Acta Sci. Pol.
Formatio Circumiectus 25 (1) 2026, 55–69
ENVIRONMENTAL PROCESSES

www.acta.urk.edu.pl ISSN 1644-0765 DOI: http://dx.doi.org/10.15576/ASP.FC/217875

ORIGINAL PAPER Accepted: 06.02.2026

FORECASTING GROUNDWATER LEVELS USING TIME-SERIES MODELS: A COMPARATIVE ANALYSIS OF SNAÏVE, ETS, AND SARIMA APPROACHES IN VIETNAM

Cong Ngoc Phan[1]✉ 0000-0001-7684-9885, Vinh Phu Hoang[1] 0000-0001-7113-2041, Messias de Carvalho[2] 0000-0001-5503-7095

[1] Vinh University, Truong Vinh Ward, Nghe An Province, Vietnam
[2] Polytechnic Institute of University Kimpa Vita, Uíge, Angola

ABSTRACT

Aim of the study
Accurate forecasting of groundwater levels is essential for sustainable water-resource management in data-limited settings. This study develops an operational, reproducible workflow for forecasting a regional shallow-groundwater index based only on historical monitoring records. Monthly groundwater-level observations from six automatic monitoring wells (P1–P6) in the southeastern coastal plain of Nghe An province (Vietnam) were obtained from the Nghe An Environmental Monitoring Center. For each month, the regional series was calculated as the arithmetic mean of the six well levels, providing a single representative indicator for the study area.

Material and methods
Monthly groundwater-level data from May 2013 to April 2025 were analyzed using three forecasting approaches: seasonal naïve (SNaïve), seasonal autoregressive integrated moving average (SARIMA), and exponential smoothing state space (ETS). The dataset was divided into a training period (May 2013–April 2024) and a testing period (May 2024–April 2025). Model performance was assessed using RMSE, MAE, and MAPE, supported by residual diagnostics and the corrected Akaike information criterion (AICc), to ensure model adequacy and parsimony.

Results and conclusions
The ETS model produced the lowest forecast errors and generated residuals closest to white noise, outperforming both SNaïve and SARIMA. These results demonstrate that the ETS offers a robust and reliable framework for forecasting groundwater levels one year ahead . The model's performance provides valuable support for irrigation planning, drought preparedness, and sustainable aquifer management in regions characterized by strong seasonal dynamics.

Keywords: groundwater level forecasting, exponential smoothing (ETS), seasonal ARIMA, seasonal naïve benchmark, Vietnam

✉e-mail: phancongngoc1402@gmail.com

INTRODUCTION

Groundwater is a critical source of freshwater for domestic use, agriculture, and ecosystem services, particularly in coastal plains affected by the monsoon, where the availability of surface water fluctuates greatly throughout the year. In such environments, the timing and magnitude of recharge and abstraction vary seasonally, and shallow aquifers often respond quickly to these changes. Many regions, including central Vietnam, rely on shallow groundwater for irrigation and household use. However, monitoring networks and hydrogeological reports are often limited, fragmented, or inaccessible to local managers. Given these constraints, robust methods for forecasting groundwater levels become an important practical tool for early warning and routine management.

Groundwaterlevel forecasting has been approached using both physically based groundwater flow models and datadriven methods. While numerical models can provide a processbased understanding and scenario analysis, they typically require detailed aquifer geometry, hydraulic parameters, timevarying pumping records, river stages, and longterm meteorological data. Compiling and calibrating such models is resourceintensive and may not be feasible for local agencies operating small monitoring networks. In many provincial monitoring programs, including those in the study area considered here, the most consistently available information is the groundwaterlevel time series itself. Datadriven forecasting therefore plays a complementary role, providing rapid, transparent predictions to support operational decisions when explanatory data are sparse.

Recent studies have demonstrated the usefulness of both classical statistical models and modern machine learning approaches for groundwater forecasting. Reviews highlight the importance of balancing accuracy, transparency, and data requirements when choosing a model, especially for operational deployment (Boo et al., 2024). While deep learning or hybrid models can achieve high accuracy when exogenous drivers are available, classical univariate models remain widely used due to their robustness, interpretability, and computational efficiency. Comparative studies show that exponential smoothing and seasonal ARIMA variants are competitive baselines in datalimited contexts (Sarma and Singh, 2022), and these models remain the standard in large forecasting benchmarks and competitions (Makridakis and Hibon, 2000; Makridakis et al., 2020). Establishing their performance for particular hydroclimatic settings is therefore valuable for practical purposes.

In central Vietnam, the seasonal climate is characterized by a rainy/monsoon season (typically May–October) and a drier season (November–April). This strong seasonal variation can cause recurring annual groundwater fluctuations through cycles of recharge, evapotranspiration, and water demand associated with cropping seasons and domestic use. When such seasonality is dominant and longterm trends are modest, univariate models that explicitly represent seasonal patterns are suitable for routine forecasting, even in the absence of detailed hydrogeological datasets.

This study focuses on the southeastern part of Nghe An province (Vietnam), where groundwater is an important local resource and an automatic monitoring network has been operating for more than a decade. However detailed information on hydrogeological settings (aquifer geometry, hydraulic properties, and groundwater flow paths) is not available for the present analysis. Accordingly, the study adopts the following applicationoriented objective: to develop and evaluate a reproducible forecasting workflow that can be implemented by local agencies using only routinely collected groundwaterlevel observations from a small network of shallow wells.

Rather than attempting to interpret the groundwater flow system spatially, this study constructs a regional monthly groundwater level index by averaging contemporaneous observations from six shallow monitoring wells. This aggregation reduces noise specific to wells (for example, due to local pumping or measurement error) and provides a single indicator that is suitable for tracking and forecasting at regional scale. This approach is intended for operational monitoring and early warning, complementing rather than replacing detailed hydrogeological investigations where such studies are possible.

The objectives of this study are therefore: 1) to characterize trends and seasonality in the regional monthly groundwater level indices; 2) to compare the predictive performance of three classical univariate forecasting methods (SNaïve, SARIMA, and ETS) using consistent training/testing splits and common evaluation metrics; and 3) to provide an evidencebased recommendations for operational models that balance accuracy with transparency, and ease of implementation in a datalimited provincial context.

MATERIALS AND METHODS

Study area

The study area is located in the southeastern part of Nghe An province, Vietnam (Fig. 1), on the lowlying coastal plain near the Lam River estuary. The terrain is generally flat, sometimes gently undulating, with elevations increasing toward the west. It features a mosaic of floodplains and slightly higher natural levees. The land is mostly used for agriculture (rice, cash crops), aquaculture ponds, and expanding urban residential areas. All of these rely partly on shallow groundwater for water supply, especially during the dry season when surfacewater resources are limited.

The climate is influenced by the monsoon, with a distinct wet season (typically May–October) and a dry season (November–April). Most of the annual rainfall occurs in the wet season, whereas potential evapotranspiration remains relatively high throughout much of the year. Seasonal rainfall, evapotranspiration, and waterdemand patterns are therefore imposing a strong annual cycle on shallow groundwater levels, with higher levels during and shortly after the monsoon, and lower levels at the end of the dry season. As detailed hydrogeological information (e.g. aquifer stratigraphy, hydraulic properties, groundwaterflow mapping) is not made available by the monitoring agency, this study treats the monitoring data as an empirical time series and focuses primarily on forecasting performance, rather than on quantitative hydrogeological interpretation.

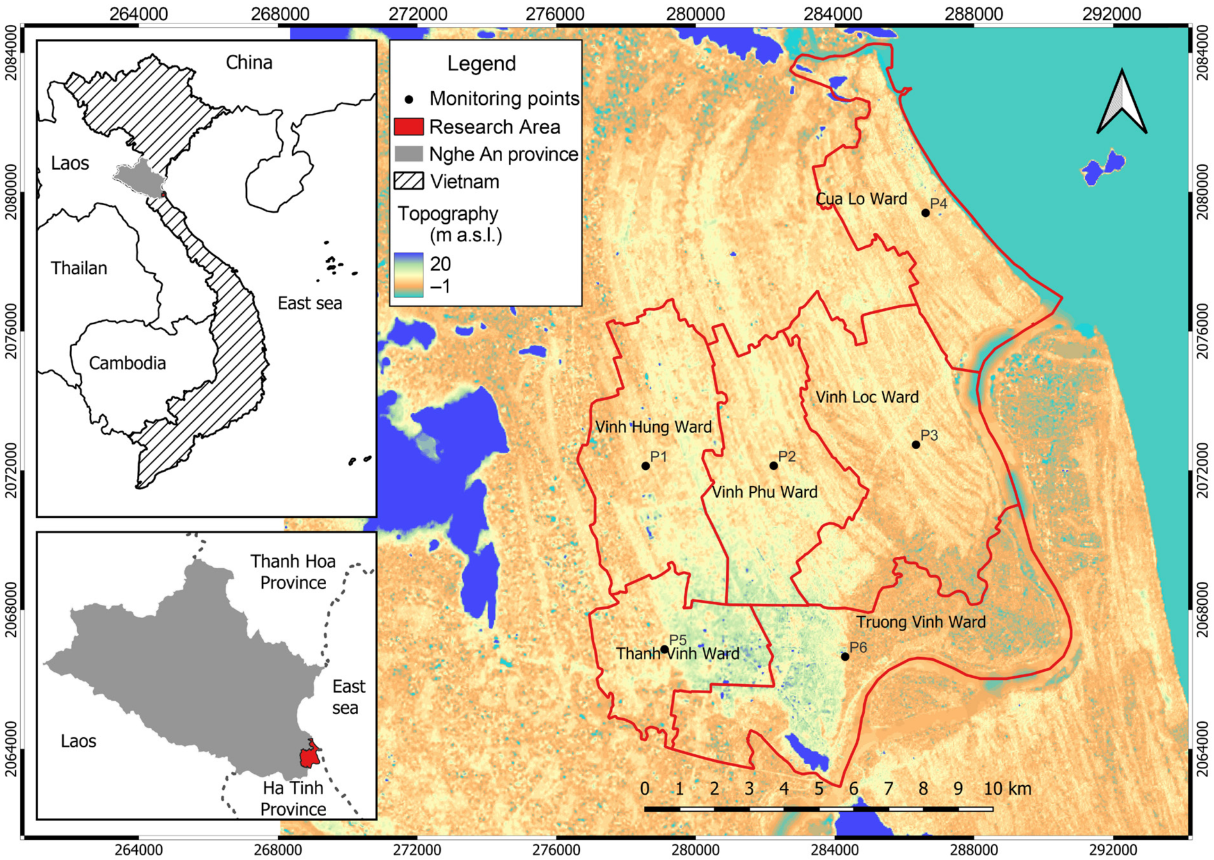

Fig. 1. Location of the study area in Vietnam and positions of the six automatic monitoring wells (P1-P6) overlaid on topography and administrative boundaries; used to define the regional monitoring index (own elaboration)

Data collection and preprocessing

Monthly groundwaterlevel data were obtained from the Nghe An Environmental Monitoring Center for six automatic observation wells (P1–P6, Fig. 1). These wells monitor shallow groundwater, with sensors installed at a depth of approximately 2 m below the ground surface. The depths were referenced to a common vertical datum. The original records consist of submonthly automatic readings, which were quality-controlled by the agency and then aggregated to provide monthly mean groundwater levels for each well.

For each month, a regional groundwater level index was computed as the arithmetic mean of the levels of six wells. This aggregation produces a single monthly time series representing the overall condition of shallow groundwater in the study area. The index is not intended to replace sitespecific analyses, but rather to provide a parsimonious indicator that can be easily forecasted and communicated. Spatial variability among individual wells is treated as part of the uncertainty, which is revisited as a limitation of the approach.

Qualitycontrol procedures included screening for missing values and anomalous spikes. Missing monthly means (less than 2% of records) were infilled using linear interpolation, which is reasonable given the strong temporal persistence of groundwater levels at monthly scales. Potential outliers were flagged using a robust Hampel filter (median ± 3 median absolute deviations) and then reviewed visually in the context of neighbouring months. Confirmed anomalies, which likely resulted from sensor malfunction or datalogging errors rather than genuine hydrological extremes, were replaced using linear interpolation between adjacent months. All preprocessing steps were applied at the level of individual wells prior to computing the regional mean, ensuring that artefacts at a single well did not propagate directly into the index.

To investigate the temporal structure of the regional index, SeasonalTrend decomposition using Loess (STL) was employed in order to separate trend, seasonal, and remainder components (Fig. 3). Longterm trend was assessed using a modified Mann–Kendall framework that accounts for serial dependence in monthly data (Hirsch and Slack, 1984; Hamed, 2008), together with Sen's slope estimator (Sen, 1968). Seasonal effects were evaluated using the Friedman ranksum test (Friedman, 1937). Finally, the remainder component was examined using autocorrelation diagnostics and the Ljung–Box test to check whether substantial structure remained unexplained after removing trend and seasonality.

For model development, the dataset was partitioned into training and testing sets. The training set covered the period from May 2013 to April 2024 (132 months), and was used for parameter estimation and model fitting. The testing set covered the period from May 2024 to April 2025 (12 months), and was used for outofsample validation. This split mirrors a realistic forecasting scenario, where models are updated annually and then employed to predict the subsequent year.

Data preprocessing, statistical analysis, and forecasting model implementation were carried out in the R programming environment. Using widely available packages (such as forecast) facilitates reproducibility and the potential transfer of the workflow to other regions. The overall research workflow is summarized in Figure 2.

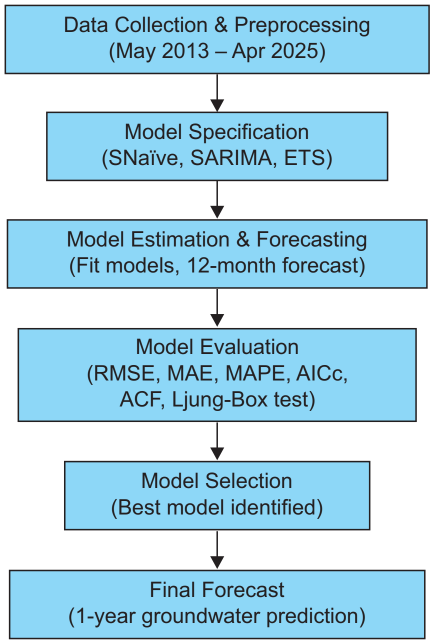

Fig. 2. Analysis workflow: data aggregation and quality control, decomposition and trend/seasonality tests, model fitting (SNaïve, SARIMA, ETS), and out-of-sample validation with residual diagnostics (own elaboration)

Model specification

Three classical univariate forecasting models were selected based on their widespread use in operational settings, the fact that they only require historical groundwaterlevel observations, and provide transparent forecasts suitable for routine monitoring:

1) *Seasonal naïve* (*SNaïve*) – a simple benchmark that assumes each month repeats the value observed in the same month of the previous year (Hyndman and Athanasopoulos, 2018). SNaïve often performs surprisingly well for strongly seasonal series, and therefore serves as a meaningful baseline against, which more complex models must demonstrate improvement.
2) *Seasonal autoregressive integrated moving average* (*SARIMA*) – a Box–Jenkins model that captures autocorrelation and seasonality through nonseasonal orders (p, d, q) and seasonal orders (P, D, Q) with seasonal period m = 12 for monthly data (Box et al., 2016). Differencing orders (d, D) remove nonseasonal and seasonal trends, while autoregressive and moving average terms model persistence and shock propagation in the residual series.
3) *Exponential smoothing state space* (*ETS*) – a state-space formulation that represents level, trend, and seasonality with additive or multiplicative error structures (Hyndman et al., 2008; De Livera et al., 2011). ETS models are particularly well-suited to series with strong, regular seasonality, and they have been shown to perform competitively in largescale forecasting experiments.

The chosen models therefore comprise a simple seasonal benchmark (SNaïve), an autoregressivemovingaverage family (SARIMA), and a structural decomposition family (ETS), providing a balanced comparison of classical univariate approaches.

Model estimation and forecasting

Each model was fitted to the training dataset using the forecast package in R. Forecasts were generated for a 12 month horizon to predict one year ahead. Model estimation procedures included automatic parameter optimization for SARIMA and ETS based on information criteria, while the SNaïve model required no parameterization. For the SARIMA, candidate models within a reasonable order space were screened using AICc and residual diagnostics. The final specification SARIMA(3,0,0)(2,1,0)12 balanced parsimony and residual adequacy. For the ETS, the automatic modelselection routine identified an ETS(A,Ad,A) structure with additive errors, damped trend, and additive seasonality.

Model evaluation criteria

Model performance was evaluated over the testing period. Three accuracy metrics – MAE, RMSE, and MAPE – are reported to facilitate comparison with prior forecasting studies (Willmott and Matsuura, 2005; Chai and Draxler, 2014). For model selection, MAE was treated as the primary criterion because it is scale-consistent, has a clear interpretation in physical units (metres), and is less sensitive to occasional large errors than RMSE. RMSE and MAPE are provided as supporting information, particularly for comparison with other studies that emphasize these metrics.

In addition to point forecast accuracy, residual adequacy was assessed using residual timeseries plots, the autocorrelation function (ACF), and the Ljung–Box test (Ljung and Box, 1978). Residual diagnostics were used to prevent the selection of models that fit the test set well, but retain significant temporal structure, as this would suggest that critical dynamics are not captured by the model, which could bias longerterm forecasts or prediction intervals.

Model selection

The final model was selected based on a balance of statistical accuracy, residual adequacy, and interpretability. While the SNaïve model serves as a baseline, SARIMA and ETS were compared in terms of their ability to capture both seasonal and interannual dynamics of groundwater levels. In line with the operational perspective, priority was given to robust outofsample performance, evaluated primarily by MAE, provided that residual diagnostics did not reveal severe misspecification. The model that performed best was then used to generate forecasts for the subsequent year, providing a tool that could be implemented in practice for groundwater management (Akaike, 1974; Burnham and Anderson, 2002).

RESULTS AND DISCUSSION

Trend and seasonality of the regional groundwater-level index

The Seasonal-Trend decomposition using Loess (STL) was applied to separate the groundwater level time series into its three fundamental components – trend, seasonality, and remainder – to provide a clearer view of the underlying temporal structure (Fig. 3). This decomposition helps identify whether variations in groundwater levels arise primarily from long-term changes or from short-term cyclical effects.

The trend component exhibited no consistent upward or downward tendency, indicating the absence of a statistically significant monotonic change over time (Fig. 3). The modified Mann–Kendall test confirmed that no significant monotonic trend existed ($\tau = 0.0155$, $p = 0.988$), and Sen's slope estimator (Sen, 1968) yielded a near-zero rate of change (3.6×10^{-6}; 95% CI: -0.00039 to 0.00055, $p = 0.988$). As Sen's slope provides a robust estimate of the rate of change based on median, its near-zero value suggests that the long-term trend magnitude was negligible. This result is supports the properties of non-parametric trend tests for serially correlated hydrological data (Kendall, 1975; Hirsch and Slack, 1984; Yue et al., 2002; Hamed, 2008).

The seasonal component revealed distinct, recurring fluctuations that reflect intra-annual cycles in groundwater levels (Fig. 3). The Friedman rank-sum test ($\chi^2 = 81.846$, $df = 11$, $p < 0.001$) confirmed statistically significant seasonal variation (Friedman, 1937). This finding aligns with the hydroclimatic conditions of the region, where wet and dry season alternations govern recharge and discharge processes, and supports prior evidence that environmental time series are often dominated by seasonal rather than long-term patterns (Makridakis et al., 1998; Hyndman and Athanasopoulos, 2018).

The remainder component, which captures the residual variability reaining after the trend component and seasonality are removed, displayed noticeable serial dependence. The Ljung–Box test ($\chi^2 = 97.031$, $df = 24$, $p < 0.001$) indicated significant residual autocorrelation (Ljung and Box, 1978), suggesting that temporal dependence was not fully accounted for by the decomposition process. The remainder component from the STL exhibits moderate autocorrelation at short lags, indicating that some short-term structures are not explained by the deterministic seasonal cycle alone. Such residual dependence is consistent with the influence of episodic meteorological events, pumping, and local management practices, all of which can affect groundwater levels over several months

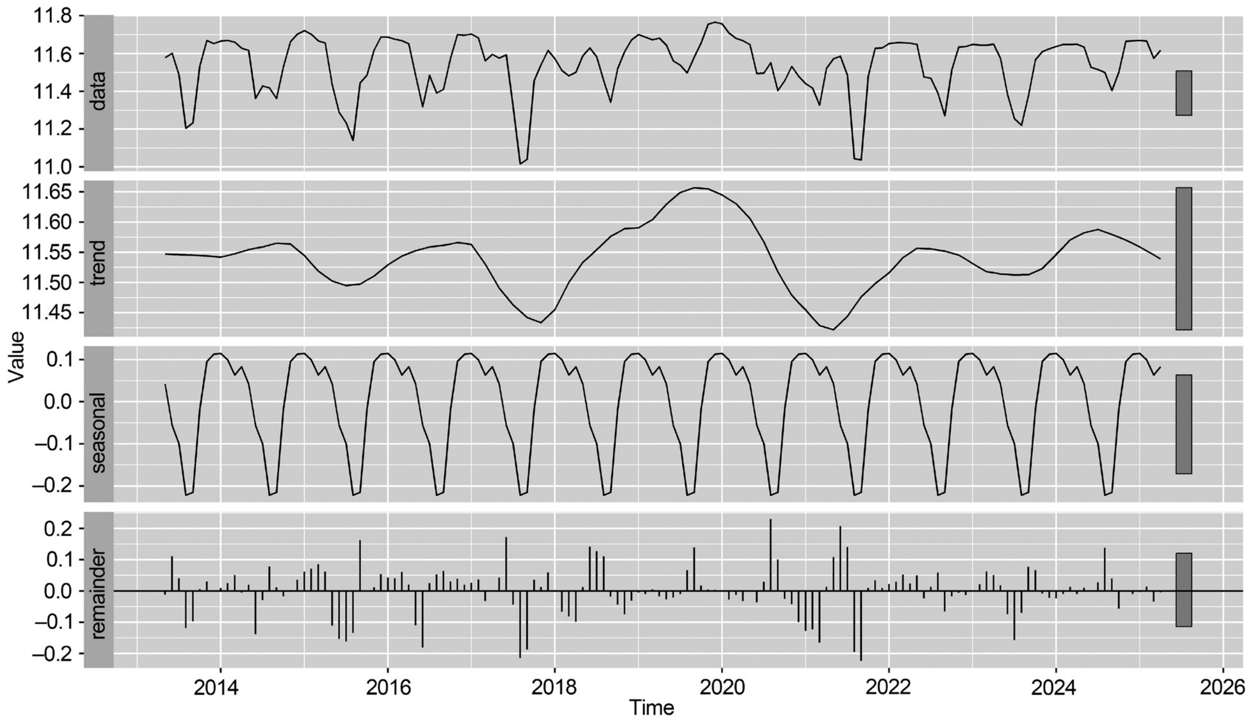

Fig. 3. STL decomposition of the regional monthly groundwater-level index: strong annual seasonality dominates, with a comparatively weak long-term trend component (own elaboration)

(Famiglietti, 2014). These features justify the use of stochastic time-series models that can explicitly represent persistence and shock propagation in the residuals (Chatfield, 2003; Box et al., 2016).

Model comparison and selection

Comparative evaluation of forecasting models

Forecasting model evaluation typically relies on a combination of error metrics and information criteria to assess predictive performance and parsimony. Table 1 presents the comparative results for three widely applied approaches: ETS, SARIMA, and SNaïve.

The ETS (A,Ad,A) ($\alpha = 0.9998$, $\beta = 10^{-4}$, $\gamma = 10^{-4}$, $\varphi = 0.972$) model achieved the lowest forecast errors across all three accuracy metrics. Specifically, the ETS reported a mean absolute error (MAE) of 0.0478, a root mean squared error (RMSE) of 0.0784, and a mean absolute percentage error (MAPE) of 0.4132. These values indicate that the ETS consistently delivered the smallest deviations from observed values, both in absolute and percentage terms. Such performance is consistent with the literature showing that ETS models are particularly effective for time series with strong level, trend, and seasonal components (Gardner, 2006; Hyndman et al., 2008).

The SARIMA (3,0,0)(2,1,0)[12] model, although producing higher forecast errors than the ETS model (MAE = 0.0668, RMSE = 0.0925, MAPE = 0.5833), achieved a substantially lower AICc value (−200.06), indicating a better likelihood-based fit after accounting for model complexity. This outcome reflects the well-established capability of SARIMA models to capture both non-seasonal and seasonal autoregressive dependencies in time-series data (Box et al., 2016). Nonetheless, the relatively larger prediction errors suggest that the SARIMA may be less effective when the data exhibit dominant seasonal patterns, in which case the ETS model can offer a more parsimonious and robust representation of the underlying structure (De Livera et al., 2011).

As expected, the SNaïve model performed poorest across all accuracy metrics, with the largest errors (MAE = 0.0908, RMSE = 0.1251, MAPE = 0.7872). While the SNaïve method is valuable as a benchmark due to its simplicity and interpretability (Makridakis et al., 1998; Hyndman and Athanasopoulos, 2018), its inability to capture either trend or complex seasonal structures was evident. This outcome is consistent with findings from large-scale forecasting competitions, where naïve methods serve as a useful baseline but are regularly outperformed by more sophisticated statistical and hybrid models (Makridakis and Hibon, 2000; Makridakis et al., 2020).

In general, the comparative evidence demonstrates that the ETS provided the most accurate forecasts, while the SARIMA offered the best likelihood-based fit as measured by AICc. The SNaïve model, although inferior in predictive accuracy, remains an important benchmark, against which the effectiveness of more advanced methods can be judged. These results reinforce broader empirical findings that no single method dominates across all criteria. Instead, model adequacy depends on the balance between accuracy, parsimony, and the structural properties of the data (Chatfield, 2000; Fildes and Petropoulos, 2015).

Residual diagnostics for forecasting model comparison

In this study, model adequacy was assessed using a combination of residual variance (Table 1), residual time plots (Fig. 4), and autocorrelation diagnostics based on the autocorrelation function (ACF) and the Ljung–Box portmanteau test (Table 2, Fig. 5), all of which are standard tools for evaluating residual independence (Ljung and Box, 1978; Box et al., 2016; Hyndman and Athanasopoulos, 2018).

Table 1. Out-of-sample forecast accuracy and training set fit statistics for SNaïve, SARIMA, and ETS models applied to the regional monthly groundwater-level index (testing: May 2024–April 2025; training: May 2013–April 2024). Residual denotes the residual variance (sigma²) from the fitted model

Model	MAE	RMSE	MAPE	AICc	Residual
SNaïve	0.0908	0.1251	0.7872	–	0.1601
ETS(A,Ad,A)	0.0478	0.0784	0.4132	41.2966	0.093
SARIMA(3,0,0)(2,1,0)[12]	0.0668	0.0925	0.5833	–200.0639	0.099

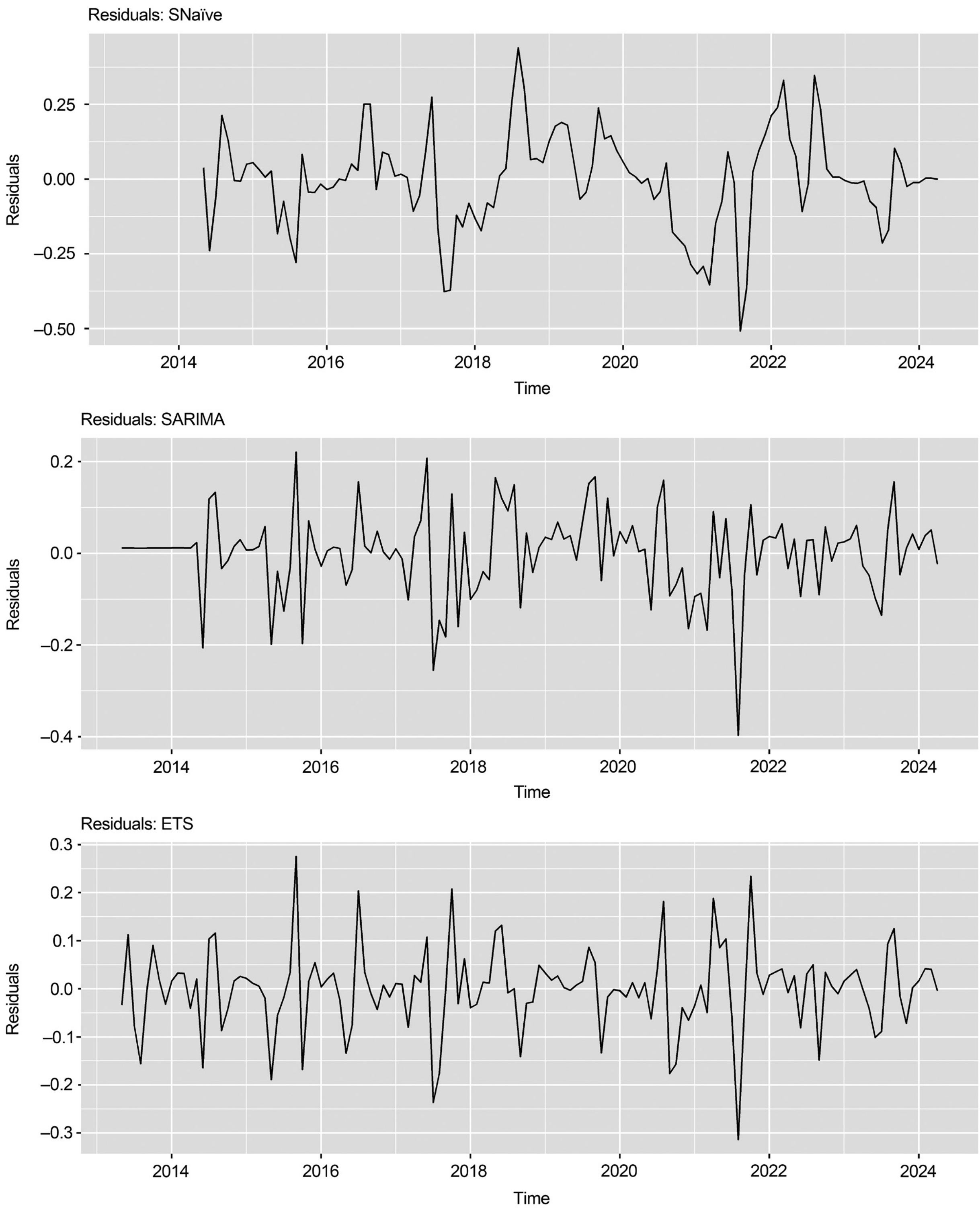

Fig. 4. Residual time-series plots for SNaïve, SARIMA, and ETS models; ETS shows reduced dispersion but some remaining structure (own elaboration)

Table 2. Ljung–Box test results for residual autocorrelation (lag = 24 months) for the three forecasting models. Q* is the test statistic; small p-values indicate remaining serial dependence in residuals.

Model	Q* (Ljung–Box, lag = 24)	df	p-value	Conclusion
SNaïve	178.72	24	$< 2.2 \times 10^{-16}$	Strong autocorrelation remains
ETS(A,Ad,A)	40.92	24	0.0169	Moderate residual autocorrelation
SARIMA(3,0,0)(2,1,0)[12]	21.58	19	0.3058	Weakest autocorrelation, but still present

The SNaïve model exhibited clear signs of inaccuracy across all diagnostics. Its residuals showed large fluctuations and persistent positive and negative deviations, especially during 2017–2018 and 2021–2022 (Fig. 4), reflecting a failure to capture both trends and seasonality. Accordingly, it produced the largest residual variance ($\sigma^2 = 0.1601$), indicating substantial unexplained variation. The ACF revealed strong and persistent autocorrelations at seasonal lags and the Ljung–Box test ($Q^* = 178.72$, df = 24, $p < 0.001$) rejected the null hypothesis of independently distributed residuals with strong confidence. Together, these results highlight the model's inability to account for recurring seasonal dependencies, and confirm that, while SNaïve remains a useful as a simple benchmark (Makridakis et al., 1998; Hyndman and Athanasopoulos, 2018), it is poorly suited to this structured seasonal series. This conclusion is consistent with broader evidence that naïve models generally underperform compared to more sophisticated approaches (Makridakis et al., 2020).

The ETS model substantially surpassed the SNaïve. It yielded the lowest residual variance ($\sigma^2 = 0.093$), with residuals tightly clustered around zero and no evident systematic pattern in the time plot (Fig. 4). This indicates that the model effectively captured the level, trend, and seasonal components of the groundwater-level series (Hyndman et al., 2002; Gardner, 2006). Residual autocorrelation was markedly reduced compared with the SNaïve, confirming the ETS's strength in modeling trend–seasonal structures. However, the Ljung–Box statistic ($Q^* = 40.92$, df = 24, $p = 0.0169$) remained significant, implying the presence of modest but systematic serial dependence. This is consistent with previous findings that ETS models, while flexible and robust, may fail to completely eliminate autocorrelation in complex or highly variable datasets (De Livera et al., 2011).

The SARIMA model displayed an intermediate level of residual variance ($\sigma^2 = 0.099$), with residuals centered around zero, but showing moderate serial dependence around 2019–2020 in the time plot (Fig. 4). In contrast to the ETS, however, the SARIMA performed best in terms of residual independence. Most autocorrelations fell within the 95% confidence bounds of the ACF (Fig. 5), and the Ljung–Box test ($Q^* = 21.58$, df = 19, $p = 0.3058$) did not reject the null hypothesis of white-noise residuals. These diagnostics indicate that the SARIMA model provided the most effective representation of short-term temporal dependence, consistent with its established capacity to model autoregressive and moving-average dynamics (Box et al., 2016). Nonetheless, minor residual structures may still persist, suggesting that further refinement or hybridization could still enhance performance for series with complex seasonal behavior (Athanasopoulos and Hyndman, 2011; De Livera et al., 2011).

Collectively, the diagnostics based on residuals reveal important contrasts in model adequacy. The SNaïve model was clearly inadequate, with high residual variance and strong seasonal autocorrelation. The ETS achieved the greatest reduction in residual variance and yielded visually pattern-free residuals, but retained statistically significant autocorrelation. The SARIMA, by contrast, provided the closest approximation to white-noise residuals according to the ACF and Ljung–Box test, albeit with slightly higher variance than the ETS. Taken together, these findings underscore the value of combining diagnostics based on variance and autocorrelation, when comparing forecasting models, and reinforce the broader methodological consensus that residual analysis offers indispensable validation beyond conventional accuracy metrics (Chatfield, 2000; Fildes and Petropoulos, 2015). They also point towards potential gains from

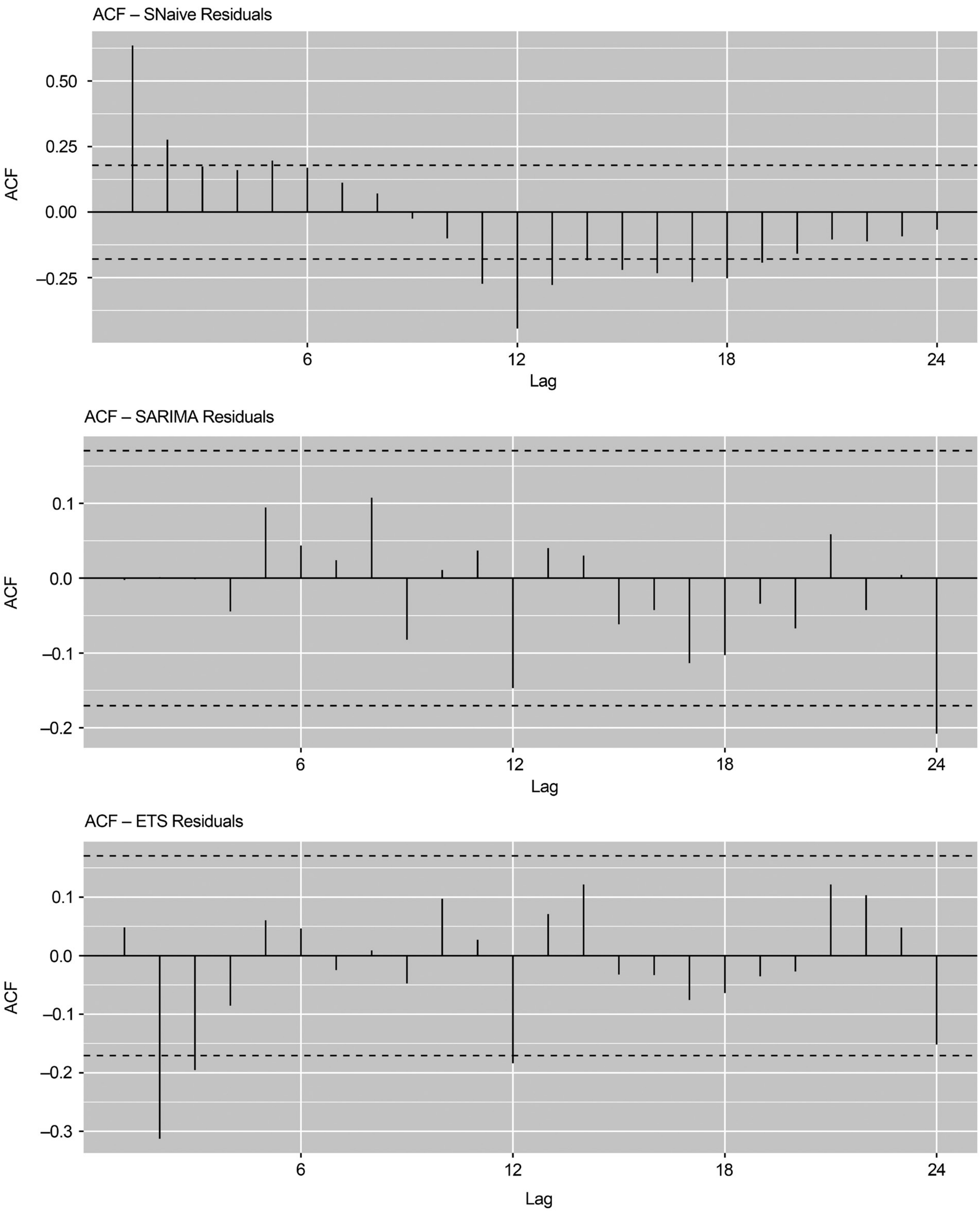

Fig. 5. Residual ACF plots for SNaïve, SARIMA, and ETS models; residual autocorrelation is reduced for ETS and SARIMA compared with the benchmark (own elaboration)

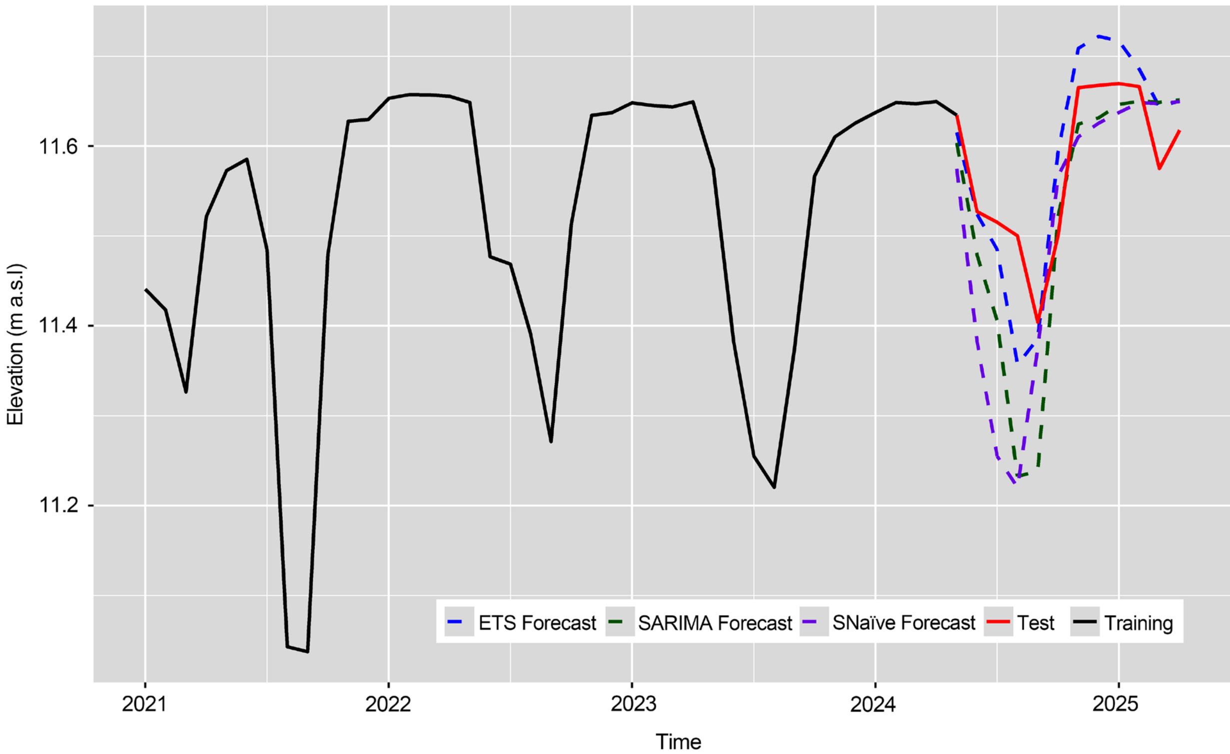

Fig. 6. Comparison of observed and one-year-ahead forecasts from SNaïve, SARIMA, and ETS models on the testing period; the ETS tracks observed seasonal fluctuations with the best accuracy (own elaboration)

more advanced frameworks – such as hybrid SARIMA – ETS models or seasonal-trend decomposition methods, such as STL – SARIMA and TBATS – that can simultaneously exploit the strengths of both exponential smoothing and ARIMA-type structures (De Livera et al., 2011; Hyndman and Athanasopoulos, 2018).

Selection the best model

Considering the primary goal of providing reliable short-term predictions of the regional groundwater-level index, the ETS model was selected for operational reasons, as it achieved the lowest out-of-sample errors while still offering acceptable residual diagnostics. However, the remaining residual autocorrelation suggests that some of the variability is driven by external factors, such as rainfall, pumping, or changes in land use, which are not explicitly represented in the present univariate framework. This is consistent with the findings of recent multivariate and machine learning studies (Wunsch et al., 2021; Sarma and Singh, 2022; Boo et al., 2024).

The preference for ETS is also supported by the broader literature on exponential smoothing. It shows that well-specified ETS models can match or outperform more complex approaches across a wide range of forecasting tasks (Hyndman et al., 2002, 2008; Gardner, 2006; De Livera et al., 2011; Taylor and Snyder, 2012). In this case, the groundwater index exhibits strong, regular seasonality and only a weak long-term trend – a pattern that is particularly well captured by the ETS(A,Ad,A) specification.

From an evaluation standpoint, using MAE as the primary selection metric is consistent with previous studies demonstrating that MAE provides a robust summary of average error and is less susceptible to occasional large deviations than RMSE (Willmott and Matsuura, 2005; Chai and Draxler, 2014). However, reporting all three measures (MAE, RMSE, MAPE) allows for a straightforward comparison with other studies on groundwater forecasting that may emphasise different metrics (Adamowski and Chan, 2011; Valipour et al., 2013; Sarma and Singh, 2022).

Forecasting one year ahead using the selected ETS model

The best ETS (A,Ad,A) model selected to predict the change in groundwater level within the next one year was developed using all available data from May 2013 to April 2025. The prediction results of the ETS model are depicted in Figure 7.

The forecasts indicate a relatively stable groundwater-level pattern between May 2025 and April 2026, with predicted values consistently fluctuating around 11.5 m a.s.l. The narrow confidence intervals (85% and 95%) suggest a high degree of reliability in the model's projections, reflecting its ability to capture the underlying level and trend in the data without significant volatility. This outcome is consistent with the strengths of ETS models, which are widely regarded as effective for series characterized by persistent level components and minimal structural breaks (Hyndman and Athanasopoulos, 2018).

From a methodological perspective, the ETS framework incorporates errors, trends, and seasonality within a state-space formulation, enabling both flexibility and interpretability. In case of our study, the predicted values suggested that the groundwater-level series is largely governed by a stable level component, with no strong evidence of pronounced trend or seasonal shifts. This aligns with prior studies demonstrating that ETS models perform well in stable time series environments with limited structural variability (De Livera et al., 2011).

Furthermore, the relatively small spread between the lower and upper bounds across both the 85% and 95% prediction intervals reinforces the model's forecasting accuracy. Narrow confidence intervals suggest reduced uncertainty and enhanced predictive power, which is particularly important for groundwater management, where reliable forecasts are crucial for planning and the sustainable allocation of resources (Taylor and Snyder, 2012). Nonetheless, it is important to recognize that, like other univariate time series models, the ETS model relies heavily on past patterns. Consequently, its accuracy may diminish in the pres-

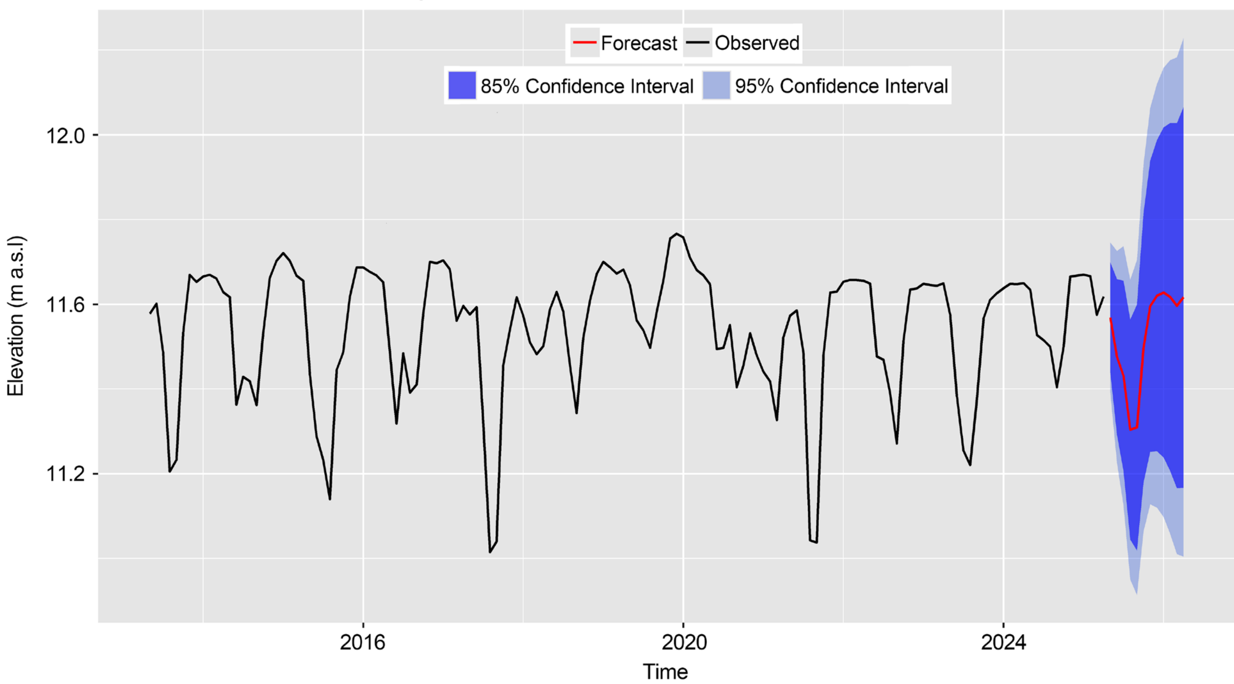

Fig. 7. Twelve-month forecast from the selected ETS model with 85% and 95% prediction intervals, illustrating expected seasonal pattern and uncertainty for operational use (own elaboration)

ence of unexpected shocks, extreme climate events, or anthropogenic disturbances.

Overall, the results of the ETS model provide strong evidence of stability and predictability in groundwater levels within the examined horizon, supporting its suitability for short-term water resource management. Future research could build on this work by comparing ETS predictions with alternative approaches, such as hybrid models (SARIMA – ETS), and machine learning models, to assess robustness under different environmental and hydrological conditions.

CONCLUSION

This study presented an applicationoriented, reproducible workflow for forecasting a regional shallow-groundwater index in southeastern Nghe An province, using only historical monitoring data from six automatic wells. Three classical univariate methods (SNaïve, SARIMA, ETS) were compared under a consistent training/testing split, paying attention to both forecast accuracy and residual adequacy. The analysis was preceded by timeseries decomposition and nonparametric trend tests, which confirmed the dominance of seasonal variability and the absence of a strong monotonic trend in the regional index over the study period.

Among the evaluated models, the ETS model provided the best outofsample accuracy and substantially reduced residual autocorrelation relative to the benchmark. Although the SARIMA model achieved slightly better values for the information criterion and somewhat whiter residuals, these advantages did not translate into superior predictions for the independent test period. Given the applied objective of providing management with reliable forecasts one year ahead, the ETS was selected as the preferred operational model. The resulting ETSbased forecasting system is transparent, computationally efficient, and can be implemented using freely available software, making it accessible to local agencies with limited technical resources.

The proposed approach nonetheless has important limitations. First, the regional index averages spatial variability and therefore cannot replace sitespecific analysis, especially in zones of intensive pumping or strong hydrogeological contrasts. Second, residual diagnostics indicate that purely univariate models cannot fully explain groundwater dynamics, which are influenced by rainfall variability, changes in land use, fluctuations in river level, and anthropogenic activities that are not explicitly represented in the present framework. Third, the evaluation considered a fixed twelvemonth forecast horizon and a single train–test split. More extensive rollingorigin validation would provide a more detailed picture of performance under different hydrological conditions.

Future work should therefore integrate readily available exogenous variables (e.g., rainfall, simple wateruse proxies, or river levels) and explore multivariate or hybrid extensions (such as ETSARIMA error models or machinelearning hybrids), while maintaining operational interpretability. Where data permit, extending the analysis from a regional index to individual or grouped wells, , would also help to quantify spatial differences in forecast skill and groundwater behaviour.

In summary, the ETS model offers a robust, transparent, and practical framework for groundwaterlevel forecasting in Nghe An Province. The methodology and findings presented here may also serve as a reference for similar hydroclimatic regions, where groundwater is monitored but ancillary data are limited. This could support incremental improvements in groundwater resource management under increasing climatic and anthropogenic pressures.

ACKNOWLEDGMENT

We gratefully acknowledge the dataset from the Center for Natural Resources and Environment Monitoring in Nghe An Province, Vietnam.

REFERENCES

Adamowski, J., Chan, H.F. (2011). A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol., 407(1–4), 28–40. https://doi.org/10.1016/j.jhydrol.2011.06.013

Akaike, H. (1974). A New Look at the Statistical Model Identification. In: Parzen, E., Tanabe, K., Kitagawa, G. (eds), Selected Papers of Hirotugu Akaike. Springer Series in Statistics. Springer, New York. https://doi.org/10.1007/978-1-4612-1694-0_16

Athanasopoulos, G., Hyndman, R.J. (2011). The value of feedback in forecasting competitions. Int. J. Forecast, 27(3), 845–849. https://doi.org/10.1016/j.ijforecast.2011.03.002

Boo, K.B.W., El-Shafie, A., Othman, F., Khan, M.M.H., Birima, A.H., Ahmed, A.N. (2024). Groundwater level forecasting with machine learning models: A review. Water Res., 252, 121249. https://doi.org/10.1016/j.watres.2024.121249

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. (2016). Time series analysis: forecasting and control, 5th ed. Wiley, Hoboken. https://doi.org/10.1111/jtsa.12194

Burnham, K.P., Anderson, D.R. (2002). Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York. https://doi.org/10.1007/b97636

Chai, T., Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geosci. Model Dev., 7(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chatfield, C. (2000). Time-series forecasting, 1st ed. Chapman & Hall/CRC, Boca Raton. https://doi.org/10.1201/9781420036206

Chatfield, C. (2003). The analysis of time series: an introduction, 6th ed. Chapman & Hall/CRC, Boca Raton. https://doi.org/10.4324/9780203491683

De Livera, A.M., Hyndman, R.J., Snyder, R.D. (2011) Forecasting time series with complex seasonal patterns using exponential smoothing. J. Am. Stat. Assoc., 106(496), 1513–1527. https://doi.org/10.1198/jasa.2011.tm09771

Famiglietti, J.S. (2014). The global groundwater crisis. Nat. Clim. Change, 4(11), 945–948. https://doi.org/10.1038/nclimate2425

Fildes, R., Petropoulos, F. (2015) Simple versus complex selection rules for forecasting many time series. J. Bus. Res., 68(8), 1692–1701. https://doi.org/10.1016/j.jbusres.2015.03.028

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc., 32(200), 675–701. https://doi.org/10.1080/01621459.1937.10503522

Gardner, E.S. (2006). Exponential smoothing: the state of the art. Part II. Int. J. Forecast., 22(4), 637–666. https://doi.org/10.1016/j.ijforecast.2006.03.005

Hamed, K.H. (2008). Trend detection in hydrologic data: the Mann–Kendall trend test under the scaling hypothesis. J. Hydrol., 349(3–4), 350–363. https://doi.org/10.1016/j.jhydrol.2007.11.009

Hirsch, R.M., Slack, J.R. (1984). A nonparametric trend test for seasonal data with serial dependence. Water Resour. Res., 20(6), 727–732. https://doi.org/10.1029/WR020i006p00727

Hyndman, R.J., Athanasopoulos, G. (2018). Forecasting: principles and practice, 2nd ed. O Texts, Melbourne. https://otexts.com/fpp2/

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D. (2002). A state space framework for automatic forecasting using exponential smoothing methods. Int. J. Forecast, 18(3), 439–454. https://doi.org/10.1016/S0169-2070(01)00110-8

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D. (2008). Forecasting with exponential smoothing: the state space approach. Springer, Berlin. https://doi.org/10.1007/978-3-540-71918-2

Kendall, M.G. (1975). Rank correlation methods, 4th ed. Griffin, London.

Ljung, G.M., Box, G.E.P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303. https://doi.org/10.1093/biomet/65.2.297

Makridakis, S., Hibon, M. (2000). The M3-Competition: results, conclusions and implications. Int. J. Forecast, 16(4), 451–476. https://doi.org/10.1016/S0169-2070(00)00057-1

Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. Int. J. Forecast, 36(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

Makridakis, S., Wheelwright, S.C., Hyndman, R.J. (1998). Forecasting: methods and applications, 3rd ed. Wiley, New York.

Sarma, R., Singh, S.K. (2022). A comparative study of data-driven models for groundwater level forecasting. Water Resour. Manag., 36, 2741–2756. https://doi.org/10.1007/s11269-022-03173-6

Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc., 63(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Taylor, J.W., Snyder, R.D. (2012). Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing. Omega, 40(6), 748–757. https://doi.org/10.1016/j.omega.2010.03.004

Valipour, M., Banihabib, M.E., Behbahani, S.M.R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J. Hydrol., 476, 433–441. https://doi.org/10.1016/j.jhydrol.2012.11.017

Willmott, C.J., Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res., 30(1), 79–82. https://doi.org/10.3354/cr030079

Wunsch, A., Liesch, T., Broda, S. (2021). Groundwater level forecasting with artificial neural networks: a comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). Hydrol. Earth Syst. Sci., 25, 1671–1687. https://doi.org/10.5194/hess-25-1671-2021

Yue, S., Pilon, P., Cavadias, G. (2002). Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. J. Hydrol., 259(1–4), 254–271. https://doi.org/10.1016/S0022-1694(01)00594-7

PROGNOZOWANIE POZIOMU WÓD GRUNTOWYCH PRZY UŻYCIU MODELI SZEREGÓW CZASOWYCH: ANALIZA PORÓWNAWCZA PODEJŚĆ SNAÏVE, ETS I SARIMA W WIETNAMIE

ABSTRAKT

Cel badania

Dokładne prognozowanie poziomu wód gruntowych jest niezbędne do zrównoważonego zarządzania zasobami wodnymi w warunkach ograniczonej dostępności danych. W niniejszym badaniu przedstawiono gotowy do użycia i możliwy do odtworzenia schemat postępowania dotyczący prognozowania regionalnego wskaźnika płytkich wód gruntowych, oparty wyłącznie na historycznych zapisach monitoringu. Miesięczne obserwacje poziomu wód gruntowych z sześciu automatycznych studni monitoringowych (P1–P6) na południowo-wschodniej równinie nadmorskiej prowincji Nghe An w Wietnamie uzyskano z Centrum Monitoringu Środowiska Nghe An. Dla każdego miesiąca regionalną serię obliczono jako średnią arytmetyczną poziomów z sześciu studni, zapewniając jeden reprezentatywny wskaźnik dla całego badanego obszaru.

Materiał i metody

Miesięczne dane dotyczące poziomu wód gruntowych z okresu od maja 2013 do kwietnia 2025 roku poddano analizie przy użyciu trzech podejść prognostycznych: sezonowego modelu naiwnego (SNaïve), sezonowego zintegrowanego modelu autoregresyjnego ruchomego średniego (SARIMA) oraz modelu wygładzania wykładniczego w przestrzeni stanów (ETS). Zbiór danych podzielono na okres treningowy (maj 2013 – kwiecień 2024) oraz okres testowy (maj 2024 – kwiecień 2025). Wydajność modeli oceniano za pomocą wskaźników RMSE, MAE i MAPE, uzupełnionych diagnostyką reszt oraz skorygowanym kryterium informacyjnym Akaikego (AICc), w celu zapewnienia adekwatności modelu oraz jego parsymonii.

Wyniki i wnioski

Model ETS dostarczył najmniej błędów prognoz oraz wygenerował reszty najbardziej zbliżone do białego szumu, przewyższając pod tym względem zarówno model SNaïve, jak i SARIMA. Wyniki te pokazują, że ETS stanowi solidne i wiarygodne narzędzie do prognozowania poziomu wód gruntowych z rocznym wyprzedzeniem. Działanie modelu może stanowić cenne wsparcie dla planowania systemów irygacji, przygotowania na okresy suszy oraz zrównoważonego zarządzania zasobami wód podziemnych w regionach o dużej dynamice sezonowości.

Słowa kluczowe: prognozowanie poziomu wód gruntowych, wygładzanie wykładnicze (ETS), sezonowy ARIMA, sezonowy naiwny benchmark, Wietnam