Summary:

This manuscript proposes an innovative modelling approach - the dual-threshold double autoregressive (DTDAR) model - designed to improve the prediction of daily streamflow under non-stationarity, long memory, and non-linearity in the field of hydrology. By integrating fractional differencing with a threshold-based structure in both the first- and second-order moments, the authors develop a long memory-threshold framework (FDTDAR) that is shown to outperform conventional models (AR-GARCH, TAR-GARCH) at multiple stations across the Yellow River Basin.

## General remarks:

1. The manuscript is well structured, comprehensive in analysis, and the methodology is methodically laid out. The proposed approach makes a meaningful contribution to the field of stochastic hydrological modelling and represents a promising alternative to existing linear and GARCH-based models. However, several aspects require clarification or revision before the manuscript can be recommended for publication.

**Responses:** We sincerely appreciate the reviewer's positive and encouraging evaluation of our work. We are grateful for your recognition of the manuscript's structure, methodological clarity, and contribution to the field of stochastic hydrological modeling. Your thoughtful comments have helped us to further improve the quality and clarity of the manuscript. We have carefully revised the manuscript in response to your questions, which are highlighted in yellow in the latest revision submitted.

2. While the DTDAR/FDTDAR model is a novel approach, it is structurally complex, with numerous parameters and thresholds. The paper would benefit from a clearer explanation of how parameter identifiability, estimation convergence, and computational burden are handled. Practitioners will benefit from a discussion on model tractability and software implementation.

**Responses:** We thank the reviewer for this important and constructive comment. We agree that the DTDAR/FDTDAR model, while novel in its structure, involves a relatively large number of parameters and thresholds, which raises concerns regarding parameter identifiability, convergence, and computational cost. To address this, we have revised the manuscript to provide a more detailed explanation of the estimation procedure.

We have clarified in the revised manuscript the constraints and initialization strategies adopted during the estimation process to ensure parameter identifiability. In addition, we have added descriptions of the convergence criteria and diagnostic methods used in the numerical optimization, as well as details on the software implementation of the model. These revisions are presented on page 13, lines 280-288 of the revised manuscript, specifically as follows: "In this study, the quasi-maximum likelihood estimation under both residual distributions was implemented using the *nlminb* function in R version 4.4.1, which is an efficient numerical optimization tool. It returns the negative value of the likelihood function; therefore, the actual likelihood value used in the estimation is the negative of the function's output. For both residual distribution scenarios, the model orders ($p_{11}$, $p_{21}$, $p_{12}$, $p_{22}$, $q_{11}$, $q_{12}$, $q_{21}$, and $q_{22}$; $p_1$, $p_2$, $q_1$ and $q_2$) were exhaustively searched in the range of 0 to 10. The initial values of the parameters were determined based on the autocorrelation functions at different lag orders up to the corresponding model order. Additionally, the *nlminb* function allows for setting constraints on parameter values, which were specified based on the theoretical requirements of the model. Specifically, the parameters associated with the first-order moments were allowed to vary over the entire real line, while the parameters related to the second-order moments were constrained to be greater than 0.".

3. The manuscript focuses primarily on comparisons with AR-GARCH and TAR-GARCH models. While these are relevant, the absence of modern nonlinear or machine learning models (e.g., LSTM, hybrid deep learning models) in the comparison set limits the extent to which the results can be considered broadly applicable. Even if not implemented, a discussion acknowledging this limitation and the rationale for focusing on DAR-type models would be appropriate.

**Responses:** We thank the reviewer for this insightful comment. We fully acknowledge that the comparison in our study has been primarily limited to traditional linear and

GARCH-based models, such as AR-GARCH and TAR-GARCH. While these models are relevant and widely used benchmarks in stochastic hydrological modeling, we agree that the absence of modern nonlinear or machine learning approaches, such as LSTM or hybrid deep learning models, may limit the generalizability of our conclusions. Our primary focus in this study is to explore the potential of DAR-type models, particularly the proposed FDTDAR structure, as a parsimonious yet effective alternative to time series models. The rationale for this focus lies in the interpretability, mathematical transparency, and relatively low data and computational requirements of DAR-type models, which are particularly important in hydrological applications with limited data availability or operational constraints. We have added a comparison paragraph with modern learning-like models in the "5. Discussion" section of the revised manuscript, although such models are not presented in this paper. Please refer to page 27, lines 516-544 of the revised manuscript for details, which reads:

"5.4 Limitations of FDTDAR models

Compared with modern machine learning models (such as long short-term memory (LSTM) networks or hybrid deep learning architectures), the DAR-type models offer a simpler and more interpretable approach to time series modeling. They are particularly advantageous in scenarios where domain interpretability, limited sample sizes, and computational efficiency are critical. Their transparent structure facilitates parameter estimation, theoretical analysis, and diagnostic evaluation, making them especially suitable for hydrological applications that typically involve noisy data, sparse observations, and operational constraints, thereby enhancing their robustness and reliability in real-world settings (Li et al., 2019; Ling, 2007).

However, the proposed FDTDAR model also has inherent limitations. First, although it can effectively capture regime-switching behavior and certain nonlinear dynamics, it may fall short in modeling the complex and high-dimensional variations commonly present in large-scale hydrological or climate datasets. Second, the modeling capacity of the FDTDAR framework is constrained by the number of thresholds and lag terms that can be practically specified, which limits its structural flexibility compared with more adaptive, data-driven models such as LSTM. Third, the prediction performance of FDTDAR models often depends heavily on the correct setting of thresholds and lag structures. This adjustment process can be challenging in practice and often requires expert domain knowledge or a large number of traversal trials, which may reduce the generalization ability of the model across different regions or datasets.

In contrast, machine learning models, particularly deep learning approaches, can learn complex patterns, nonlinearities, and long-term dependencies directly from the data without strong prior assumptions about model structure (van Cranenburgh et al., 2022). These models often demonstrate outstanding predictive performance in many benchmark tasks. However, their " black-box " nature raises concerns about interpretability and transparency, which are critical in scientific and decision-making contexts such as hydrology (Beven, 2020; Hosseini et al., 2025). In addition, the high computational cost, demand for large training datasets, and sensitivity to hyperparameter tuning may restrict their applicability in resource-limited environments or real-time forecasting systems.

Given these trade-offs, this study focuses on DAR-type models as a theoretically grounded and operationally tractable alternative to traditional AR-GARCH models. While the proposed FDTDAR framework strikes a balance between model simplicity and nonlinear expressiveness, we acknowledge that future research should include comprehensive comparisons with state-of-the-art machine learning models. Such efforts would help further assess the strengths and limitations of the proposed approach and explore the potential of hybrid frameworks that combine the interpretability of statistical models with the flexibility of machine learning techniques.".

4. While the use of average interval width (AIW) and containing ratio (CR) are appropriate to assess the prediction uncertainty, the manuscript lacks detail on how prediction intervals were constructed. Clarification on whether these are based on analytical variance, bootstrapping, or Monte Carlo simulations is necessary.

**Responses:** We agree that a clear explanation of the construction method of the prediction intervals is crucial to assess the reliability of uncertainty analysis. We added explanations about AIW and CR values in the section "3.6 Comparative evaluation methods" in the revised manuscript, but due to space limitations, we did not reflect their specific calculation formulas in the main text, but cited the reference source. The description of lines 324-326 on page 15 of the revised manuscript is: "The interval forecasting performance was evaluated using the Average Interval Width (AIW) and Coverage Rate (CR), with the calculation formulas detailed in Wang et al. (2023b).".

Wang, H., Song, S., Zhang, G., Ayantoboc, O.O., 2023b. Predicting daily streamflow with a novel multi-regime switching ARIMA-MS-GARCH model. Journal of Hydrology: Regional Studies, 47: 101374. https://doi.org/10.1016/j.ejrh.2023.101374

The AIW used in this study is expressed as RIW in this reference, which provides a detailed process for calculating RIW and CR values.

5. The analysis clearly shows that the Student's t-distribution improves predictive performance over the Gaussian assumption. The authors are encouraged to provide more discussion on how degrees of freedom were selected, and whether any skewed or generalized t-distributions were considered or could be more appropriate for heavy-tailed hydrological data.

**Responses:** Thank you for this thoughtful comment. In the revised manuscript, we have added an explanation of how the degrees of freedom parameter was selected. Specifically, the freedom was estimated jointly with other model parameters through maximum likelihood estimation. To ensure numerical stability and meaningful interpretation, we imposed a constraint that freedom must be greater than 2, which guarantees the existence of finite variance. This is described in the revised manuscript on page 13, line 288-289: "The degrees of freedom of the student's t distribution are estimated jointly with the other model parameters, and a constraint of greater than 2 is imposed to ensure numerical stability.".

Regarding considerations for skewed or generalized t-distributions, we agree that such extensions may be more appropriate for modeling asymmetric and fat-tailed behavior in hydrological data. While our current focus is on symmetric distributions, we have added a discussion in Section 5.4 (pages 27-28, lines 545-550) acknowledging this limitation and suggesting that future work could explore skewed or generalized t-distributions to further improve the flexibility and robustness of the model in capturing extreme events. The specific description is: "In terms of the model residuals, although the student's t distribution used in this study effectively captures heavy tails, it does not account for potential asymmetry in the distribution of hydrological residuals. In real-world streamflow processes, especially during flood or drought events, residuals may exhibit skewness in addition to heavy tails. Therefore, more flexible distributions, such as the skewed or generalized t-distributions, may offer a better fit for such cases. While these distributions were not explored in the present study, they represent a promising direction for future research aimed at enhancing the model's adaptability to extreme or asymmetric behavior.".

Apart from these general comments, the authors should also take into consideration a few minor points.

## Minor remarks:

1. Terms such as FDTDAR-n and FDTDAR-t should be introduced earlier and used consistently.

**Responses:** Thank you for pointing this out. In the revised manuscript, we mentioned the expression of FDTDAR-n and FDTDAR-t models in the last paragraph of the Introduction, which are the two most important models in this paper. At the same time, we carefully checked other terms in the text to ensure that they are displayed when they are first mentioned. We also carefully reviewed the manuscript to ensure that these terms are used consistently in all subsequent sections.

2. Several grammatical and syntactic issues are present throughout the manuscript. A round of professional language editing is recommended.

**Responses:** Thanks for your suggestion. We have already asked Professor Thian Yew Gan to help us with professional revisions. Prof Gan is internationally renowned for his many innovative, multidisciplinary contributions to our understanding in hydrology, hydroclimatology, cryosphere, remote sensing of environment, and water resources management. He is a pioneer in research regarding climate change impact to water resources, and developed many practical engineering tools/models for hydrologic forecasting, and innovative algorithms to retrieve large-scale spatial information from remotely sensed data.

3. Recent advances in time series forecasting using deep learning could be briefly referenced to contextualize the DTDAR approach.

**Responses:** Thank you for your valuable suggestion. We agree that briefly mentioning the strengths of deep learning in time series forecasting helps to better clarify the motivation and positioning of the DTDAR approach. In the revised manuscript, we have added a paragraph in the discussion section outlining the rationale for choosing DAR-type models, along with a reflection on their limitations compared to machine learning approaches. We also clarified that, although machine learning methods are powerful, they typically require large datasets, substantial computational resources, and may lack interpretability, which are especially critical in scientific and decision-making contexts such as hydrology. In addition, we objectively acknowledge that future research should include comprehensive comparisons with state-of-the-art machine learning models.