

LC-SPACE-04-EO-2019-2020 “Copernicus evolution – Research activities in support of cross-cutting applications between Copernicus services”

Action acronym: **G3P**
Full title: Global Gravity-based Groundwater Product
Grant Agreement No: 870353



Global Gravity - based Groundwater Product

Deliverable 5.6: G3P-GDI InfoSequia integration and evaluation report – Revision 1

Date: 10.03.2023

Authors: Sergio Contreras & Amelia Fernández

1. Change Record

Name	Author(s)	Date	File name
Submission 1	Contreras	26.12.2022	G3P_D5_6_G3P G3P-GDI InfoSequia integration and evaluation report_final
Revision 1	Haas	10.03.2023	G3P_D5_6_G3P_G3P-GDI_InfoSequia_report_R1

2. Table of Contents

1. Change Record	2
2. Table of Contents	3
3. Overview and summary	4
4. InfoSequia	5
4.1 System overview	5
4.2 System architecture.....	5
5. Groundwater Drought	6
5.1 Introduction and objectives	6
5.2 Drought Index.....	7
5.3 Operationalization.....	8
5.3.1 ARIMA/SARIMA models	8
5.3.2 XGBRegression model	10
5.4 Study region	11
5.5 In-situ data collection.....	12
6. Results.....	13
6.1 Temporal patterns of G3P-GDI and relationship with meteorological drought indices	13
6.2 ARIMA/SARIMA	13
6.3 XGBR model.....	15
7. Conclusions	18
8. References	18

3. Overview and summary

Note: This report is based on G3P version 1.3. This version had not yet been fully gap-filled and error characterized. Accordingly, uncertainty estimates are missing in this report. Please refer to D4.1 – Product Report for further information on version history and uncertainty.

This document is a deliverable in Work Package 5.6 of G3P as described in the Description of Action¹:

This task aims to evaluate a groundwater drought index based on the G3P product and integrate it into InfoSequia, an operational Drought Early Warning System developed by FutureWater. The development of this new groundwater index is being prototyped, calibrated and tested at the 0.5deg. pixel level in the continental Spain. Due to the low latency of current G3P-GWSA products, and in order to make the G3P-GDI product suitable for InfoSequia operational purposes, two methodological approaches based on Time Series Analysis (ARIMA/SARIMA) and ensemble gradient boosting (XGBR) were evaluated as potential candidates for being adopted as nowcasting tools for InfoSequia.

¹ G3P Description of Action – Part A, p. 9

4. InfoSequia

4.1 System overview

InfoSequia is an operational Drought and Early Warning System developed by FutureWater able to provide seasonal outlooks of risk of critical failure on the productivity of croplands at the district level, or the availability of water resources at basin scale.

Seasonal outlooks of crop and water supply failures are delivered monthly and are estimated by InfoSequia rest on machine learning techniques which are trained with a comprehensive set of predictors computed at different timescales which includes satellite-based drought indices, atmospheric oscillation indices and other indicators derived from hydrological modelling, reanalysis datasets, or ground measurements. The set of predictors generated by InfoSequia cover the main components of a drought and provide the most reliable and accurate picture of the drought status of a region including its magnitude, severity, spatial extent, and persistence.

The system is characterized by its modularity, robustness, flexibility and the quality of its outcomes (e.g., data with high spatial and temporal resolution, short revisit times). The system is very suitable for being coupled with other monitoring or forecasting technologies, or for helping different market segments by addressing their needs, including:

- more efficient and transparent management of droughts (triggering alerts and prioritizing of actions) time and costs saving by simplifying reporting commitments (useful for water management authorities, and disaster response and management agencies)
- adjustment of irrigation quotas and restrictions at the district or province level (agricultural sector – irrigation boards and extension officers).
- early detection of scheme failures (e.g., crop yield damages and water shortage events) which allows a more accurate planning of financial and insurance needs, and to optimize the allocation of human resources for field-based inspections (insurance/re-insurance).
- triggering of ex-ante cash transfer and voucher programs (civil defense and humanitarian-aid agencies and response organizations)

InfoSequia is provided at different service levels (light vs full versions) and deployment options (e.g., open-standard vs restricted-authenticated, web-mapping vs PaaS technologies), depending on user requirements.

4.2 System architecture

InfoSequia is composed of three main modules:

- 1) The IS-MONITOR module, which includes the set of algorithms and subroutines that compute the drought indices and indicators from satellite and climate data collected from native and cloud-hub repositories
- 2) The IS-4CAST, which includes the machine-learning routine which produces the seasonal outlooks of risk of failure of a particular system.
- 3) The IS-VIEWER, which includes all the front-end functionalities and solutions specifically designed for accessing the data generated in IS-MONITOR and IS-4CAST modules. Data can be delivered via bulletins, a web-mapping interface, or via FTP. Also, other options like access through existing Platform-as-a-Service (PaaS) are available.

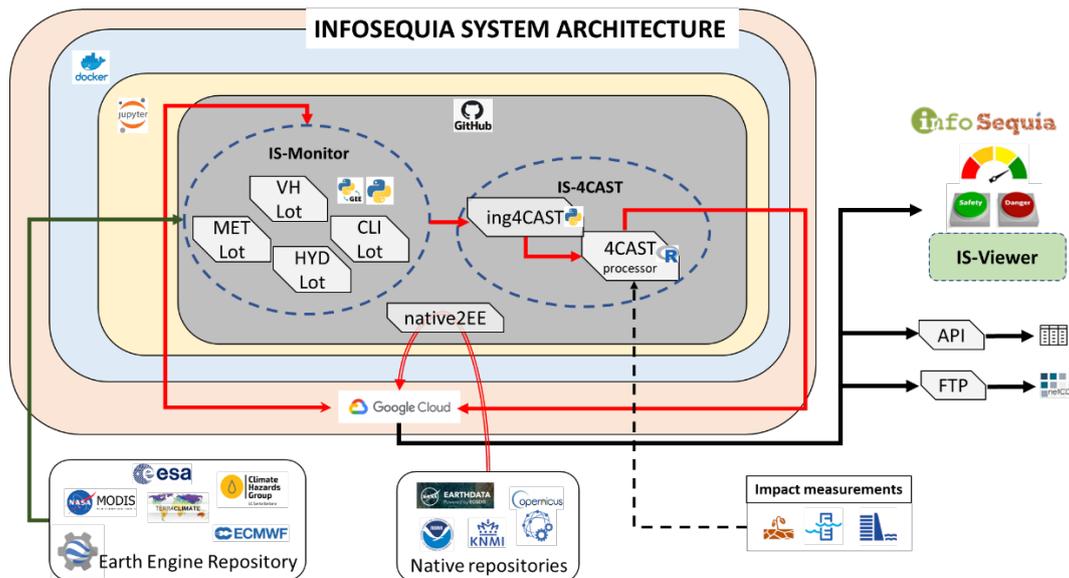


Figure 1. Functional block diagram of InfoSequia.

5. Groundwater Drought

5.1 Introduction and objectives

Groundwaters play a critical role for the water security of regions, and especially drylands where the access to these resources may be an important source of water for mitigating the impacts of water in the surface. Groundwater drought refers to a period of decreased groundwater levels that results in water-related problems². The magnitude and severity of groundwater level changes will depend on local conditions and the interplay between groundwater and climate or human factors. To monitor groundwater drought is then crucial to evaluate the vulnerability of a region to surface water shortages. In addition to other drought and climate indices, a groundwater drought index may be considered as a potential predictor to support seasonal forecasting of drought impacts.

Groundwater level observations are the basis for the quantification of the onset, magnitude, and severity of groundwater droughts. However, the suitability and reliability of groundwater level observations to support monitoring purposes will depend on several factors like:

- Length of record, which is of high importance to set up the reference or “normal” needed to compute the relative anomalies
- Frequency and latency at which observations are taken and delivered (e.g. near-real time monitoring systems vs large time-lag system)
- Perturbance factors or non-control factors as intensive pumping or construction defects (Taylor and Alley, 2001) that may affect the right characterization of true groundwater drought conditions
- Density or spatial coverage of the monitoring network across the aquifer or region of interest

² <https://water.usgs.gov/ogw/drought/>

Because of these limitations, in some regions the usage of water level observations is not fully recommended and need to be replaced by other methods like those ones which rest on satellite-based measurements. For those cases, estimates of groundwater level fluctuations derived from GRACE gravimetry anomaly measurements have been shown a very promising method to support drought monitoring efforts (Thomas et al., 2017).

By using a residual water-balance approach, the G3P system is able to isolate the groundwater signal from the total water storage changes derived from the raw GRACE signal. The final G3P product refers to the G3P-GWSA, i.e. the GroundWater Storage Anomaly defined as absolute deviations of total groundwater storage changes in equivalent water height at a particular month relative to the long-term monthly average in a baseline time period. The GWSA is computed in G3P as,

$$GWSA_t = TWSA_t - SWEA_t - RZSMA_t - GWEA_t - SWSA_t$$

Equation 1

being, TWSA the Total Water Storage Anomaly (mm), SWEA Snow Water Equivalent Anomaly (mm); RZSMA Root Zone Soil Moisture Anomaly (mm), GWEA the Glacier Water Equivalent Anomaly (mm), SWSA the Surface Water Storage Anomaly (mm), and t indicates the time. Detail explanations on auxiliary datasets and methodologies employed for estimating the isolate components in the right side of equation are provided in the technical reports of the G3P project.

This work aims to bridge the lack of groundwater indicators in InfoSequia by ingesting the G3P product into its operational workflow. This process requires to address two important tasks:

- To establish a conceptual procedure to compute a groundwater drought index derived from the G3P product.
- To explore and to find a suitable method able to provide up-to-date values of GP3 data in an attempt to meet the technical requirements of InfoSequia.

To address both tasks, the G3P-GWSA product v1.3 resulting from the application of an optimal Gaussian filtering solution to the isolate components of the water balance was used. In this version, the groundwater storage anomalies refer to a baseline time period ranging from 2002-04 to 2020-03.

5.2 Drought Index

After a review of different options available in the scientific literature, the methodology suggested by (Thomas et al., 2014, 2017) was adopted to retrieve a groundwater drought index. This is computed as the normalized water storage residuals from a monthly climatology, in which negative residuals would describe deficits or relatively dry conditions, while positive values would do for surpluses or wet conditions.

Monthly climatology is used to remove the influence of seasonality in groundwater storage changes. It is computed as:

$$C_i = \frac{\sum_1^{n_i} GWSA_i}{n_i}, i = 1, \dots, 12$$

Equation 2

After, the monthly climatology is subtracted from the GWSA to retrieve a groundwater storage deviation (GSD). Finally the groundwater drought index is computed by normalizing the GSD timeseries using the Z-score approach.

$$GSD_i = GWSA_i - C_i, i = 1, \dots, 12$$

Equation 3

$$GDI_i = \frac{GSD_i - \overline{GSD}_i}{s(GSD_i)}$$

Equation 4

Being the \overline{GSD}_i the average of the net groundwater storage deviation, and $s(GSD_i)$ the standard deviation along the timeseries.

5.3 Operationalization

At this stage, the nowcasting of G3P-GWSA product is a prerequisite of InfoSequia in order to sort out the high latency of the G3P-GSWA product which is generated once per year. In this study, two methods have been tested and evaluated to nowcast or extend forward the GRACE signature: the ARIMA/SARIMA method, and the eXtreme Gradient Boosting Regression (XGBR) algorithm.

The overall performance of both methods was tested at the pixel level along the region of study (continental Spain) and using the G3P-GDI generated for continental Spain.

5.3.1 ARIMA/SARIMA models

The first method evaluated and tested as a nowcasting candidate tool of G3P-GDI values in based on the Time Series Analysis (TSA) approach. Models that adopt this approach rest on the basic assumption that autocorrelation plays a prominent role in forecasting the behaviour of a target variable. Among the TSA models, the ARIMA (AutoRegressive Integrated Moving Average) model is one the most widely used and recognized statistical methods for forecasting time series because their high accuracy and efficiency in representing various types of time series. The SARIMA model is a variant that includes the seasonality component in the formulation. Both type of models have been widely applied in a large number of hydrological applications (Valipour, 2015) and drought analysis (Mishra and Singh, 2011). ARIMA and SARIMA models are formally represented as:

$$ARIMA(p, d, q)$$

$$SARIMA(p, d, q)(P, D, Q)[s]$$

where p is the order of the non-seasonal autoregressive model (number of autoregressive terms), q is the order of non-seasonal moving average model (number of lagged forecast errors in the prediction equation), and d is the number of non-seasonal differences needed

for stationarity. P, Q and D have similar meaning but for the seasonal component, and s is the periodic term.

The AutoRegressive component (AR) indicates the correlation of the variable against itself according to the equation Equation 5.

$$m_t = p_0 + p_1 m_{t-1} + p_2 m_{t-2} + p_3 m_{t-3} \dots p_n m_{t-n}$$

Equation 5

where the vector $[p_0, \dots, p_n]$ represents the regression coefficients for the different time lags. The optimal values of p are determined by analyzing the Autocorrelation function (ACF) and the Partial Autocorrelation function (PACF). The ACF shows the correlation between the current and the past values of the same variable. The PACF measures the direct correlations between past values and current values.

The Integrated component (I) represents any differencing that has to be applied in order to make the data stationary. For example, a differencing factor ($d=1$) would mean a lag of i.e. $m_t - m_{t-1}$. Usually, a dickey fuller test is done to check the stationarity of the data.

Final the Moving Average (MA) term refers to how past forecast errors during the historical training period (or hindcast) are taken in the model to reduce the error of the forecast. A MA model follow the equation Equation 6

$$m_t = q_0 + \varepsilon_{t-1} + q_2 \varepsilon_{t-2} + q_3 \varepsilon_{t-2} \dots q_n \varepsilon_{t-1}$$

Equation 6

In which the ε is call an error and represents the random residual deviations between the model and the target variable. Usually, the matrix of error are established using iterative techniques like Maximum Likelihood Estimation.

SARIMA stands for Seasonal-ARIMA and it includes the contribution of seasonality to the forecast³. When the effect of seasonality is evident, ARIMA models usually fails making advisable to include the seasonal effect to increase the robustness of the forecast estimation. The Autoregressive (AR), Integrated (I), and Moving Average (MA) parts of the model (p,d,q) remain as that of ARIMA. The seasonal parts (P,D,Q) are also deduced from the ACF y PACF plots.

The typical workflow applied for an ARIMA/SARIMA method involves three steps:

- *Exploratory analysis*. It aims to check the stationarity of a timeseries⁴ by identifying the potential presence of consistent temporal patterns like long-term trends, cycles, or seasonality. The decomposition of a timeseries into its main components (trend-cycle component, seasonal component, and error component) is called the ETS decomposition. This analysis is performed in InfoSequia by applying the *statsmodels* library developed in Python.

- *Choosing and fitting models*. The selection of a model depends on the strength of the correlation between the target variable and a set of potential predictors. In TSA, predictors refer to the target variable itself but lagged in time. Usually, several models with different configuration settings are fitted for a training period. To support the decision on which model

³ <https://neptune.ai/blog/arima-sarima-real-world-time-series-forecasting-guide>

⁴ TSA models require that data is stationary or independent to time influence .

best fit the variable, InfoSequia uses the Auto-ARIMA functionality in Python. This package is able to generate the optimal set of parameter (p, d, q , and P, D, Q) values which would provide the better forecasting. Auto-ARIMA works over a hyper-parameter space which is defined by the range of values suitable for p , d and q parameters. A grid search algorithm is then implemented over the space and the best combination is found as that one that minimize a performance indicator (e.g. Akaike Information Criterion). Here, we set a range of potential values from 0 to 3 for the p/P and q/Q values, and a maximum of 3 for d parameter. For the grid search algorithm, InfoSequia applies a stepwise algorithm which is usually faster and more effective than exhaustive or random algorithms.

- *Performance evaluation.* Finally, and once a model is selected and its parameters estimated, the performance of the model to predict the target variable is quantified for a testing (blinded) period.

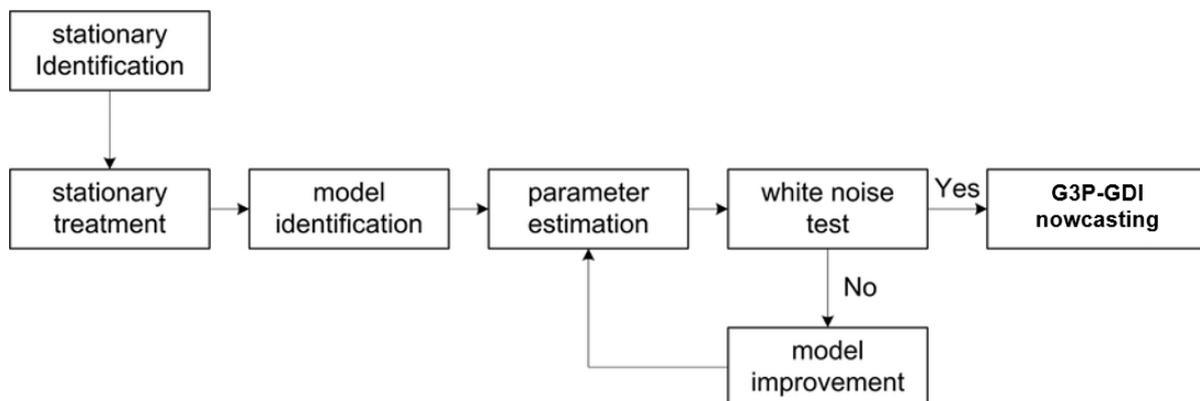


Figure 2. Flowchart of methodological steps in ARIMA/SARIMA models

5.3.2 XGBRegression model

The eXtreme Gradient Boost Regression model (XGBR) is a type of Boosting algorithm. Boosting algorithms are sequential ensemble algorithms that that converts weak learners⁵ to strong learners which usually improves the prediction accuracy by decreasing bias (Zhang et al., 2020). Boosting algorithms pay most attention to the samples with highest prediction errors and increase their weights in the next iteration which makes that the algorithm learns from previous mistakes. The final prediction results a weighted combination of predictions across the sequence of learners (Figure 3).

⁵ A weak learner refers to a learning algorithm that only predicts slightly better than randomly.

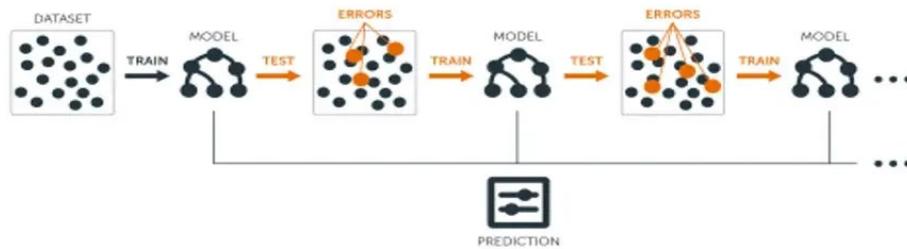


Figure 3. Typical flowchart followed used in boosting algorithms.

Extreme Gradient Boosting is an advanced implementation of the Gradient Boosting. It is a powerful and very scalable and accurate algorithm which improves greatly the overall performance and reduce the overfitting issues.

5.4 Study region

The study area covers continental Spain, which spans over an area of approximately 500,000 km². The region is characterized by its large climate diversity, where up to 5 climate Köppen-Geiger zones exist. Hot-summer (*Csa*) and warm-summer (*Csb*) Mediterranean climates mostly dominate the entire country along the South and Eastern Mediterranean coastline as well as the central plains located in the inland, whereas Oceanic climate (*Cfb*) prevails in the North and North-West coastline, and warm-summer Continental (*Dfb*) governs the North-Eastern sectors. The mean annual precipitation is around 680 mm, but it varies strongly in time and space⁶. Meteorological drought periods are very frequent, while extreme floods usually happen during the autumn period along the Mediterranean coastline fringe as result of convective rainfall events.

The retrieval of the G3P-GDI and the application of the nowcasting algorithms was applied at 0.5deg. pixel level. The G3P's 1deg and 0.5deg grid for the continental Spain are shown in Figure 4. Each 0.5deg quadrant was coded following a two capital letters which indicate the row and column position in the 1deg grid, and 2-number digits that indicates relative position inside the 1 deg. quadrant (11 for the upper-left, and 22 for the lower-right positions).

⁶ https://en.wikipedia.org/wiki/Climate_of_Spain

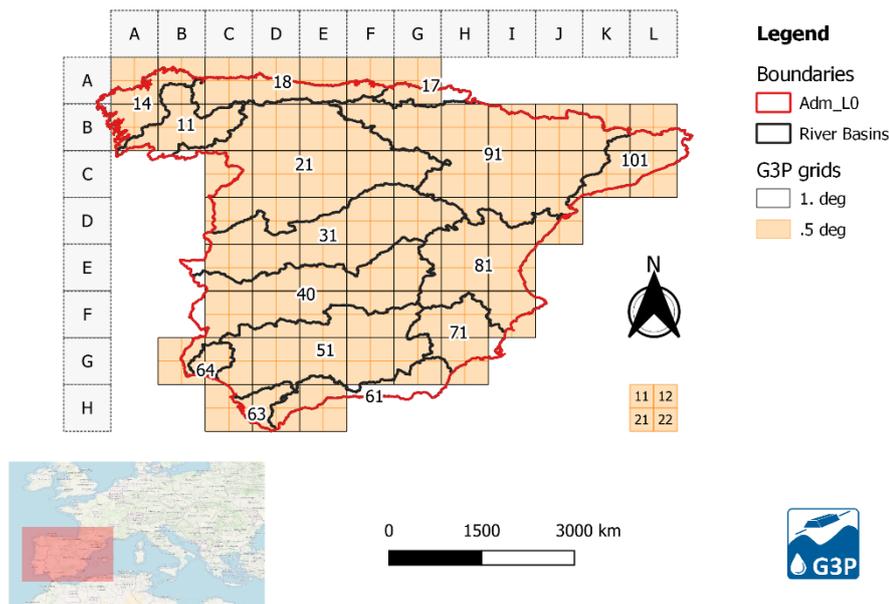


Figure 4. Division of the study area by GRACE quadrants of 1 deg and 0.5 deg. River Basins districts (from N-S): 14: Galicia-Costa, 18: Cantábrico Occidental, 17: Cantábrico Oriental, 11: Miño-Sil, 21: Duero, 91: Duero, 101: Cuencas Internas de Cataluña, 31: Guadiana, 81: Júcar, 40: Guadiana, 64: Tinto, Odiel y Piedras, 51: Guadalquivir, 71: Segura, 63: Guadalete y Barbate, 61: Cuenca Mediterránea Andaluza.

5.5 In-situ data collection

Similarly, a GroundWater Drought Index was also derived at the 0.5deg. pixel level using water level measurements collected from the Spanish National Water Monitoring System. A detailed description of this dataset is provided in Deliverable D4.2. For this particular case, ground measurements were taken for verifying the patterns and magnitude of the relationships between the G3P-GDI, meteorological drought indices and the ground measurements. The strength of these relationships were quantified through the Person correlation matrix measured between drought indices.

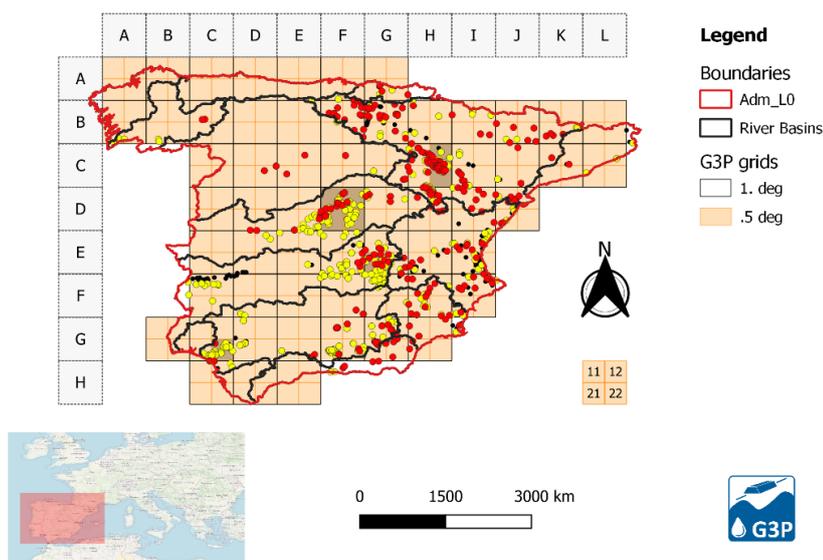


Figure 5. Distribution of qualified wells from which groundwater level changes were extracted at the 0.5deg pixel level. 0.5 deg. quadrants selected for evaluating relationships between drought indices are in grey shadow. From North to South, CH12 and CH22 (Ebro valley); DF11, DF12, DF21 and DF22 (Headwater of Tajo river basin); EG21 (Headwater of Guadiana river basin); GC22 (Doñana National Park at the outlet of Guadalquivir river).

6. Results

6.1 Temporal patterns of G3P-GDI and relationship with meteorological drought indices

In general, patterns of Pearson correlation found between the ground-truth groundwater drought index and the G3P and meteorological drought indices are extremely weak, with values most of the time lower than 0.3 or a little higher for long timescale aggregations. The highest correlation were found at the Ebro valley and particularly in pixels CH12 and CH22 (Figure 6). In those locations, correlation increased with timescale aggregation reaching the highest values at 6 and 12 months).

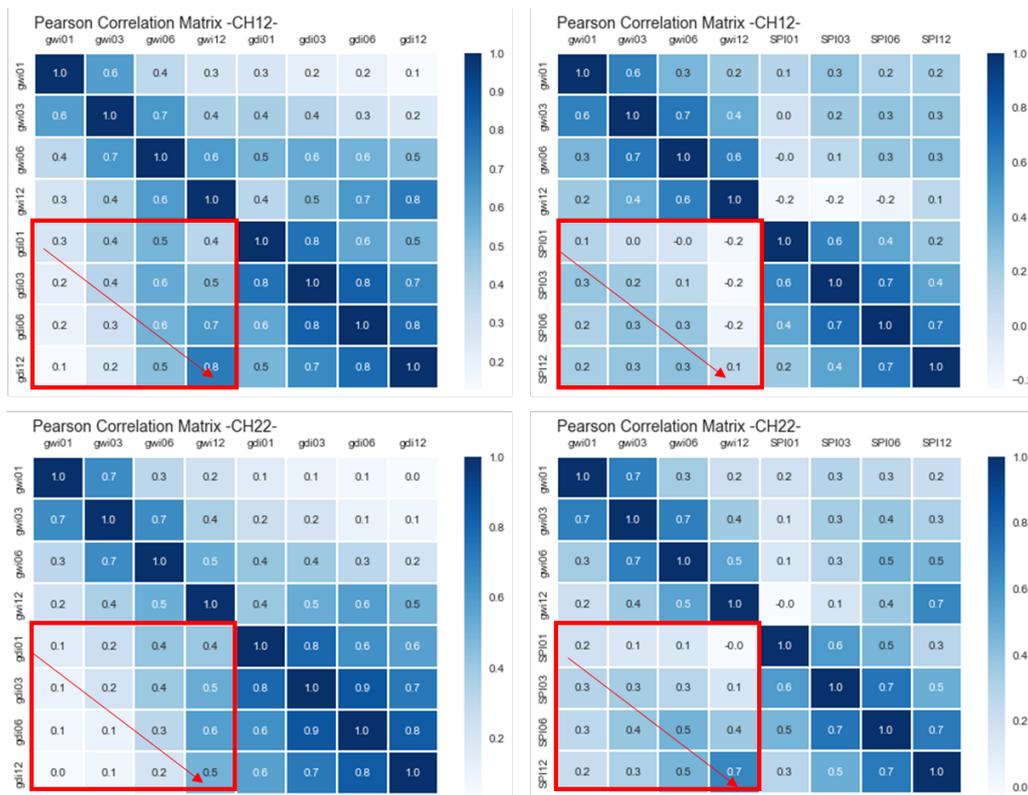


Figure 6. Pearson correlation matrices between drought indices: gwi = groundwater drought index retrieved at the pixel level from water well observations; gdi = groundwater drought index retrieved from G3P-GWSA product; SPI = Standardized Precipitation Index computed by InfoSequia from CHIRPS dataset.

6.2 ARIMA/SARIMA

This section shows the results derived from the application of the typical TSA workflow for the generation of ARIMA/SARIMA models for the study region.

Figure 7 shows the ETS components of four G3P-GDI timeseries extracted in the study region. Decomposition was applied over a subdataset extracted from the original one. The splitting strategy adopted resulted in:

- Original dataset: Abr'2002 – Mar'2020
- Period for training-testing: Jan'03 – Dec'2019
- Training period: Jan'2003 – Dec'2016 (~80% of the training-testing period)
- Testing period: Jan'2017 – Dec'2020

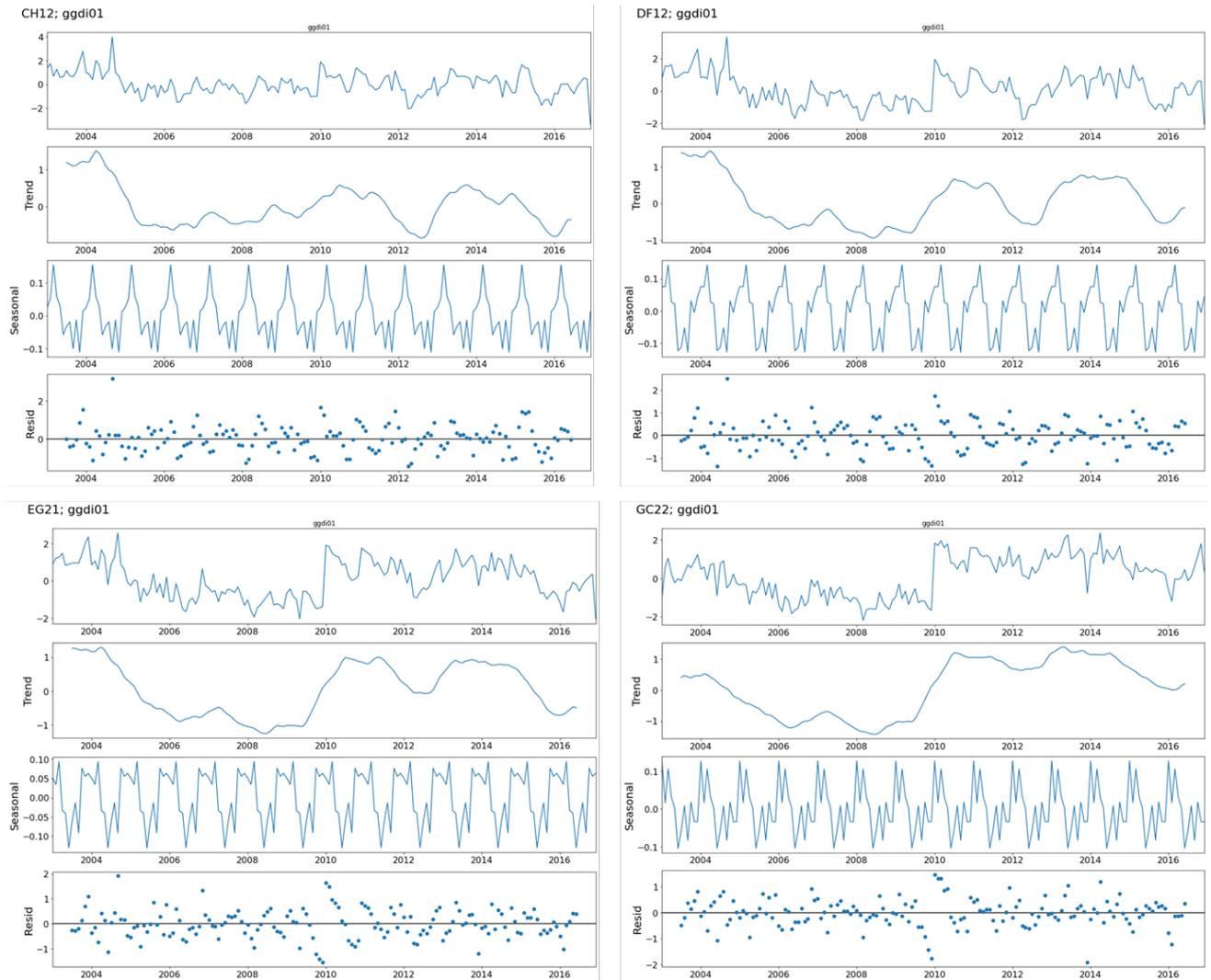


Figure 7. Error/Trend/Seasonal Decomposition of timeseries of the G3P-GDI for selected pixels.

Results from the application of the Auto-ARIMA functionality are shown in Table 1 for the set of 0.5deg pixels selected to illustrate the process. Once selected the model, its performance to predict actual values was evaluated for the training (hindcast) and the testing periods. In Figure 8, the estimated values retrieved from the SARIMA models would indicate a clear problem of overfitting due to the large differences in the ability of the model to predict actual values in the train vs testing period. Root Mean Square Error values in the testing period were higher than in the training period by 50-80%.

Table 1. Best model configurations retrieved for the selected pixels.

ID	Stepwise
CH12	ARIMA(1,0,0)(3,1,1)[12]
CH22	ARIMA(1,0,0)(3,1,0)[12]
DF11	ARIMA(1,0,1)(0,1,1)[12]
DF12	ARIMA(1,0,1)(0,1,1)[12]
DF21	ARIMA(1,0,1)(0,1,1)[12]
DF22	ARIMA(1,0,1)(3,1,0)[12]
EG21	ARIMA(2,0,1)(3,1,0)[12]
GC22	ARIMA(3,0,0)(0,1,1)[12]

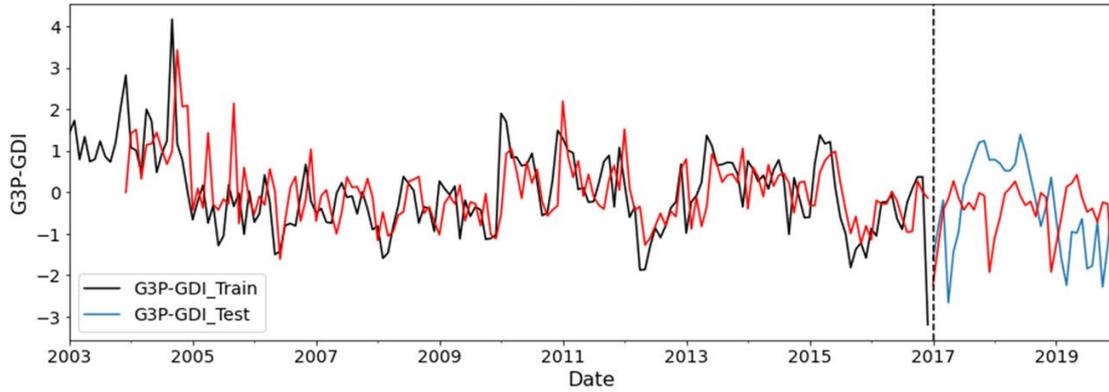
6.3 XGBR model

Figure 9 show the comparison between actual and predicted values from XGBR models for the training and testing periods. In this case SPI values at timescales of 1-, 3- and 6-months were used as potential predictors of the monthly G3P-GDI product. Additionally, monthly averages for the period of analysis were also added as auxiliary predictors. Results confirms the higher performance of the XGBR method against the SARIMA approach.

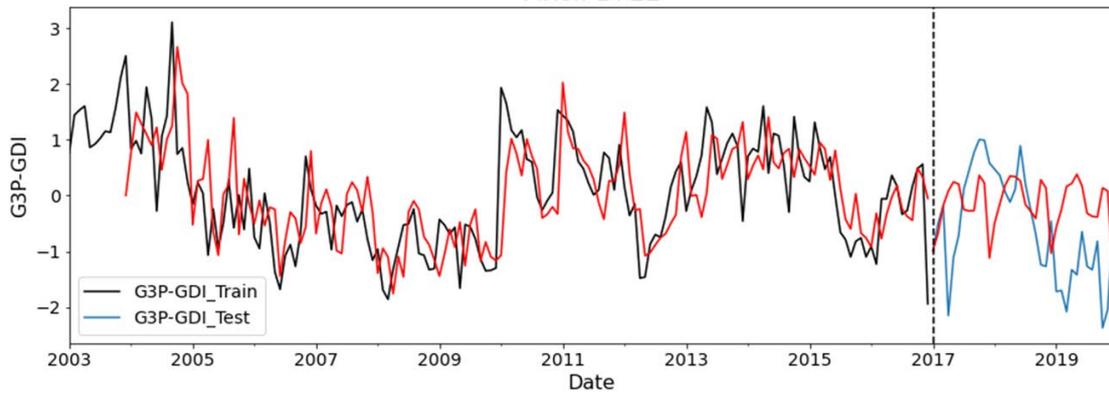
Table 2. Statistical metrics (MSE = Mean Square Error, RMSE = Root Mean Square Error) retrieved from XGBR models applied to the G3P-GDI timeseries at 1-, 3- and 6-month timescales.

GDI	ID	MSE_train	MSE_test	RMSE_train	RMSE_test
GDI01	CH12	0.052	1.446	0.228	1.202
	CH22	0.035	1.609	0.187	1.268
	DF11	0.052	1.092	0.228	1.045
	DF12	0.078	1.189	0.279	1.09
	DF21	0.054	1.113	0.232	1.055
	DF22	0.046	1.434	0.214	1.197
	EG21	0.053	1.5	0.23	1.225
	GC22	0.112	0.635	0.335	0.797
GDI03	CH12	0.052	1.769	0.228	1.33
	CH22	0.039	1.763	0.197	1.328
	DF11	0.051	1.089	0.226	1.044
	DF12	0.055	1.21	0.235	1.1
	DF21	0.06	1.212	0.245	1.101
	DF22	0.072	1.389	0.268	1.179
	EG21	0.067	1.583	0.259	1.258
	GC22	0.085	0.845	0.292	0.919
GDI06	CH12	0.096	2.232	0.31	1.494
	CH22	0.063	1.531	0.251	1.237
	DF11	0.065	1.39	0.255	1.179
	DF12	0.073	1.181	0.27	1.087
	DF21	0.066	1.035	0.257	1.017
	DF22	0.073	1.143	0.27	1.069
	EG21	0.052	1.372	0.228	1.171
	GC22	0.15	0.765	0.387	0.875

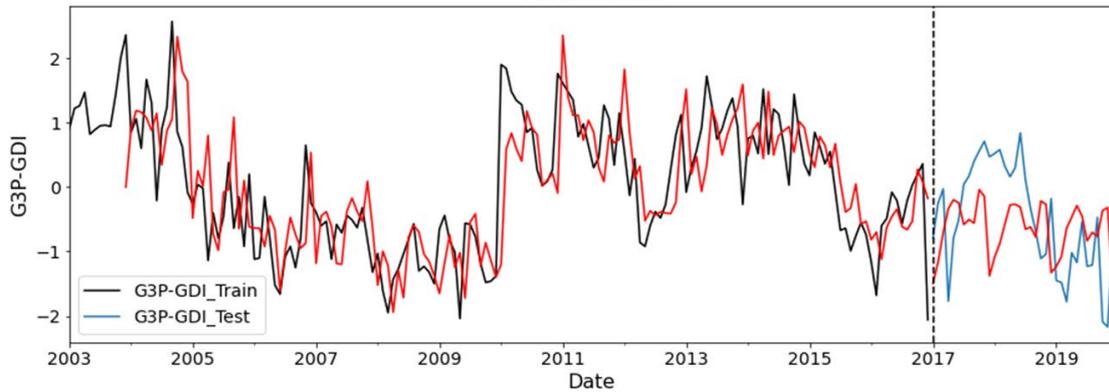
Figure 9
Pixel: CH22



Pixel: DF22



Pixel: EG21



Pixel: GC22

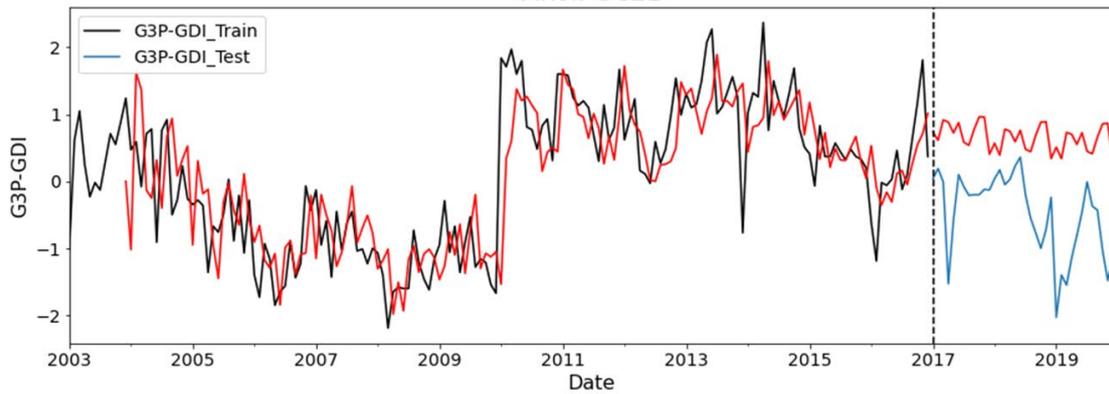


Figure 8. G3P-GDI01 original values (black for training period, blue for testing period) and predicted values from best SARIMA model (red) found for four selected pixels.

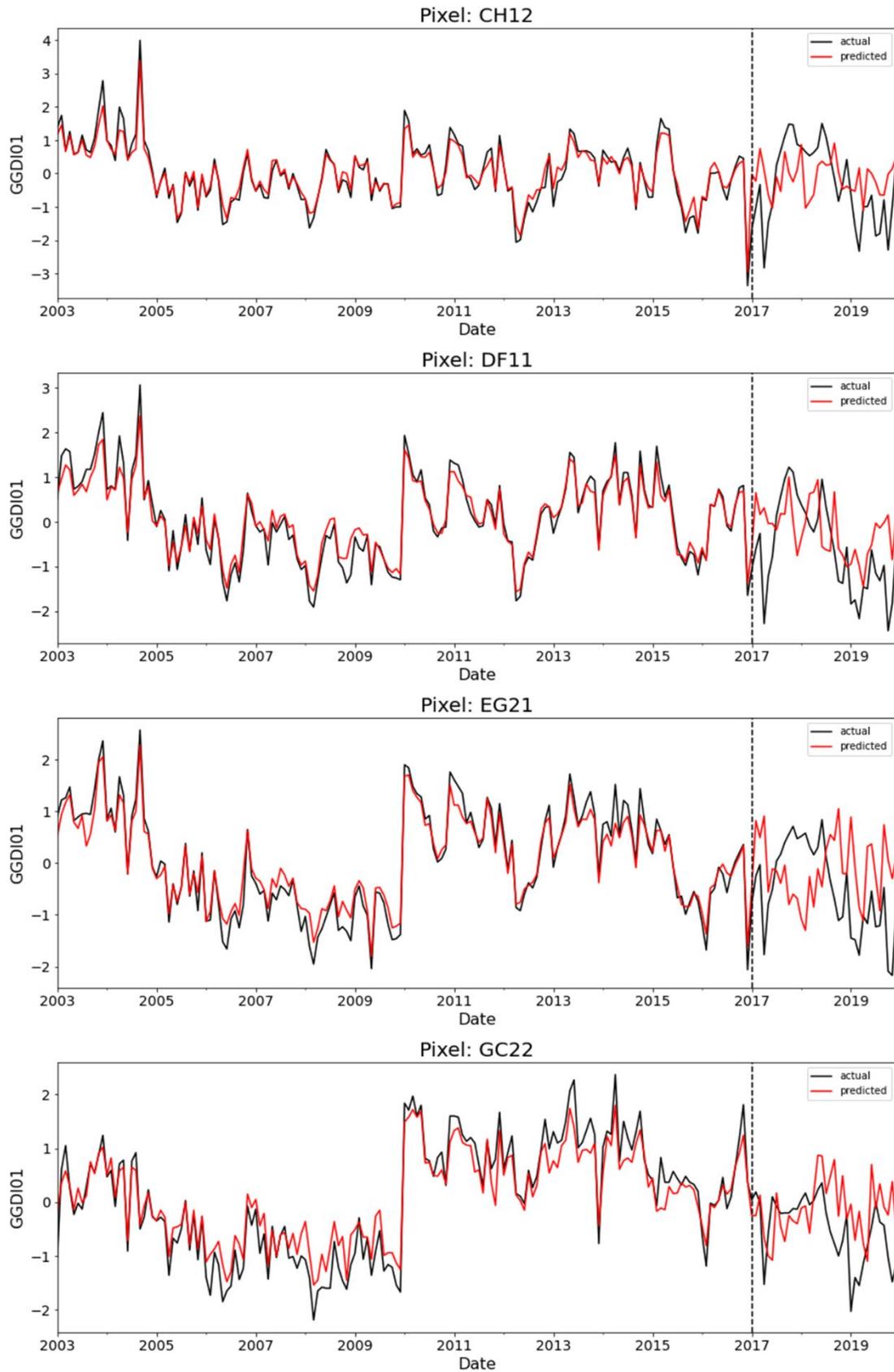


Figure 9. G3P-GDI01 actual values (black) and predicted values from the XGBR model (red). Vertical dashed line splits the period into training dataset (left side) and testing dataset (right side).

7. Conclusions

In this report a methodology to compute a GroundWater Drought Index (GDI) has been proposed and successfully implemented to the G3P-GWSA product. Additionally, due to the low latency of the current G3P version products, here two methodological approaches resting on TSA (ARIMA/SARIMA models) and ensemble boosting algorithms (Extreme Gradient Boosting Regression) were evaluated for being used as nowcasting tools into the InfoSequia service. The XGBR method, using Standardized Precipitation Index at different timescale aggregations and mean monthly values of the target variable, showed much better performance than the best ARIMA/SARIMA models fitted. Additionally, XGBR method demonstrated to be much faster and computationally more efficient than TSA algorithms.

8. References

- Mishra, A.K., Singh, V.P., 2011. Drought modeling – A review. *J. Hydrol.* 403, 157–175. <https://doi.org/10.1016/j.jhydrol.2011.03.049>
- Taylor, C.J., Alley, W.M., 2001. *Ground-Water-Level Monitoring and the Importance of Long-Term Water-Level Data*. Denver.
- Thomas, A.C., Reager, J.T., Famiglietti, J.S., Rodell, M., 2014. A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophys. Res. Lett.* 41, 1537–1545. <https://doi.org/10.1002/2014GL059323>
- Thomas, B.F., Famiglietti, J.S., Landerer, F.W., Wiese, D.N., Molotch, N.P., Argus, D.F., 2017. GRACE Groundwater Drought Index: Evaluation of California Central Valley groundwater drought. *Remote Sens. Environ.* 198, 384–392. <https://doi.org/10.1016/J.RSE.2017.06.026>
- Valipour, M., 2015. Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorol. Appl.* 22, 592–598. <https://doi.org/10.1002/met.1491>
- Zhang, Y., Ma, J., Liang, S., Li, X., Li, M., 2020. An Evaluation of Eight Machine Learning Regression Algorithms for Forest Aboveground Biomass Estimation from Multiple Satellite Data Products. *Remote Sens.* 12, 4015. <https://doi.org/10.3390/rs12244015>