

Geosciences

Prediction of flooded area based on time series generated by images of satellites

Predição de área alagada com base em séries temporais geradas por imagens de satélites

Heverton Barbosa do Nascimento^I , Paulo Renato Alves Firmino^{II} ,
Francisco Alixandre Ávila Rodrigues^{II} , Jair Paulino de Sales^{II} ,
Rosilda Benício de Souza^{II} 

^ICentro Universitário Católica de Quixadá, Ceará, CE, Brasil

^{II}Universidade Federal do Cariri, Juazeiro do Norte, CE, Brasil

ABSTRACT

Digital image processing methodologies can be applied to different problems related to sustainable development, including the monitoring of water bodies through multi-temporal satellite images. Monitoring and predicting flooded areas are relevant procedures for understanding waterbody dynamics and supporting studies related to floods, droughts and water-resource management. This study proposed a methodological approach to map and predict variations in the flooded area of Kirwin Lake, USA, based on Landsat 8 satellite images collected between January 2013 and November 2022. The flooded area was estimated through the Otsu image segmentation method, and the resulting monthly time series was modeled using Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), Exponential Smoothing (ETS), and Holt-Winters models. The results showed evidence of annual seasonality in the flooded-area time series. Among the evaluated models, Holt-Winters presented the best predictive performance, with the lowest RMSE, MAPE and Theil's U coefficient values, indicating greater ability to represent the seasonal behavior of the flooded area and to produce predictions closer to the observed values. Therefore, the proposed approach proved to be useful for transforming satellite images into a predictive monitoring tool, contributing to future studies on hydrological monitoring and reservoir surface-area dynamics.

Keywords: Remote sensing; Hydrological monitoring; Reservoir dynamics; Threshold segmentation; Otsu algorithm; Water surface variability

RESUMO

Metodologias de processamento digital de imagens podem ser aplicadas a diferentes problemas relacionados ao desenvolvimento sustentável, incluindo o monitoramento de corpos hídricos por meio de imagens de satélite multitemporais. O monitoramento e a predição de áreas alagadas são procedimentos relevantes para compreender a dinâmica de corpos d'água e subsidiar estudos relacionados a inundações, secas e gestão de recursos hídricos. Este estudo propôs uma abordagem metodológica para mapear e prever variações na área alagada do Lago Kirwin, nos Estados Unidos, com base em imagens do satélite Landsat 8 coletadas entre janeiro de 2013 e novembro de 2022. A área alagada foi estimada por meio do método de segmentação de imagens de Otsu, e a série temporal mensal resultante foi modelada utilizando os modelos Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), Exponential Smoothing (ETS) e Holt-Winters. Os resultados evidenciaram sazonalidade anual na série temporal da área alagada. Entre os modelos avaliados, o Holt-Winters apresentou o melhor desempenho preditivo, com os menores valores de RMSE, MAPE e coeficiente U de Theil, indicando maior capacidade de representar o comportamento sazonal da área alagada e produzir previsões mais próximas dos valores observados. Portanto, a abordagem proposta mostrou-se útil para transformar imagens de satélite em uma ferramenta preditiva de monitoramento, contribuindo para estudos futuros sobre monitoramento hidrológico e dinâmica da área superficial de reservatórios.

Palavras-chave: Sensoriamento remoto; Monitoramento hidrológico; Dinâmica de reservatórios; Segmentação por limiarização; Algoritmo de Otsu; Variabilidade da superfície hídrica

1 INTRODUCTION

The analysis of waterbodies based on digital image processing stands out in the sustainable development context. Landslide mapping (Intrieri *et al.*, 2018), sealed surface estimates (Lei; Weng, 2018), detection of changes in forests (Xu; Wang, 2020), and river flood monitoring (Khan *et al.*, 2014; Muhadi *et al.*, 2020) are some examples.

Recently, deep learning techniques have been applied to detect and delineate water bodies in high-resolution satellite images, using tools like Detectron2 (Fernandes; Fialho; Patriarca, 2024). Moreover, artificial intelligence-based approaches have been developed for near real-time monitoring of contaminants in coastal waters, integrating remote sensing data and onboard satellite processing, such as in the Phisat-2 mission (Razzano *et al.*, 2024). Another relevant application is the estimation of water turbidity using remotely piloted aircraft (RPA) images combined with machine learning techniques like

Artificial Neural Networks and Random Forest, providing an efficient water quality monitoring system (Andrade *et al.*, 2024).

According to Zhou *et al.* (2020), monitoring the surface of lakes and rivers is essential to guide actions aimed at mitigating damage to exposed populations. It is necessary to constantly take images of regions holding waterbodies. This process allows requires constant image acquisition from the same region, at different moments, to gather the so-called “multi-temporal images”. The flooded areas in water bodies based on multi-temporal images are an important aspect to be assessed. Using the segmentation method is a natural way to estimate these areas. Zhou *et al.* (2020) used the classic method known as Otsu (Otsu, 1979) to identify the surface of lakes. Muhadi *et al.* (2020) introduced a study that compared image segmentation techniques used to extract water information based on digital images to monitor floods in Malaysia. In addition, recent studies on sustainable development address that risks involved in wetlands can significantly affect populations living in Sri Lanka wetlands (Samaraweera, 2024).

Therefore, monitoring and forecasting the flooded areas in water bodies are essential procedures to provide warnings on both floods and water shortages. The present study aims to propose a method to assess and predict flooded areas in water bodies.

2 TIME SERIES MODELS

Overall, a time series can be represented by a graphic that shows patterns, such as upward and downward moves (trend) or a pattern that repeats itself (seasonal variation) (Hyndman *et al.*, 2008). Information to set observations of a time series ($y_1, \dots, y_n$) is an attempt to predict a future observation in time $n + h$ (Hyndman *et al.*, 2008).

Thus, y_t is an observation in time t , and χ_t is the state vector containing unobserved components that describe the level, trend, and seasonality of the

series. In the state-space model of linear innovations, the measurement equation, presented in Equation (2.1), describes the relationship between the unobserved state vector and the observed value of the series (Hyndman et.al, 2008). The transition equation, presented in Equation (2.2), describes the evolution of the state vector over time (Hyndman et al., 2008). The model can be written as follows:

$$y_t = w'x_t - 1 + Et \quad (2.1)$$

$$x_t = F x_{t-1} - 1 + gEt \quad (2.2)$$

Where Et is a white noise series and F , g , and w are coefficients of the state-space representation. Hence, the time series y_t is an additive model represented by the combination of components trend T , seasonality S , and error E , which is not captured by the model at period t , according to Equation (2.3) (Hyndman *et al.*, 2008).

$$y = T + S + E \quad (2.3)$$

Accordingly, T corresponds to the long-term direction of the series, S is the pattern repeating itself within a known periodicity and E is the unpredicted behavior of the series (Hyndman *et al.*, 2008). Time series data can show a wide variety of patterns, so it is interesting to split this series into several components; each one of them would represent an underlying pattern category (Hyndman & Athanasopoulos, 2018). This decomposition can be additive or multiplicative, according to Equations (2.4) and (2.5), respectively:

$$y_t = S_t + T_t + R_t \quad (2.4)$$

$$y_t = S_t \times T_t \times R_t \quad (2.5)$$

Where in y_t are data, S_t is the seasonal component, T_t is the component of the trend cycle and R_t is the remaining component – all these components are gathered in period t . The use of each one of these components will depend on variations in the seasonal pattern or on the cycle of trends (Hyndman & Athanasopoulos, 2018).

Data transformation until variations in the series seem stable over time, followed by additive decomposition, is an alternative to multiplicative decomposition (Hyndman & Athanasopoulos, 2018). Thus, whenever a log transformation is used, it will become equivalent to a multiplicative decomposition, as shown in Equation (2.6).

$$\log y_t = \log S_t + \log T_t + \log R_t \quad (2.6)$$

Accordingly, the decomposition of the time series recorded for the flooded area was carried out to analyze this area's behavior in detail. It was essential to assess the time series to generate a data prediction for the flooded area in the waterbody, therefore, it was necessary to build models to better disclose them.

2.1 Autoregressive integrated moving average

The ARIMA, proposed by George Box and Gwilym Jenkins, back in 1979 (BOX, 2013), was the first adopted formalism. When it comes to architecture terms, the study followed ARIMA (p, d, q), according to which, p is the number of autoregressive parameters (AR), d is the number of differences to stabilize the time series (I) and q informs the number of moving average parameters (MA).

Equations (2.7), (2.8) and (2.9) describe models AR AR(p), i.e., ARIMA (p, 0, 0), models MA MA(q), i.e., ARIMA (0, 0, q), and the Autoregressive moving average models ARMA (p, q), i.e., ARIMA (p, 0, q) (Abonazel; Darwish, 2022).

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad (2.7)$$

$$y_t = c - \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (2.8)$$

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (2.9)$$

Wherein ϕ , θ and ε_t are parameters AR(p), MA(q) and random error or white noise, respectively. Thus, the main idea of this model lies in the assumption that the linearity of the time series oftentimes follows normal statistical distribution

(Farsi *et al.*, 2021), because it was created to analyze and predict stationary data of the time series.

2.2 Seasonal autoregressive integrated moving average

Seasonal ARIMA, or SARIMA, was the second adopted model; it was proposed to predict seasonal time series, based on the ARIMA presentation mode $(p, d, q) (P, D, Q)$, according to which, parameters (P, D, Q) are seasonal equivalents of (p, d, q) that describe the non-seasonal part (Farsi *et al.*, 2021). Hence, SARIMA can be expressed through Equation (2.10):

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)w_t \quad (2.10)$$

Where in P presents the order of the seasonal AR model, D points out the number of seasonal differences, Q refers to the seasonal MA order, s is periodicity, w_t is the white noise value within the period t and B is the counter shift operator (Farsi *et al.*, 2021).

2.3 Exponential smoothing

The Exponential Smoothing model (ETS) was the third model to be adopted; it can be used to predict data without a clear trend or seasonal pattern (Hyndman & Athanasopoulos, 2018). This model continuously revises a prediction based on recent information about estimated data. Therefore, it regards weighed measurements that give an exponentially higher weight to observations about recent predictions than those set for older predictions (Hyndman *et al.*, 2008). This process focuses on three components that, altogether, form a time series (T) , (S) and (E) ; therefore, the association among them is additive or multiplicative, although some variations are observed (Hyndman *et al.*, 2008).

The classification of these models starts with the trend component, which regards the combination of a level term (l) to a growth term (b) . These terms can be

combined into five future-trend types (Hyndman *et al.*, 2008). T_h is a trend predicted at the next h periods and φ is a damping parameter ($0 < \varphi < 1$); The five possible trend specifications used in the ETS framework are presented in Equations (2.11) to (2.15). Equation (2.11) represents the absence of trend, Equation (2.12) represents the additive trend, Equation (2.13) represents the damped additive trend, Equation (2.14) represents the multiplicative trend, and Equation (2.15) represents the damped multiplicative trend. (Hyndman *et al.*, 2008).

$$\text{None: } T_h = l \quad (2.11)$$

$$\text{Additive: } T_h = l + bh \quad (2.12)$$

$$\text{Smoothing Additive: } T_h = l + (\varphi + \varphi^2 + \dots + \varphi^h)b \quad (2.13)$$

$$\text{Multiplicative: } T_h = lb^h \quad (2.14)$$

$$\text{Smoothing Multiplicative: } T_h = lb^{(\varphi + \varphi^2 + \dots + \varphi^h)} \quad (2.15)$$

A seasonal component and an error component are introduced after the trend component is defined; it is done in an additive or multiplicative way (Hyndman *et al.*, 2008). In case the error is ignored, there will be fifteen exponential smoothing methods involving both the association trend and the seasonal components (Hyndman *et al.*, 2008).

2.4 Holt-winters

Holt (1957) expanded his studies on simple exponential smoothing to predict data with trends; he created the so-called Holt method. Then, along with Winters (1960), he developed the Holt-Winters method to capture seasonality (Hyndman *et al.*, 2008) - this was the fourth model adopted for the present study. This is the additive model by Holt-Winters with additive trend and seasonality. Their multiplicative method is based on additive trend and multiplicative seasonality, and it is seen as one of the models used in exponential smoothing to disregard the error (Hyndman & Athanasopoulos, 2018).

The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations — one for the level ℓ_t , one for the trend b_t , and one for the seasonal component s_t , with corresponding smoothing parameters α , β and γ (Hyndman & Athanasopoulos, 2018). The Holt-Winters additive and multiplicative formulations are presented in Equations (2.16) and (2.17), respectively. Equation (2.16) represents the additive formulation, in which the seasonal effect is added to the level and trend components. Equation (2.17) represents the multiplicative formulation, in which the seasonal effect is proportional to the level of the series (Hyndman & Athanasopoulos, 2018):

$$\text{Additive method: } y_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)} \quad (2.16)$$

$$\text{Multiplicative method: } y_{t+h|t} = (\ell_t + hb_t)s_{t+h-m(k+1)} \quad (2.17)$$

Where k is the integer part of $(h-1)/m$, which ensures that the estimates of the seasonal indices used for forecasting come from the final year of the sample. Use m to denote the period of the seasonality (Hyndman & Athanasopoulos, 2018).

2. 5 Assessment measurements

Formalisms ARIMA, ETS, SARIMA and Holt-Winters were used at the data modeling stage and applied to flooded area images based on the detailed assessment of residuals of these images. This procedure was necessary to predict likely changes in the herein assessed flooded area of the waterbody using the best-performing model. A prediction error estimator and the relative quality of the statistical models for a given data set were used to make the scientific analysis of the best model to the herein assessed time series (Pho *et al*, 2019), the so-called Akaike Information Criterion – AIC. This criterion was essential to assess the training data (Cavanaugh; Neath, 2019). This methodology is useful for selecting regression predictors and setting the order of an

ARIMA model. The Akaike Information Criterion used to compare the fitted models is presented in Equation (2.18) (hyndman & Athanasopoulos, 2018):

$$AIC = T \log\left(\frac{SSE}{T}\right) + 2(k + 2) \quad (2.18)$$

In this expression, T is the number of observations adopted to the estimate, SSE imply the sum of squared residuals, and k is the number of predictors in the model. The model with the minimal AIC value is often the best model for the prediction (Hyndman & Athanasopoulos, 2018).

The Shapiro-Wilk test was also used to assess whether an empirical distribution is similar to a normal distribution when the p-value is compared to a 0.05 significance level (González; Cosmes, 2019). Subsequently, the behavior of the series was analyzed through autocorrelation and partial autocorrelation correlograms fed with specification thresholds for significance level (Alsharif; Younes; Kim, 2019).

Predictions for all models were plotted along with the assessed time series. The accuracy of each model was analyzed through prediction-performance measurements. *Root Mean Square Error* (RMSE) was the first adopted measurement; it measures the root of the Mean Squared Error (MSE), which, in its turn, concerns the subtraction of what had happened, of the projected value, and , which led to the error (Hodson, 2022). The RMSE calculation is shown in Equation (2.19).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^{\wedge})^2} \quad (2.19)$$

The benefit of using it lies in the fact that the scale of the error changes to correspond to the ideal work measurement unit when the square root is calculated since it enables assessing the power of the prediction made for a data set (Kumar *et al.*, 2022). Since RMSE is scale-dependent, there is no universal threshold to classify its value as low or high. In this study, RMSE was interpreted comparatively among the fitted models, considering that lower RMSE values indicate smaller prediction errors in the same unit

of the flooded-area series. Therefore, the model with the lowest RMSE was considered to have better predictive performance, especially when this result was consistent with the MAPE and Theil's U coefficient values. (Hodson, 2022).

The second adopted measurement was the Mean Absolute Percentage Error (MAPE), which measures the mean absolute percentage error by dividing the difference between the observed value and the predicted value (the result of the error) by the observed value (Elsaraiti; Marabet, 2022) – the signal is ruled out. This calculation is shown in Equation (2.20).

$$MAPE = \frac{1}{n} \sum \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2.20)$$

Finally, Theil's U coefficient, which allows the relative comparison between formal prediction methods and simplistic methods, was used (Lopez-Garcia; Martinez-Damian; Arana-Coronado, 2022). This measurement is defined according to Equation (2.21).

$$U \text{ de Theil} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i)^2} + \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i)^2}} \quad (2.21)$$

If Theil's coefficient is equal to 0, the prediction is perfect; if it is equal to 1, the accuracy of the assessed prediction method is equal to that of the simplistic random walk method, whose prediction for the next instant in time meets the last observation of the series. In case it is higher than 1, the model is not useful for predictive purposes (Lopez-Garcia; Martinez-Damian; Arana-Coronado, 2022). Therefore, it is possible to observe that this relative prediction measurement embodies non-negative values.

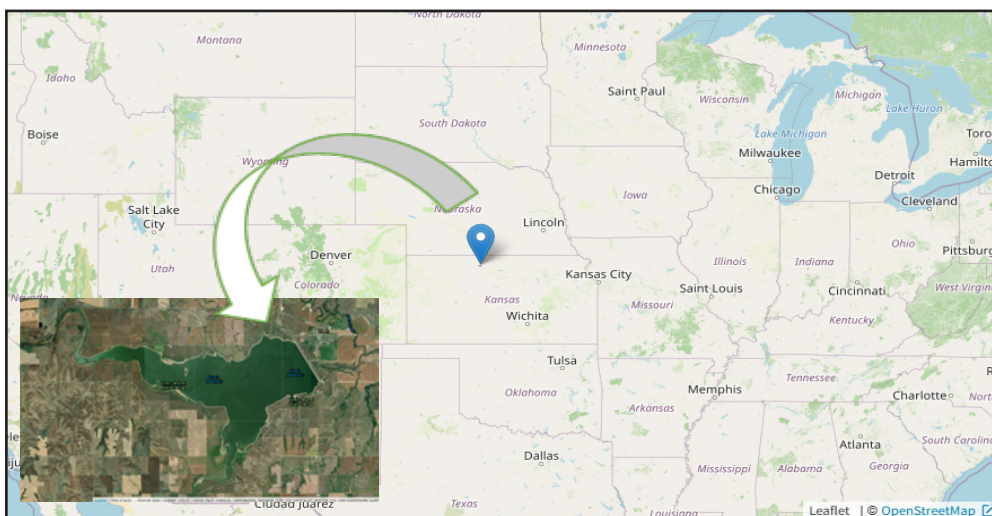
3 MATERIALS AND METHODS

The present study is directed to the collection of images in the region of the Kirwin reservoir, in Phillips County, Kansas, USA. Kirwin Lake (USA) was the water body selected for application, which was monitored monthly between 2013 and 2022. Lake images

were taken at each assessed period, and its flooded area was estimated through the Otsu method, based on information provided by the resolution of the selected images. Then, the time series models guided by the image processing of the lake were adopted to describe its temporal behavior and to predict the future flooded area. Autoregressive integrated moving average (ARIMA) (Hyndman & Athanasopoulos, 2018), seasonal ARIMA (SARIMA) (Box et.al, 2015), exponential smoothing (ETS) (Hyndman et.al, 2008), and Holt-Winters (Hyndman & Athanasopoulos, 2018) were the predictors taken to comparison. Holt-Winters recorded the best performance rates.

As shown in Figure 1, it is located close to Kirwin City, at coordinates 39°39'22"N 99°09'40"W, Northern Kansas (USGS, 2022). The Kirwin National Wildlife Refuge is located on the lakeshore" ou "around the lake, and is one of the most important wildlife preservation units in the USA (Jorgensen; Stullken, 1981).

Figure 1 – Kirwin Reservoir, Kansas State, USA



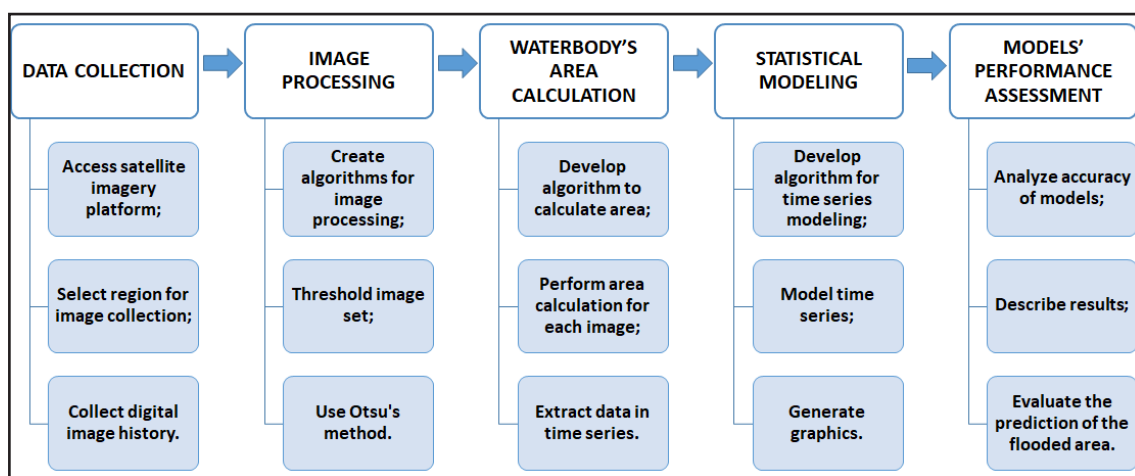
Source: Authors (2023)

The U.S. Bureau of Reclamation built this refuge and still runs it to control floods and irrigation in the region. The refuge has a built area of 22.55 Km², and a capture area of 3,540 Km² (Jorgensen; Stullken, 1981).

3.1 Proposed methodology

The herein developed methodological trajectory is divided into 5 stages: (i) data collection, (ii) image processing, (iii) waterbody's area calculation, (iv) statistical modeling, and (v) models' performance assessment. Stages (ii), (iii), (iv), and (v) were carried out in R language (R core team, 2022). These stages are introduced in the flowchart depicted in Figure 2.

Figure 2 – Methodological trajectory

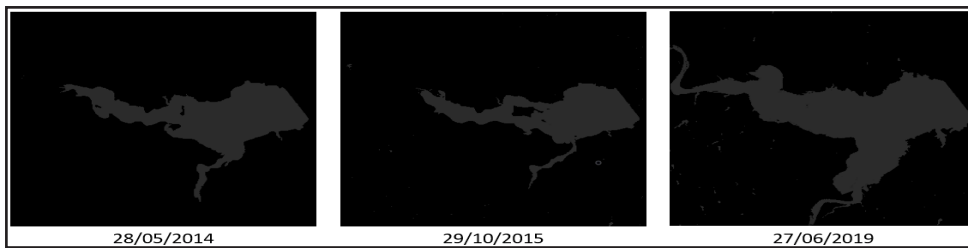


Source: Authors (2023)

3.2 Image database

Landsat satellite images taken from the collection of scientific products were used for dataset selection purposes; these images are available in the section associated with the extension of surface water. The rate of 10% cloud coverage and shading was included as additional criteria. Results are requested after all these parameters are inserted and, subsequently, it is possible to download images in DIAG.TIF format, which is ideal for the Otsu method, because it is expressed in a grey scale. Images of the herein assessed reservoir were rescaled because they recorded 5,000mm x 5,000mm in dimension and showed other waterbodies. The ideal dimension for the present study was 350mm x 350mm, because it put the reservoir in the center of the image, as shown in Figure 3.

Figure 3 – Rescaled images of the Kirwin reservoir on site of USGS, at different moments



Source: Authors (2023)

In total, 119 images of the Kirwin reservoir were collected between 2013 and 2022, and it allowed estimation of the flooded area in this region, based on the collected data. An algorithm was built in R software to process the images.

3.3 The Otsu method

An algorithm was created in R software 4.2.1 (R core team, 2022) to meet the image processing stages. The images were saved in a computer drive and processed according to the Otsu method (Otsu, 1979). It is an image segmentation method that analyzes grey image level histograms to find optimum segmentation thresholds. This method sets the optimum segmentation threshold by maximizing maximizing between-class variance, which regards pixel intensity variation in each class (Pandey *et al.*, 2017). It was necessary to create a function that applies image threshold through the Otsu method given the need for mass processing procedures to set the pixels corresponding to the continent and those that set water occupation in the study site. The separation of these pixels allowed changing their color to white in sites forming the flooded area.

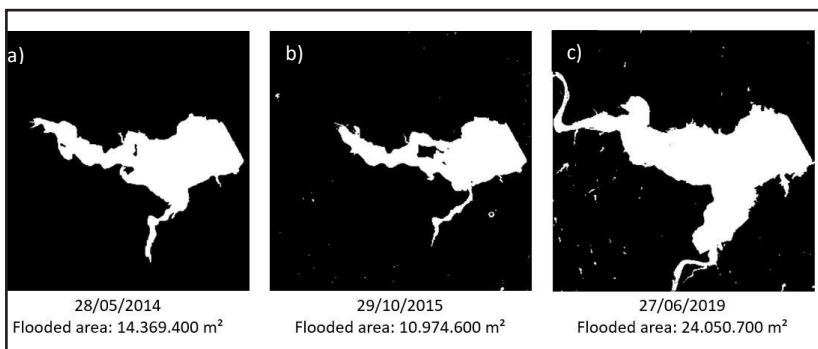
3.4 Estimation of the flooded area

Each pixel records a score of 1, and it enables summing all of them in the sequence of the study. If one takes into consideration that image resolution in the present research was 30m, then, each pixel has an area of 900 m². Thus, the

sum of all pixels, multiplied by this value leads to an estimated flooded area in the waterbody, as shown in Figure 4.

Figure 4a shows a flooded area of 14,369,400 m² in May 2014. In figure 4b, the same evaluated lake region presented a flooded area of 10,974,600 m² in October 2015. In Figure 4c, the Lake Kirwin region, had a flooded area of 24,050,700 m² in June 2019. The application of codes for capturing images allowed changing values of the flooded area expressed in m², every month, starting in January 2013 and ending in November 2022 – allowing the creation of the time series.

Figure 4 – Images of the Kirwin reservoir after application of the Otsu method, at different times. In image a) the flooded area in May 2014 is shown, in image b) the flooded area in October 2015 and in image c) the flooded area in June 2019



Source: Authors (2023)

3.5 Time series models

The image processing algorithm was added to a new script in R language to meet the statistical modeling stage and to carry out the statistical analysis of the time series.

In total, 107 observations were separated for training and 12 for test. It was done to seek modeling to predict changes in the herein assessed flooded area in the waterbody. Thus, it was possible to assess how close the tested models were to the actual results.

Subsequently, data in the study are consolidated and it allows plotting the 12-step ahead prediction based on parameters of the best analyzed model to identify

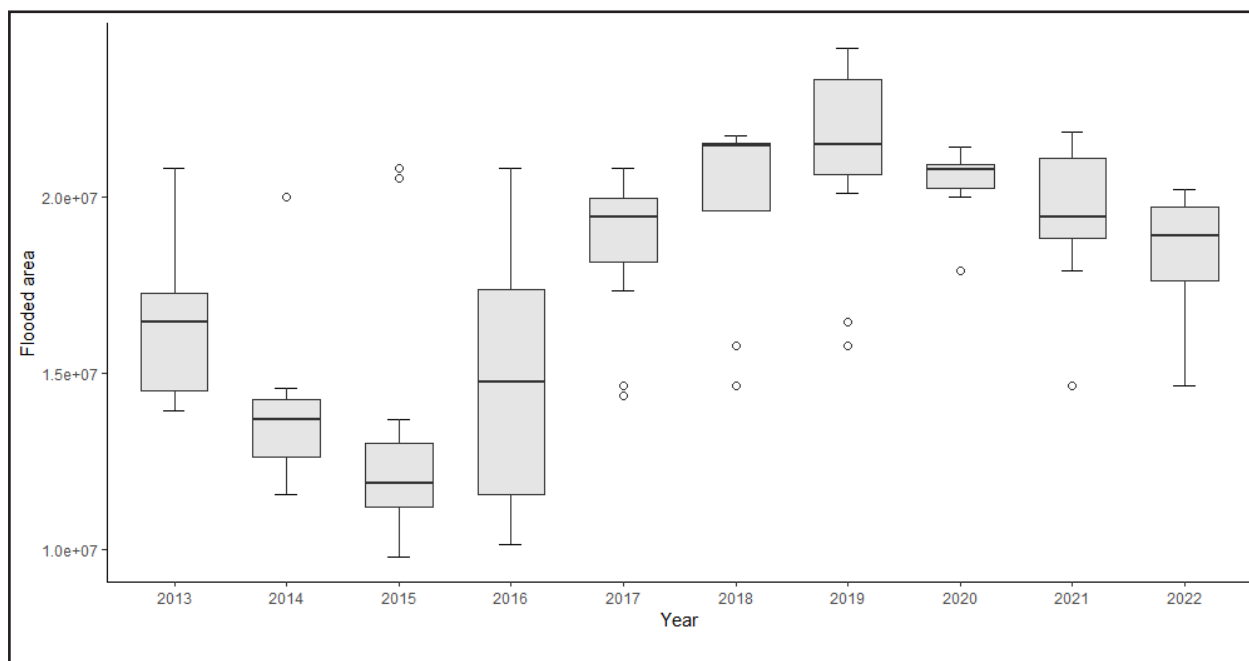
likely changes in the flooded area. It would allow assessing changes in the flooded area analyzed in the current study.

All files used in the present article are available at the repository of the author at https://github.com/Hevertonbarbosa/dados_artigo_PRODER_hbn.git; it contributes to the reproducible research culture.

4 RESULTS AND DISCUSSION

A box plot was created through the annual evolution shown in Figure 5, according to a detailed analysis of the monthly series recorded for the flooded area in the assessed waterbody.

Figure 5 – Box plot of the annual evolution of the flooded area in Kirwin reservoirs.



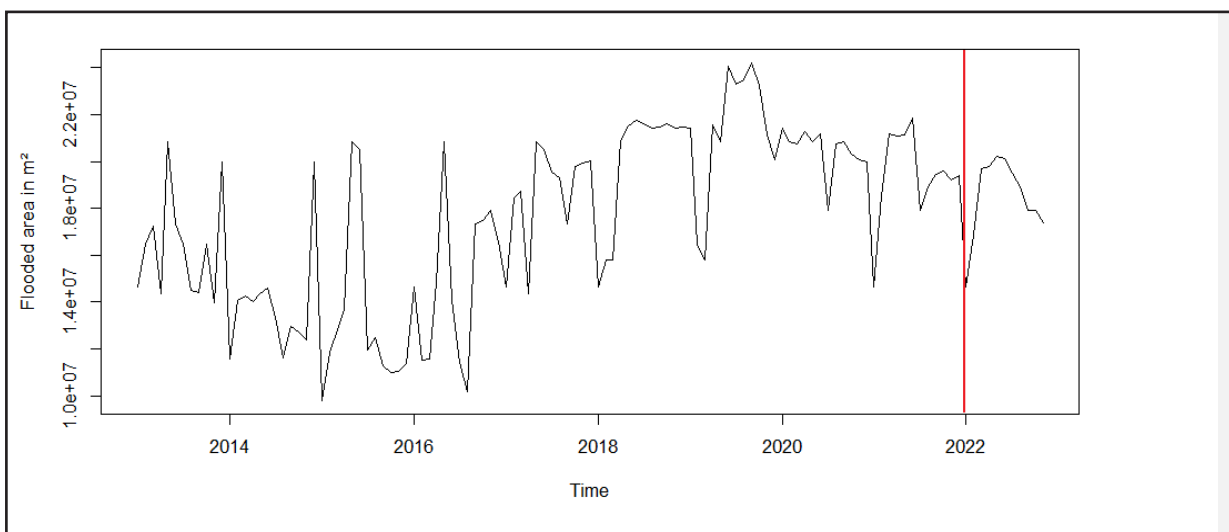
Source: Authors (2023)

It is possible to observe that, from 2016 onwards, there was greater variation in the flooded area of Kirwin Reservoir, followed by an increase in its median values. The median values increased between 2017 and 2019, indicating a relevant change in the flooded area during this period. In 2020, lower variability and a decrease in

median values were observed, and this downward trend remained in 2021 and 2022. Outliers were also observed in 2015, 2017, 2018, and 2019, indicating months with higher flooded-area values in the waterbody.

These variations may be associated with the hydrological and operational dynamics of Kirwin Reservoir. Its water levels fluctuate from year to year according to rainfall, snow runoff, drought periods, and irrigation use (U.S. FISH AND WILDLIFE SERVICE, [s.d.]). In addition, the reservoir is part of an agricultural and irrigation system in the North Fork Solomon River Valley, where water releases below Kirwin Dam are an important source for irrigation and drought conditions can affect water availability (Jorgensen; Stullken, 1981). Thus, the differences observed in the flooded area may reflect the combined influence of seasonal climate variability and water-management demand. Although land-use changes in the contributing basin may also influence runoff and reservoir inflow, this factor was not directly assessed in the present study. Figure 6 sketches the month time series and shows data separation into training and test data; values recorded for the flooded area selected for the test group are observed after the vertical red line. Models were trained with data observed to the left of the vertical line.

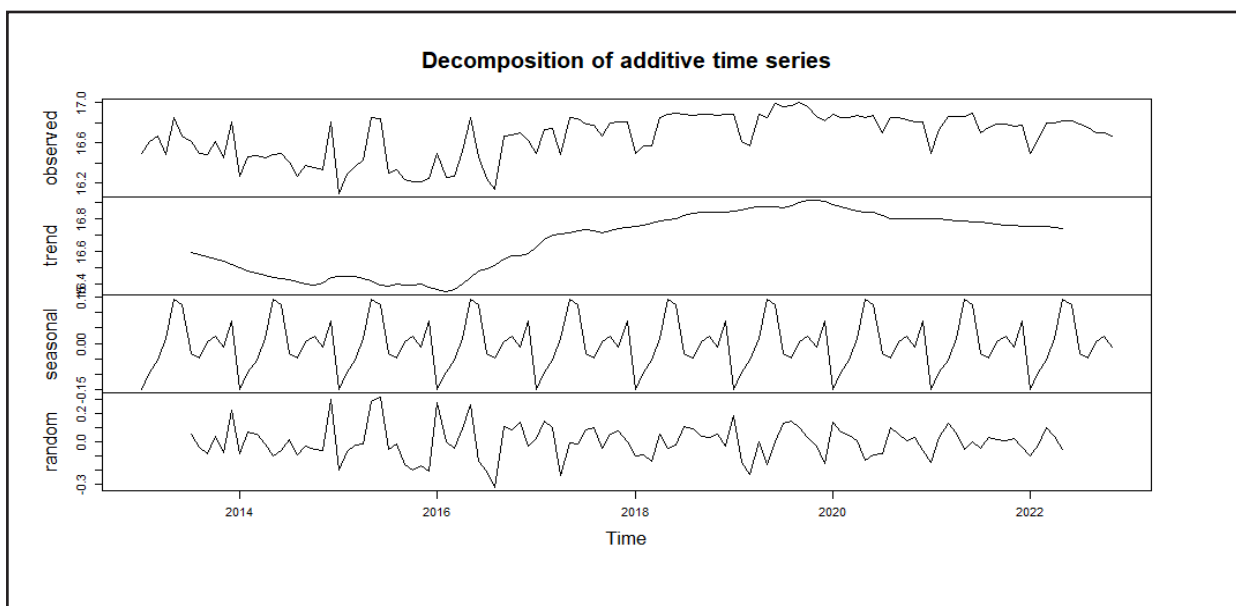
Figure 6 – Time series of the flooded area in Lake Kirwin. The vertical line separates training data (to the left) from test data (to the right).



Source: Authors (2023)

The flooded area showed a mean growing trend within the assessed period. Data logarithmical transformation was carried out to reduce the influence of outliers on the modeling phase. Time series patterns were analyzed in a decomposed way to improve the process of assessing the time series. It was done by decomposing its components, as shown in Figure 7. The trend of this series was extracted, and it presented the growing trend of the flooded area between 2016 and 2020. Then, one can observe annual seasonality in this series, and it reflects the influence of a given external factor. Then, annual seasonality can be observed in the series, which may be associated with hydroclimatic factors, especially rainfall distribution, snow runoff, drought periods, and irrigation-related water use. Kirwin Reservoir is part of an irrigation and flood-control system (Jorgensen; Stullken, 1981). Its water levels may also vary according to rain, snowmelt runoff, and irrigation demand (U.S. FISH AND WILDLIFE SERVICE, [s.d.]). Therefore, the seasonal component may reflect the combined influence of climatic and water-management dynamics. The final section depicts the erratic behavior deriving from interruptions in the assessed series.

Figure 7 – Decomposition of the time series recorded for flooded area in Lake Kirwin



Source: Authors (2023)

4.1 Modeling and prediction

The preliminary analysis of the ARIMA model, based on the autoarima function (R CORE TEAM, 2022), for training data, generated the best ARIMA model, which was that of order (0,1,2) and of the lowest AIC (-79.15). However, the assessment of its residuals – the difference between the original and the adjusted series (ATIQUE et al., 2019) - allowed identifying that it could be enhanced due to the existing significant autocorrelation in lag 12. Accordingly, some relevant information remains in residuals whose projection can be enhanced. It was necessary to work on this model to present the seasonality that was not identified. Thus, seasonality 12 was taken into account and the best seasonal model was assessed by running the autoarima function once more. SARIMA (0,1,2) (1,0,0)₁₂ was the best tested model; it presented an AIC of -83.9, which was the lowest one among all seasonal models tested at this stage. Therefore, the SARIMA model was also used.

Accordingly, and due to seasonality identification, the third model (ETS) was included in the analysis, as well as a fourth model known as Holt-Winters. From this point onwards, all analyses and results were based on assessing the residuals of these models; it was done to highlight the best model among the assessed ones and to allow predicting the flooded area in that region. Table 1 shows the results observed for parameters recorded for each assessed model.

Table 1 – Results of models set for the time series recorded for flooded area

(Continued...)

MODEL	SPECIFICATION	COMPONENT	PARAMETER	ESTIMATE
ARIMA	ARIMA (0,1,2)	Moving average	ma1	-0.595
ARIMA	ARIMA (0,1,2)	Moving average	ma2	-0.185
SARIMA	SARIMA (0,1,2)(1,0,0) ₁₂	Moving average	ma1	-0.185
SARIMA	SARIMA (0,1,2)(1,0,0) ₁₂	Moving average	ma2	-0.166
SARIMA	SARIMA (0,1,2)(1,0,0) ₁₂	Seasonal autoregressive	sar1	0.261

(Conclusion)

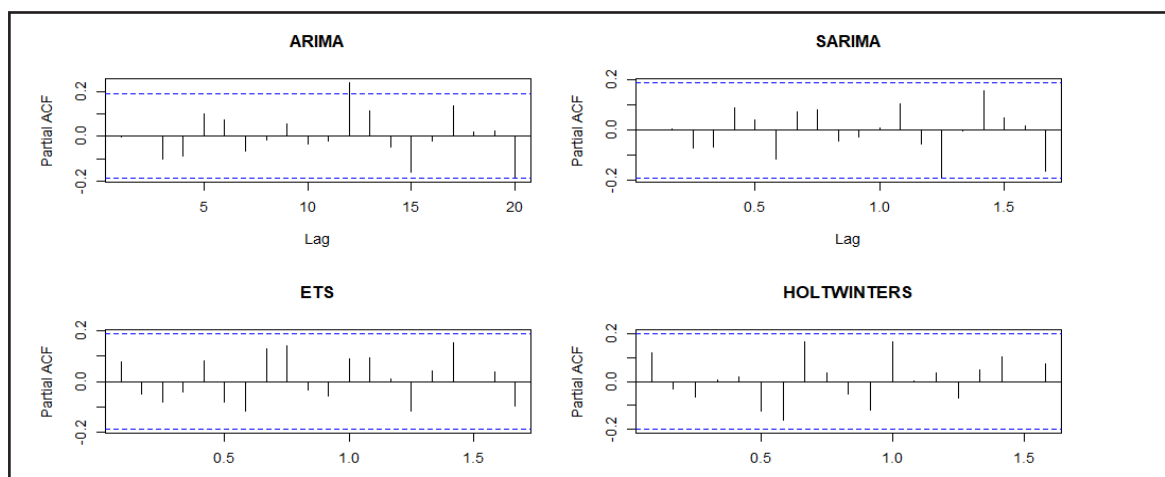
MODEL	SPECIFICATION	COMPONENT	PARAMETER	ESTIMATE
ETS	ETS (A,N,A)	Smoothing parameter	alpha	0.293
ETS	ETS (A,N,A)	Level	l	16.680
ETS	ETS (A,N,A)	Standard deviation	sigma	0.148
HOLT-WINTERS	Additive trend and seasonality	Smoothing parameter	alpha	0.239
HOLT-WINTERS	Additive trend and seasonality	Trend parameter	beta	0.060
HOLT-WINTERS	Additive trend and seasonality	Seasonal parameter	gamma	0.255
HOLT-WINTERS	Additive trend and seasonality	Level	a	16.759
HOLT-WINTERS	Additive trend and seasonality	Trend	b	-0.002

Note: The seasonal components estimated for the Holt-Winters model were: $s_1 = 0.0745$; $s_2 = -0.1294$; $s_3 = -0.0630$; $s_4 = -0.0284$; $s_5 = 0.0411$; $s_6 = 0.1162$; $s_7 = 0.1261$; $s_8 = -0.0040$; $s_9 = 0.0009$; $s_{10} = 0.0231$; $s_{11} = 0.0367$; $s_{12} = -0.0065$.

Source: Authors (2023)

Correlograms were used to enhance the analysis applied to find the best model to analyze the randomness of residuals in each assessed model. Figure 8 presents the correlogram of partial autocorrelation of residuals in the analyzed models.

Figure 8 – Partial autocorrelation of residuals in the used models

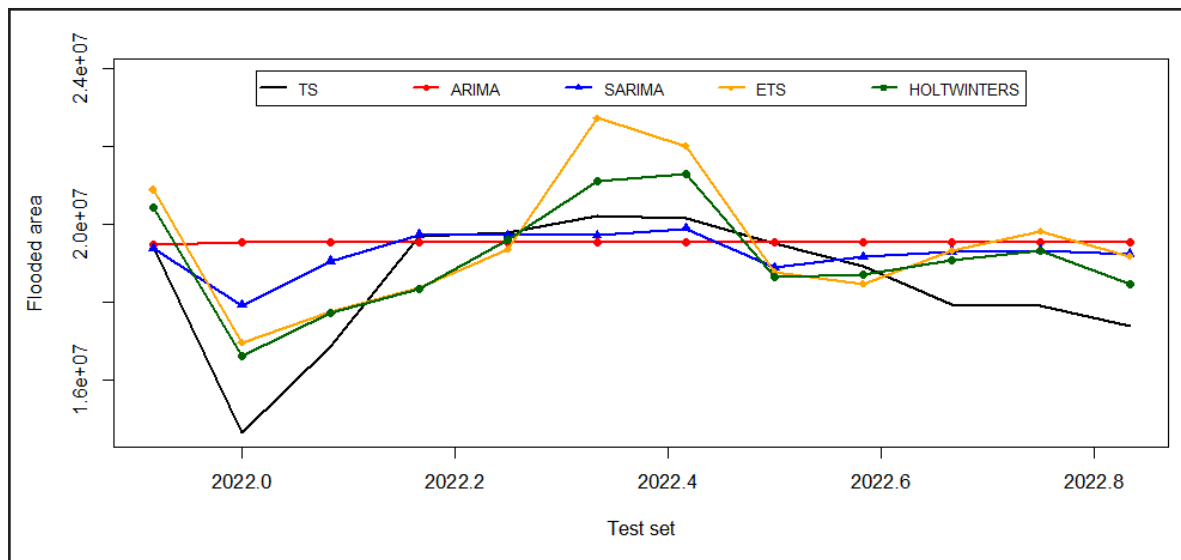


Source: Authors (2023)

Figure 8 shows significant autocorrelations in the residuals of ARIMA and SARIMA models. It also presents an interesting result for the analyses based on the availability of data within the specification thresholds to assess models ETS and Holt-Winters. To better analyze these points, a prediction was generated and compared to data separated for the test. It was done through analysis based on prediction at a time horizon equal to 12 steps ahead.

As observed in Figure 9, predictions generated through SARIMA and Holt-Winters presented performance close to test data of the assessed time series.

Figure 9 – Comparison of predictions for models assessed with test data



Source: Authors (2023)

Some measurements shown in Table 2 were used to analyze the accuracy of the assessed models. It was done to evidence the model accounting for the best performance.

The Holt-Winters model presented the best predictive performance for the test time series. It recorded the lowest RMSE among the evaluated models, indicating the smallest absolute prediction error in the flooded-area unit. This result was also consistent with the lowest MAPE and Theil's U coefficient values, confirming the best relative performance of the Holt-Winters model in comparison with ARIMA, SARIMA and ETS.

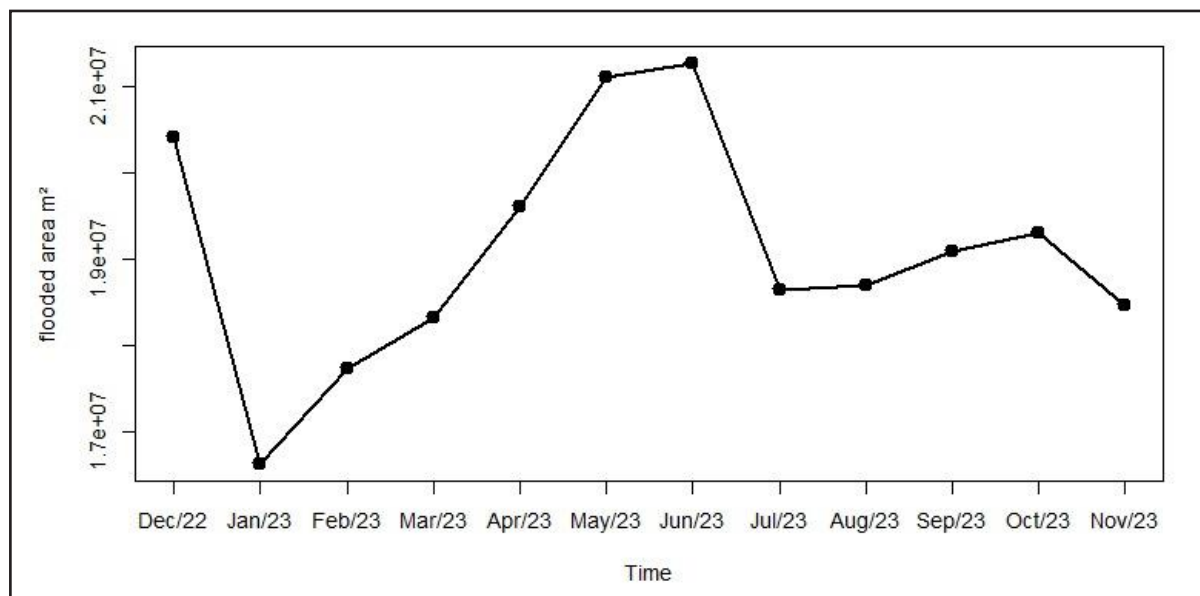
Table 2 – Evaluation measurements applied to the assessed prediction models

MODEL	RMSE (km ²)	MAPE (%)	THEIL
ARIMA	1.879.976	7,672	1,126
SARIMA	1.400.625	5,872	0,625
ETS	1.573.033	7,868	0,730
Holt-Winters	1.118.389	5,669	0,371

Source: Authors (2023)

Accordingly, all data of the time series were taken into account and a prediction was made based on the best model assessed through accuracy measurements. Figure 10 shows the final prediction by taking into consideration the 12-step-ahead time horizon.

Figure 10 – Final prediction for the flooded area by taking into consideration 12 months ahead



Source: Authors (2023)

It is possible to observe that data from the flooded area are similar to the last data collected in the last 12 months. Thus, there was evidence of a low probability of

an increase in flooded area in the following year at Kirwin Lake region. A highlight of this image was observed in January 2023, which demonstrated a notable decrease in the flooded area as a result of Kirwin Lake's presence in the area during the winter, when the majority of it is covered in ice.

The increase projected for May and June 2023 may be interpreted as a seasonal response captured by the Holt-Winters model, associated with the transition from spring to early summer. During this period, rainfall events and runoff may temporarily increase the reservoir surface area, although drought conditions and higher evaporative demand can still limit water availability. This interpretation is consistent with drought reports for May 2023, which indicated recent rainfall and partial drought improvement in northwestern Kansas, but also persistent drought conditions across the state (National Integrated Drought Information System, 2023). In addition, Kirwin Reservoir is part of an irrigation and flood-control system, so its flooded area may also be influenced by water-management demand and irrigation releases (Jorgensen; Stullken, 1981).

5. CONCLUSION

This study proposed a methodological approach to map and predict variations in the flooded area of Kirwin Lake, USA, based on Landsat 8 satellite images collected between January 2013 and November 2022. The proposed procedure combined digital image processing through the Otsu method with time-series modeling, allowing the transformation of multi-temporal satellite images into a monthly flooded-area series.

The results showed evidence of annual seasonality in the flooded-area time series. Among the evaluated models, Holt-Winters presented the best predictive performance, with the lowest RMSE, MAPE and Theil's U coefficient values. This result indicates that the Holt-Winters model was more efficient in representing the seasonal behavior of the flooded area and in producing predictions closer to the observed values.

Therefore, the combination of satellite image processing and Holt-Winters forecasting proved to be a useful strategy for monitoring and predicting flooded-area dynamics in water bodies. The proposed approach can support future studies related to hydrological monitoring, water-resource management and early assessment of possible changes in reservoir surface area. Future research may apply the same methodological structure to other reservoirs and compare its performance with additional forecasting models and environmental variables.

REFERENCES

- Abonazel, M. R., & Darwish, N. M. (2022). Forecasting confirmed and recovered COVID-19 cases and deaths in Egypt after the genetic mutation of the virus: ARIMA Box-Jenkins' approach. *Communications in Mathematical Biology and Neuroscience*, 2022, Article 17. <https://scik.org/index.php/cmbn/article/view/6888>
- Alsharif, M. H., Younes, M. K., & Kim, J. (2019). Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea. *Symmetry*, 11(2), Article 240. <https://doi.org/10.3390/sym11020240>
- Andrade, L. C. de, Lopes, T. O., Medeiros, N. das G., Ferreira, I. O., Santos, A. de P. dos, & Poz, W. R. D. (2024). Estimativa da turbidez da água utilizando imagens de RPA's associadas às técnicas de Machine Learning. *Revista de Geociências do Nordeste*, 10(1), 506–517. <https://doi.org/10.21680/2447-3359.2024v10n1ID34612>
- Atique, S., Noureen, S., Roy, V., Subburaj, V., Bayne, S., & Macfie, J. (2019). Forecasting of total daily solar energy generation using ARIMA: A case study. In 2019 *IEEE 9th Annual Computing and Communication Workshop and Conference* (pp. 114–119). IEEE. <https://doi.org/10.1109/CCWC.2019.8666481>
- Box, G. (2013). Box and Jenkins: Time series analysis, forecasting and control. In *A very British affair: Six Britons and the development of time series analysis during the 20th century* (pp. 161–215). Palgrave Macmillan.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.
- Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics*, 11(3), Article e1460. <https://doi.org/10.1002/wics.1460>
- Elsaraiti, M., & Merabet, A. (2022). Solar power forecasting using deep learning techniques. *IEEE Access*, 10, 31692–31698. <https://doi.org/10.1109/ACCESS.2022.3160484>

- Farsi, M., Hosahalli, D., Manjunatha, B. R., Gad, I., Atlam, E. S., Ahmed, A., Elmarhomy, G., & Ghoneim, O. A. (2021). Parallel genetic algorithms for optimizing the SARIMA model for better forecasting of the NCDC weather data. *Alexandria Engineering Journal*, 60(1), 1299–1316. <https://doi.org/10.1016/j.aej.2020.10.052>
- Fernandes, S., Fialho, A., & Patriarca, I. (2024). Detecção e delimitação de corpos de água em imagens de satélite de alta resolução com aprendizagem profunda. *Revista Internacional Mapping*, 32(214), 10–24. <https://doi.org/10.59192/mapping.442>
- González-Estrada, E., & Cosmes, W. (2019). Shapiro–Wilk test for skew normal distributions based on data transformations. *Journal of Statistical Computation and Simulation*, 89(17), 3258–3272. <https://doi.org/10.1080/00949655.2019.1658763>
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. <https://otexts.com/fpp2/>
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22. <https://doi.org/10.18637/jss.v027.i03>
- Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer.
- Intrieri, E., Raspini, F., Fumagalli, A., Lu, P., Del Conte, S., Farina, P., Allievi, J., Ferretti, A., & Casagli, N. (2018). The Maoxian landslide as seen from space: Detecting precursors of failure with Sentinel-1 data. *Landslides*, 15(1), 123–133. <https://doi.org/10.1007/s10346-017-0915-7>
- Jorgensen, D. G., & Stullken, L. E. (1981). *Hydrology and model of North Fork Solomon River Valley, Kirwin Dam to Waconda Lake, north-central Kansas* (Irrigation Series No. 6). University of Kansas.
- Khan, S. I., Hong, Y., Gourley, J. J., Khattak, M. U., & De Groeve, T. (2014). Multi-sensor imaging and space-ground cross-validation for 2010 flood along Indus River, Pakistan. *Remote Sensing*, 6(3), 2393–2407. <https://doi.org/10.3390/rs6032393>
- Kumar, A., Alsadoon, A., Prasad, P. W. C., Abdullah, S., Rashid, T. A., Pham, D. T. H., & Nguyen, T. Q. V. (2022). Generative adversarial network (GAN) and enhanced root mean square error (ERMSE): Deep learning for stock price movement prediction. *Multimedia Tools and Applications*, 81(3), 3995–4013. <https://doi.org/10.1007/s11042-021-11670-w>
- Lei, Z., & Weng, Q. (2018). Impervious surface estimation by integrated use of Landsat and MODIS time series in Wuhan, China. In Q. Weng (Ed.), *Remote sensing time series image processing* (pp. 121–134). CRC Press.

- López-García, M. del R., Martínez-Damián, M. Á., & Arana-Coronado, J. J. (2022). Predictores del precio de maíz blanco en Jalisco y Michoacán. *Revista Mexicana de Ciencias Agrícolas*, 13(2), 261–272. <https://doi.org/10.29312/remexca.v13i2.2740>
- Muhadi, N. A., Abdullah, A. F., Bejo, S. K., Mahadi, M. R., & Mijic, A. (2020). Image segmentation methods for flood monitoring system. *Water*, 12(6), Article 1825. <https://doi.org/10.3390/w12061825>
- National Integrated Drought Information System. (2023, May 19). *Drought status update for the North Central U.S.* Drought.gov. <https://www.drought.gov/drought-status-updates/drought-status-update-north-central-us-2023-05-19>
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. <https://doi.org/10.1109/TSMC.1979.4310076>
- Pandey, P., Singh, P. K., & Chandra, S. (2017). A review on image segmentation techniques. *Pattern Recognition and Image Analysis*, 27(2), 259–284.
- Pho, K. H., Ly, S., Ly, S., & Lukusa, T. M. (2019). Comparison among Akaike information criterion, Bayesian information criterion and Vuong's test in model selection: A case study of violated speed regulation in Taiwan. *Journal of Advanced Engineering and Computation*, 3(1), 293–303. <http://dx.doi.org/10.25073/jaec.201931.220>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Razzano, F., Di Stasio, P., Mauro, F., Meoni, G., Esposito, M., Schirinzi, G., & Ullo, S. L. (2024). AI techniques for near real-time monitoring of contaminants in coastal waters on board future Φsat-2 mission. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 16755–16766. <https://doi.org/10.1109/JSTARS.2024.3455992>
- Samaraweera, H. U. S. (2024). Exploring complexities of disaster risk and vulnerability: Everyday lives of two flood-affected communities in Sri Lanka. *Sustainable Development*, 32(2), 1376–1385. <https://doi.org/10.1002/sd.2723>
- United States Geological Survey. (2022). *Provisional Landsat 8 Surface Reflectance Code (LaSRC) product (Version 3.3)*. <https://www.usgs.gov/media/files/landsat-8-collection-1-land-surface-reflectance-code-product-guide>
- U.S. Fish and Wildlife Service. (n.d.). Kirwin National Wildlife Refuge: About us. Retrieved May 19, 2026, from <https://www.fws.gov/refuge/kirwin/about-us>
- Xu, Z., & Wang, Y. (2020). Radar satellite image time series analysis for high-resolution mapping of man-made forest change in Chongming Eco-Island. *Remote Sensing*, 12(20), Article 3438. <https://doi.org/10.3390/rs12203438>
- Zhou, S., Kan, P., Silbernagel, J., & Jin, J. (2020). Application of image segmentation in surface water extraction of freshwater lakes using radar data. *ISPRS International Journal of Geo-Information*, 9(7), Article 424. <https://doi.org/10.3390/ijgi9070424>

Authorship contributions

1 – Heverton Barbosa do Nascimento

Bachelor of Mechanical Production Engineering (URCA, 2010). Master of Sustainable Regional Development (PRODER/UFCA, 2023).

<https://orcid.org/0000-0002-4111-7905> • heverton.nascimento@hotmail.com

Contribution: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

2 – Paulo Renato Alves Firmino

PhD in Production Engineering (UFPE, 2009). Professor of Statistics at UFCA since 2014.

<https://orcid.org/0000-0002-3308-2650> • paulo.firmino@ufca.edu.br

Contribution: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – review & editing.

3 – Francisco Alixandre Ávila Rodrigues

PhD in Teleinformatics Engineering (UFC, 2017). Assistant Professor at UFCA.

<https://orcid.org/0000-0002-1675-7775> • alixandre.avila@ufca.edu.br

Contribution: Contribution: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – review & editing.

4 – Jair Paulino de Sales

PhD in Computer Science (UFPE, 2024). Assistant Professor at CCSA/UFCA.

<https://orcid.org/0000-0002-2206-1547> • jair.paulino@ufca.edu.br

Contribution: Validation, Visualization.

5 – Rosilda Benício de Souza

PhD in Biometry and Applied Statistics (UFRPE, 2015). Associate Professor at UFCA.

<https://orcid.org/0000-0002-7433-595X> • rosilda.benicio@ufca.edu.br

Contribution: Supervision, Validation, Visualization.

How to quote this article

do Nascimento, H. B., Firmino, P. R. A., Rodrigues, F. A. A., de Sales, J. P., & de Souza, R. B. (2026). Prediction of flooded area based on time series generated by images of satellites. *Ciência e Natura*, Santa Maria, v. 48, e91144, 2026. DOI: 10.5902/2179460X91144. Available in: <https://doi.org/10.5902/2179460X91144>.