaller than -2.0 (Mori et al. 2007). The SPI was calculated for annual values on the basis of 30-year precipitation sums from the CHIRPS dataset (Funk et al. 2015), using the following formula:

$$SPI = \frac{p_{period} - \text{mean of } p_{period}}{sd \text{ of } p_{1990-2020}} \quad (1)$$

The Southern Oscillation Index (SOI) quantifies the strength of the El Niño phenomenon on a scale from -4 to +4. The index expresses the difference in the air pressure between two stations divided by the monthly standard deviation (see Formula 2). The values were obtained from the National Oceanic and Atmospheric Administration (NOAA 2022) and mean values were used in the analysis.

$$SOI = \frac{\text{Standard.Tahiti} - \text{Standard.Darwin}}{MSD} \quad (2)$$

The Normalized Difference Vegetation Index (NDVI) can be calculated due to the high absorption by vegetation in the red (R) spectrum and the simultaneous lack of absorption in the near-infrared (NIR) spectrum (Schmidt & Barron 2020; Pet-torelli et al. 2011). It is the most commonly used vegetation index (Guo et al. 2015) and expresses plant health using the following formula:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (3)$$

NDVI data was obtained using the Landsat 5 and 7 dataset in Google Earth Engine (Google Developers 2022).

All variables were then calculated for yearly mean values on the department (*departamento*) level using a Zonal Statistics tool in QGIS 3.12.

### 2.2.2 Exploratory Data Analysis

A first step in a statistical analysis is usually an exploratory data analysis. For this purpose, we calculated a correlation matrix, using Spearman's  $\rho$  to check for correlations between the explanatory variables (cf. Figure 2). High coefficients are usually a good indicator for multicollinearity, which is a common problem in regression analyses. In our data, there was only one variable combination with a correlation coefficient  $> 0.7$ , which is a common threshold (Dormann et al. 2013). The relation between precipitation and SPI showed a correlation coefficient of 0.94, which is sensible as the SPI is based on precipitation values. Consequently, precipitation was not considered as a separate variable in the analysis.



**Figure 2.** Correlation matrix. The Spearman correlation coefficient was used to calculate statistical similarity between the explanatory variables.

### 2.3 Regression analysis

The data cleaning and further analyses were conducted using Python 3.8. The data basis used in this study had to be cleansed by treating missing and outlying values, as these could affect the reliability and plausibility of relationships between variables. Missing values in yerba mate production were imputed using region-based mean values. Further outliers were detected and excluded using an interquartile range of 0.98 (Wong & Wang 2003). The values were normalised

using the standard scaler in scikit-learn (Pedregosa et al. 2011). The independent and dependent variables were tested for skewness and split into training and testing data. For this, the training data size was set to 70%, a common threshold (Dobbin & Simon 2011).

### 2.3.1 OLS Model

One of the most common and simplest linear regression models is the Ordinary Least Squares model (OLS), which can be described by the following formula:

$$y = \alpha_0 + X\beta + \epsilon \quad (4)$$

where  $y$  is the dependent variable, " $\alpha_0$ " is a  $n \times 1$  vector of ones associated with the constant term parameter  $\alpha$ ,  $X$  denotes an  $n \times K$  matrix of explanatory variables associated with the  $K \times 1$  parameter vector  $\beta$  and  $\epsilon$  is an error term (Halleck Vega und Elhorst 2015).

The explanatory variables were ordered according to theoretical relevance in order to add them in a stepwise approach. The defined order was *SPI*, *SOI*, *NDVI*, *temp*. The resulting specifications were then tested for their suitability using statistical measures, such as the Akaike information criterion (AIC), Bayesian information criterion (BIC) and the coefficient of determination ( $R^2$ ). Additionally, the variance inflation factor (VIF) was calculated to check for multicollinearity.

### 2.3.2 Random Forest Regression

A common ML algorithm applied for the prediction of numerical values is the random forest regression (RF). The RF is a so-called ensemble of regression trees where the predictions of several decision trees are averaged (Strobl et al. 2009). Using a randomised search cross validation, numerous models were calculated, varying parameters including the number of trees, the depth of trees, the minimum samples per split and the minimum samples per leaf. The result of the validation was the best-performing parameter combination for the data basis (Koehrsen 2018). As the model performance generally varies with sample size, we performed an additional k-fold cross validation ( $n\_splits$ : 10) in order to evaluate the average performance of our models.

As a measure of model quality, several indicators were used. The Mean Squared Error (MSE) and Mean Absolute Error (MAE) are among common quality measures (Bachmair et al. 2016). Further, the percentage of correctly predicted values and the relative importance of the variables in the database were considered for the evaluation of the models. For the RF, we used the normalised variables.

The calculated RF model was then used to predict yerba mate productivity using different values for the explanatory variables. A value range with 0.1 steps was defined for all variables, stretching between the minimum and maximum values of the respective variable, with one standard deviation added to each. Predictions were made based on all possible combinations.

## 3. RESULTS

### 3.1 OLS Model

Different specifications for a simple OLS model were calculated. The results can be found in Table 1. Based on the AIC and BIC, model 3 seemed to be the most appropriate for the data basis. The specification that added temperature had a practically identical AIC, but the added variable was not statistically significant. For both of these models, both SOI and NDVI had positive correlation coefficients and were statistically significant ( $p \leq 0.05$ ). The SPI, on the other hand, was not statistically significant in any of the model specifications, which was also reflected by its strongly fluctuating coefficients.

The VIF was well below the commonly applied threshold of 10 for each variable, implying low multicollinearity. The condition number, a similar measure, also remained below the popular threshold of 30 (Lever et al. 2016).

	model 1	model 2	model 3	model 4
SPI	-15.593 (0.6261)	12.649 (0.6969)	-0.702 (0.9572)	0.037 (0.9977)
SOI	-	177.849 (0.0017)	101.109 (1e-05)	102.143 (1e-05)
NDVI	-	-	902.450 (2e-74)	704.289 (2e-05)
temp	-	-	-	39.652 (0.2164)
AIC	3649.35	3641.37	3304.27	3304.70
BIC	3652.57	3647.82	3313.95	3317.61
$R^2$	0.0013	0.0534	0.8471	0.8484
Adjusted $R^2$	-0.0041	0.0431	0.8446	0.8451
Condition number	1	1.86	2.39	29.44

**Table 1.** Results of OLS models. The p-values are included in brackets after the correlation coefficients. The second part of the table shows statistical measures. For easier legibility, the values of the indices were multiplied by 10.

### 3.2 Random Forest Regression

The randomised search cross validation revealed that a model with 200 estimators was the most accurate. Details on the parameters and performance can be obtained from Table 2. In both models, the order of importance of the variables was the same, with the respective values differing slightly. The SOI was the most important variable within the models. The average accuracy, i.e. the number of correctly classified data points, of the RF models was 71.9.

We predicted yerba mate productivity for different variable values. For the highest predictions, the SOI values were always in the low negative range, i.e. light La Niña events. At the same time, they all had high NDVI values. The SPI values, on the other hand, did not show such a clear trend. For the very highest predictions, however, the SPI values were negative, although the range of values fluctuated strongly. The highest prediction was made for a SPI of -0.3, a NDVI of 0.6 and a SOI of -0.2 (cf. Table 3).

Model	Accuracy	MSE	MAE	Variable importance
Default model	70.97	0.5208	0.5310	SOI: 0.38 SPI: 0.34 NDVI: 0.28
<i>n_estimators: 100, min_samples_split: 2, min_samples_leaf: 1, max_features: auto, max_depth: None, bootstrap: True</i>				
Optimised model	71.92	0.5393	0.5485	SOI: 0.36 SPI: 0.32 NDVI: 0.31
<i>n_estimators: 200, min_samples_split: 2, min_samples_leaf: 1, max_features: sqrt, max_depth: 50, bootstrap: True</i>				

**Table 2.** Results of Random Forest Regression. The default, baseline model is listed, as well as the best model according to the Randomised Search Cross validation.

In the case of the lowest predictions, the statements are not quite so clear. Only the NDVI values were consistently very low. For the other two variables, there were very strongly fluctuating values, although these were mostly negative. The lowest prediction was made for a SPI of -1.2, a NDVI of 0.4 and a SOI of -0.6. Interestingly, these were not the respective extreme values for the variables.

SPI	NDVI	SOI	Predicted productivity
-0,3	0,6	-0,2	6104,22
-0,3	0,7	-0,2	6104,22
-0,5	0,6	-0,2	6101,23
-0,5	0,7	-0,2	6101,23
-1,1	0,6	-0,2	6097,68
-1,0	0,4	-0,6	1891,20
-0,9	0,4	-0,6	1884,86
-1,1	0,4	-0,6	1872,94
-0,8	0,4	-0,6	1864,16
-1,2	0,4	-0,6	1858,89

**Table 3.** Prediction of yerba mate productivity. The top 5 highest and lowest predictions are shown in the table with the respective values for each variable.

#### 4. DISCUSSION

According to the results of the OLS model, there is a statistically significant relationship between yerba mate productivity and the SOI and NDVI. If the NDVI increased by 0.1, the productivity would also rise by 902 kg/ha. The relationship with the SOI was also positive. If the SOI increased by 0.1, the productivity would rise by 101 kg/ha. The interpretation of this value is not quite simple, which is mainly due to the nature of the linear model. It is doubtful that a strong La Niña event delivers higher yields than under normal conditions. However, one could interpret that a drier event is better for mate cultivation than too wet conditions, i.e. El Niño. There may be a connection to the possible irrigation of the

plants, which was not controlled for in our model due to lack of data.

The RF performed well considering the difficulty of predicting numerical values ranging from 0 to 7,500 kg/ha. An accuracy value of over 70% was satisfying, given that the MAE values were smaller than 0.7.

Considering the predicted yerba mate productivity for the SPI, NDVI and SOI, a more complex image was given on the relationship between the variables. Predicting the expected production for certain SPI, NDVI and SOI values was helpful to understand the model's interpretation of the data. The great importance of the SOI combined with the fact that this variable can be predicted up to six months in advance could be relevant for future research.

The RF results indicated that the maximum yerba mate production is to be expected during weak La Niña events. El Niño events, which tend to be weak in north-eastern Argentina, on the other hand, were estimated to provide rather low yerba mate production. The relationship between vegetation vitality and productivity also appeared clear, with high NDVI values being related to high predicted production. For the SPI, the correlations were not quite so clear, as the values fluctuated strongly. However, negative SPI values, which represent drier phases, were most likely to be present in high forecasts. However, similar SPI values were also associated with very low predictions. The interpretation of this result was not entirely clear. It may even indicate a threshold value at which mate productivity tips.

The methodology presented here should only be understood as a first approach to the topic, as the analysis relied only on very basic statistical methods. Other, non-linear types of regression models or ML algorithms could provide more insightful statements in future research on this topic. For this, data with a higher spatial resolution or an explicit geo-reference with regard to the cultivation areas would probably be advantageous.

Several shortcomings of the methodology should be noted. As it was already mentioned, we were not able to control for irrigation practises on yerba mate plantations due to missing data. Furthermore, the extension of our research area to other regions with yerba mate plantations could be sensible. Particularly, an extension to the south of Brazil, the largest producer of yerba mate (FAO 2021), could enhance the quality of the models. Since the employed variables, in particular the SOI, were not geographically explicit and not restricted to the plantations, a more pixel-based approach could be adequate for future studies. Then, information on soil types or topography could also be considered. In order to make a more accurate statement regarding droughts, one could also use a logistic regression which is based on a binary classification of the dependent variable. This was tested for our data, but it did not yield any significant results and was therefore discarded. However, it could be considered sensible for another case study.

## 5. CONCLUSION

According to the regression analysis performed in this paper, the relationship between the yerba mate productivity and the SOI and NDVI are both of high significance in north-eastern Argentina. Future monitoring of SOI values could be of importance for the prediction of the Argentinian yerba mate production, as the highest productivity values were predicted for weak La Niña events. Even though there were some limitations, the chosen methodology offers great potential for further use in other geographical regions, considering other crops and the integration of additional remote sensing information.

## REFERENCES

- Apurv, Tushar; Sivapalan, Murugesu; Cai, Ximing (2017): Understanding the Role of Climate Characteristics in Drought Propagation. In: *Water Resour. Res.* 53 (11), pp. 9304–9329. DOI: 10.1002/2017WR021445.
- Bachmair, Sophie; Stahl, Kerstin; Collins, Kevin; Hannaford, Jamie; Acreman, Mike; Svoboda, Mark et al. (2016): Drought indicators revisited: the need for a wider consideration of environment and society. In: *WIREs Water* 3 (4), pp. 516–536. DOI: 10.1002/wat2.1154.
- Barker, L. J.; Hannaford, J.; Chiverton, A.; Svensson, C. (2016): From meteorological to hydrological drought using standardised indicators. In: *Hydrol. Earth Syst. Sci.*, pp. 2483–2505. DOI: 10.5194/hessd-12-12827-2015.
- BBC (2021): South America's drought-hit Paraná river at 77-year low. In: *BBC News*, 09.01.2021. <https://www.bbc.com/news/world-latin-america-58408791> (03.01.2022).
- Bracesco, N.; Sanchez, A. G.; Contreras, V.; Menini, T.; Gugliucci, A. (2011): Recent advances on *Ilex paraguariensis* research: minireview. In: *Journal of ethnopharmacology* 136 (3), pp. 378–384. DOI: 10.1016/j.jep.2010.06.032.
- Butiuk, Ana P.; Martos, María A.; Adachi, Osao; Hours, Roque A. (2016): Study of the chlorogenic acid content in yerba mate (*Ilex paraguariensis* St. Hil.): Effect of plant fraction, processing step and harvesting season. In: *Journal of Applied Research on Medicinal and Aromatic Plants* 3 (1), pp. 27–33. DOI: 10.1016/j.jarmap.2015.12.003.
- Byakatonda, Jimmy; Parida, B. P.; Kenabatho, Piet K.; Molafhi, D. B. (2019): Prediction of onset and cessation of austral summer rainfall and dry spell frequency analysis in semi-arid Botswana. In: *Theor Appl Climatol* 135 (1-2), pp. 101–117. DOI: 10.1007/s00704-017-2358-4.
- Cabré, María Fernanda; Solman, Silvino; Núñez, Mario (2016): Regional climate change scenarios over southern South America for future climate (2080-2099) using the MM5 Model. Mean, interannual variability and uncertainties. In: *ATM*. DOI: 10.20937/ATM.2016.29.01.04.
- Croge, Camila Pereira; Cuquel, Francine Lorena; Pintro, Paula Toshimi Matumoto (2021): Yerba mate: cultivation

## ACKNOWLEDGEMENTS

We want to thank our reviewers for their helpful input. Further, we would like to acknowledge Melina Hofmann for her tremendous support.

systems, processing and chemical composition. A review. In: *Sci. agric. (Piracicaba, Braz.)* 78 (5). DOI: 10.1590/1678-992x-2019-0259.

Dobbin, K.; Simon, R. (2011): Optimally splitting cases for training and testing high dimensional classifiers. In: *BMC Medical Genomics* 4 (31), pp. 1–8. DOI: 10.5194/gi-2016-11-RC2.

Dormann, Carsten F.; Elith, Jane; Bacher, Sven; Buchmann, Carsten; Carl, Gudrun; Carré, Gabriel et al. (2013): Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. In: *Ecography* 36 (1), pp. 27–46. DOI: 10.1111/j.1600-0587.2012.07348.x.

Eibl, B.; Fernandez, R. A.; Kozarik, J. C.; Lupi, A.; Montagnini, F.; Nozzi, D. (2000): Agroforestry systems with *Ilex paraguariensis* (American holly or yerba mate) and native timber trees on small farms in Misiones, Argentina. In: *Agroforestry Systems* 48, pp. 1–8.

FAO (2021): FAOSTAT. <https://www.fao.org/faostat/en/#data> (03.01.2022).

Funk, Chris; Peterson, Pete; Landsfeld, Martin; Pedreros, Diego; Verdin, James; Shukla, Shraddhanand et al. (2015): The climate hazards infrared precipitation with stations--a new environmental record for monitoring extremes. In: *Scientific data* 2, pp. 150066. DOI: 10.1038/sdata.2015.66.

Global Drought Observatory (2022): Drought Impact Report. Hg. v. JRC. <https://edo.jrc.ec.europa.eu/gdo/php/index.php?id=2022&lon=-54.410087269512&lat=-26.547968157833&refDate=2021-07-21&ndx=RDrl> (03.01.2022).

Google Developers (2022): Landsat Collections in Earth Engine. <https://developers.google.com/earth-engine/datasets/catalog/landsat> (03.01.2022).

Guo, Wei; Lu, Dengsheng; Wu, Yanlan; Zhang, Jixian (2015): Mapping Impervious Surface Distribution with Integration of SNNP VIIRS-DNB and MODIS NDVI Data. In: *Remote Sensing* 7 (9), pp. 12459–12477. DOI: 10.3390/rs70912459.

Gutierrez, Luciano (2017): Impacts of El Niño-Southern Oscillation on the wheat market: A global dynamic analysis. In:

- PloS one 12 (6), e0179086. DOI: 10.1371/journal.pone.0179086.
- Halleck Vega, Solmaria; Elhorst, J. Paul (2015): THE SLX MODEL. In: JOURNAL OF REGIONAL SCIENCE 55 (3), pp. 339–363. DOI: 10.1111/jors.12188.
- Hayes, Michael; Svoboda, Mark; Wall, Nicole; Widhalm, Melissa (2011): The Lincoln Declaration on drought indices. Universal Meteorological Drought Index Recommended. In: Bulletin of the American Meteorological Society 92 (4), pp. 485–488. <http://www.jstor.org/stable/26226865>.
- Heck, C. I.; Mejia, E. G. de (2007): Yerba Mate Tea (*Ilex paraguariensis*): a comprehensive review on chemistry, health implications, and technological considerations. In: Journal of food science 72 (9), R138–51. DOI: 10.1111/j.1750-3841.2007.00535.x.
- INDEC (2022): Censo 2010. Instituto Nacional de Estadística y Censos de la República Argentina. <https://www.indec.gov.ar/indec/web/Nivel4-Tema-2-41-135> (03.01.2022).
- Koehrsen, Will (2018): Hyperparameter Tuning the Random Forest in Python - Towards Data Science. In: Towards Data Science, 01.10.2018. <https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74> (03.01.2022).
- Kogan, Felix; Guo, Wei; Yang, Wenze (2019): Drought and food security prediction from NOAA new generation of operational satellites. In: Geomatics, Natural Hazards and Risk 10 (1), pp. 651–666. DOI: 10.1080/19475705.2018.1541257.
- Lever, Jake; Krzywinski, Martin; Altman, Naomi (2016): Logistic regression. In: Nat Methods 13 (7), pp. 541–542. DOI: 10.1038/nmeth.3904.
- Li, Y.; Ye, W.; Wang, M.; Yan, X. (2009): Climate change and drought: a risk assessment of crop-yield impacts. In: Clim. Res. 39, pp. 31–46. DOI: 10.3354/cr00797.
- Lloyd-Hughes, Benjamin; Saunders, Mark A. (2002): A drought climatology for Europe. In: Int. J. Climatol. 22 (13), pp. 1571–1592. DOI: 10.1002/joc.846.
- Marengo, Jose A.; Ambrizzi, Tercio; Da Rocha, Rosmeri P.; Alves, Lincoln M.; Cuadra, Santiago V.; Valverde, Maria C. et al. (2010): Future change of climate in South America in the late twenty-first century: intercomparison of scenarios from three regional climate models. In: Clim Dyn 35 (6), pp. 1073–1097. DOI: 10.1007/s00382-009-0721-6.
- Ministerio de Agricultura, Ganadería y Pesca (2022): Estimaciones Agrícolas. <https://datosestimaciones.magyp.gob.ar/> (03.01.2022).
- Mishra, Ashok K.; Singh, Vijay P. (2010): A review of drought concepts. In: Journal of Hydrology 391 (1–2), pp. 202–216. DOI: 10.1016/j.jhydrol.2010.07.012.
- Morid, Saeid; Smakhtin, Vladimir; Bagherzadeh, K. (2007): Drought forecasting using artificial neural networks and time series of drought indices. In: Int. J. Climatol. 27 (15), pp. 2103–2111. DOI: 10.1002/joc.1498.
- Mukherjee, Sourav; Mishra, Ashok; Trenberth, Kevin E. (2018): Climate Change and Drought: a Perspective on Drought Indices. In: Curr Clim Change Rep 4 (2), pp. 145–163. DOI: 10.1007/s40641-018-0098-x.
- NOAA (2022): Southern Oscillation Index (SOI). <https://www.ncdc.noaa.gov/teleconnections/enso/soi> (03.01.2022).
- Pedregosa, F.; Varoquaux, G.; Vincent, M.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M. & É. Duchesnay (2011): Scikit-learn: Machine learning in Python. In: Journal of Machine Learning Research 12, p. 2825–2830.
- Pettorelli, N.; Ryan, S.; Mueller, T.; Bunnefeld, N.; Jedrzejewska, B.; Lima, M.; Kausrud, K. (2011): The Normalized Difference Vegetation Index (NDVI): unforeseen successes in animal ecology. In: Clim. Res. 46 (1), pp. 15–27. DOI: 10.3354/cr00936.
- Podestá, Guillermo; Letson, David; Messina, Carlos; Royce, Fred; Ferreyra, R. Andrés; Jones, James et al. (2002): Use of ENSO-related climate information in agricultural decision making in Argentina: a pilot experience. In: Agricultural Systems 74 (3), pp. 371–392. DOI: 10.1016/S0308-521X(02)00046-X.
- Pol, Maria; Binyamin, Jacqueline (2014): Impact of Climate Change and Variability on Wheat and Corn Production in Buenos Aires, Argentina. In: AJCC 03 (02), pp. 145–152. DOI: 10.4236/ajcc.2014.32013.
- Portela, Maria Manuela; dos Santos, João Filipe; Silva, Artur Tiago; Benitez, Julián Baez; Frank, Carlos; Reichert, José Miguel (2015): Drought analysis in southern Paraguay, Brazil and northern Argentina: regionalization, occurrence rate and rainfall thresholds. In: Hydrology Research 46 (5), pp. 792–810. DOI: 10.2166/nh.2014.074.
- Potop, Vera; Možný, Martin; Soukup, Josef (2012): Drought evolution at various time scales in the lowland regions and their impact on vegetable crops in the Czech Republic. In: Agricultural and Forest Meteorology 156, pp. 121–133. DOI: 10.1016/j.agrformet.2012.01.002.
- Rodell, M.; House, P. R.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.-J. et al. (2004): The Global Land Data Assimilation. In: American Meteorological Society, pp. 381–394.
- Rusticucci, Matilde; Vargas, Walter (2002): Cold and warm events over Argentina and their relationship with the ENSO phases: Risk evaluation analysis. In: Int. J. Climatol. 22 (4), pp. 467–483. DOI: 10.1002/joc.743.
- Salimi, Hana; Asadi, Esmail; Darbandi, Sabereh (2021): Meteorological and hydrological drought monitoring using several drought indices. In: Appl Water Sci 11 (2). DOI: 10.1007/s13201-020-01345-6.

Schmidt, Sebastian; Barron, Christopher (2020): Mapping Impervious Surfaces Precisely—a GIS-Based Methodology Combining Vector Data and High-Resolution Airborne Imagery. In: *J geovis spat anal* 4 (1). DOI: 10.1007/s41651-020-00055-6.

Souza, A.; Rúbia, C.; Barros, L.; Calhelha, R.; Santos Buelga, C.; Peralta, R. et al. (2015): Phytochemical and bioactive properties of *Ilex paraguariensis*. An in-vitro comparative study between the whole plant, leaves and stems. In: *Food Research International* 78, pp. 286–294. DOI: 10.1016/j.foodres.2015.09.032.

Stagge, James H.; Tallaksen, Lena M.; Gudmundsson, Lukas; van Loon, Anne F.; Stahl, Kerstin (2015): Candidate Distributions for Climatological Drought Indices (SPI and SPEI). In: *Int. J. Climatol.* 35 (13), pp. 4027–4040. DOI: 10.1002/joc.4267.

Strobl, Carolin; Malley, James; Tutz, Gerhard (2009): An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. In: *Psychological methods* 14 (4), pp. 323–348. DOI: 10.1037/a0016973.

Tallaksen, Lena M.; van Lanen, Henny A.J. (2004): *Hydrological Drought. Processes and Estimation Methods for Streamflow and Groundwater*: Elsevier.

Thomasz, Otto Esteban; Vilker, Ana Silvia; Rondinone, Gonzalo (2018): The economic cost of extreme and severe droughts in soybean production in Argentina. In: *CyA* 64 (1), p. 86. DOI: 10.22201/fca.24488410e.2018.1422.

Ulla, T. (2016): El Niño floods in Argentina. A story of displacement and vulnerability. In: *The State of Environmental Migration*, pp. 123–147.

van Loon, A. F.; van Huijgevoort, M. H. J.; van Lanen, H. A. J. (2012): Evaluation of drought propagation in an ensemble mean of large-scale hydrological models. In: *Hydrol. Earth Syst. Sci.* 16 (11), pp. 4057–4078. DOI: 10.5194/hess-16-4057-2012.

Vargas, Walter M.; Naumann, Gustavo; Minetti, Juan L. (2011): Dry spells in the River Plata Basin: an approximation of the diagnosis of droughts using daily data. In: *Theor Appl Climatol* 104 (1-2), pp. 159–173. DOI: 10.1007/s00704-010-0335-2.

Velazco, Santiago José Elías; Galvão, Franklin; Keller, Héctor Alejandro; Bedrij, Natalia Alejandra (2018): Cerrados in Argentina? Structure, diversity and biogeography of the woody component of a savanna in the Misiones Province. In: *Rodriguésia* 69 (2), pp. 335–349. DOI: 10.1590/2175-7860201869206.

Wong, A.K.C.; Wang, Yang (2003): Pattern discovery: a data driven approach to decision support. In: *IEEE Trans. Syst., Man, Cybern. C* 33 (1), pp. 114–124. DOI: 10.1109/TSMCC.2003.809869.