Referee Report

General comments

The manuscript proposes a data-driven framework based on Generalized Additive Models (GAMs) to reconstruct historical irrigation water withdrawals (IWW) across 498 French catchments. The model is calibrated using observations from 2008–2022 and subsequently applied to reconstruct IWW during 2000–2007. The reconstructed time series are then used to investigate long-term trends and their relationships with structural and hydroclimatic factors.

The topic is relevant to hydrological modelling and water resources management, and the availability of consistent historical irrigation withdrawal estimates would certainly be valuable. The manuscript is generally well structured, and the national-scale application addresses a meaningful practical problem.

However, I have substantial concerns regarding both the scientific contribution and the methodological validity of the proposed analysis.

My first concern is the limited novelty of the methodology. The proposed framework is fundamentally based on an established statistical modelling approach, with a decomposition that builds upon the authors' previous work. The introduction largely motivates this choice by discussing the limitations of Random Forests, but these limitations are already well known, and overcoming the shortcomings of a specific algorithm does not, in itself, constitute a methodological contribution. Moreover, the manuscript does not demonstrate, through appropriate comparisons, that the proposed approach provides a meaningful advancement over existing alternatives.

My second concern is that the experimental evaluation does not adequately support the main scientific conclusions. In particular, strong predictive performance across catchments does not necessarily imply accurate reconstruction of temporal variability. More critically, the conclusions about widespread significant trends are derived from model-generated time series, without adequately accounting for reconstruction uncertainty and the assumptions embedded in the model formulation.

Overall, I find that the manuscript does not currently establish sufficient methodological or scientific novelty, while several limitations raise concerns about the reliability of its principal findings. Addressing these issues would require substantially reconsidering the methodological contribution and the experimental analysis, rather than simply clarifying the presentation. For these reasons, I do not recommend publication of the manuscript.

Specific comments

1. Limited methodological novelty and unclear contribution (Introduction and Section 2)

The main methodological contribution appears to consist of adopting a Generalized Additive Model for reconstructing irrigation water withdrawals and separating their structural and interannual variability. However, GAMs are a well-established statistical methodology, and their ability to capture nonlinear additive relationships while maintaining interpretability is extensively documented. Similarly, the decomposition between structural and interannual components builds upon the authors' previous work (Zarpas et al., 2026b). I struggle to identify what constitutes the

actual methodological advancement proposed by this manuscript. The use of a GAM instead of a different regression algorithm does not, by itself, introduce a novel modelling framework.

The spatially structured random effect represents a potentially more specific methodological element, but its formulation largely relies on established regularization concepts and, as discussed below, raises mathematical questions.

Finally, if the intended contribution is primarily the historical reconstruction of irrigation withdrawals rather than a new methodology, the manuscript should demonstrate that this reconstruction provides new and independently supported scientific insights. I do not believe the current results sufficiently establish this contribution either.

2. Questionable motivation based on Random Forest limitations (Introduction, lines 39–53)

The introduction motivates the proposed approach primarily through the limitations of Random Forests in extrapolation tasks. While these limitations are well known, I find this argument insufficient to justify the proposed methodology. Random Forests are only one possible approach to modelling irrigation withdrawals, and their inability to extrapolate beyond the training range does not imply that GAMs provide a reliable alternative under nonstationary historical conditions.

Furthermore, GAMs do not inherently guarantee extrapolation capabilities. Their behaviour outside the training domain depends on the selected smoothers, boundary conditions, and assumptions concerning the relationships between predictors and the target variable.

The authors should explain why a GAM represents an advancement over other established statistical and spatiotemporal modelling approaches. This would also require meaningful comparisons with appropriate benchmarks. Without such evidence, the methodological choice seems largely motivated by known weaknesses of another algorithm rather than by a clearly identified research gap.

3. Temporal reconstruction is not equivalent to temporal forecasting (Sections 2.2 and 2.3)

The manuscript repeatedly emphasizes the temporal extrapolation capabilities of the proposed framework. However, the actual task consists of learning a relationship between IWW and contemporaneous structural and hydroclimatic predictors during 2008–2022, and applying this relationship to historical predictors from 2000–2007.

The model does not explicitly learn the temporal evolution of IWW, nor does it appear to incorporate previous IWW values or an explicit temporal dynamic. Rather, it estimates historical values from explanatory variables that are assumed to retain stable relationships with IWW over time, which is essentially a historical backcasting exercise. Although this is a legitimate modelling objective, it should not be confused with learning a temporal forecasting model.

More importantly, the manuscript provides limited evidence that the estimated predictor–response relationships remain stable between the training and reconstruction periods. This assumption is central to the reliability of the proposed approach, particularly considering potential changes in agricultural practices, irrigation infrastructure, water management, and climatic conditions.

4. Lack of convincing experimental comparisons and evaluation of temporal variability (Sections 2.3 and 4.1)

Despite the emphasis on methodological development, the experimental evaluation does not include a systematic comparison with alternative approaches. In particular, the Random Forest methodology used to motivate the study is not included as a directly evaluated benchmark. Neither are simpler statistical models or reduced formulations of the proposed GAM. Consequently, the reported predictive performance does not establish whether the proposed framework provides any meaningful improvement over existing or simpler approaches.

Furthermore, the main predictive metrics are calculated by pooling observations across all catchments. Since catchments exhibit substantial differences in their average irrigation withdrawals, these metrics may be dominated by the reconstruction of spatial variability rather than the interannual dynamics that the manuscript aims to investigate. A model capable of reproducing catchment-specific average withdrawal levels could potentially obtain strong pooled performance while providing limited information about temporal anomalies.

5. Statistical validity of the trend analysis (Sections 2.4 and 4.4)

A particularly important issue concerns the interpretation of trends estimated from model-generated time series.

The authors fit the GAM using the available observations and subsequently use its predictions over the full 2000–2022 period to estimate long-term trends, reporting statistically significant positive trends in approximately 92% of catchments. However, the predicted values are not independent observations. Their temporal evolution is determined by the selected predictors and the fitted functional relationships, while the uncertainty associated with reconstructing historical IWW is not propagated into the trend significance assessment.

Consequently, the reported p-values do not directly establish the statistical significance of trends in actual irrigation withdrawals. Rather, they characterize trends in the time series generated by the fitted model, conditional on the selected model and predictors. This distinction is fundamental because the identification of widespread positive trends is among the main conclusions of the manuscript.

Moreover, the authors do not adequately establish whether these trends are robust to the inclusion of the reconstructed period or to alternative model specifications.In my opinion, the presented evidence is insufficient to support the claim that widespread positive historical trends in IWW have been reliably identified.

6. The climatic and structural decomposition is largely imposed by the model (Sections 2.1, 2.2 and 4.4.2)

The manuscript emphasizes the capacity of the proposed framework to distinguish structural changes from climate-driven variability in irrigation withdrawals. However, Equation (5) explicitly assigns structural and agricultural predictors to the base component, while meteorological predictors are assigned exclusively to the relative deviation component.

As a consequence, the finding that interannual variability and most reconstructed trends are associated with the relative deviation component is not an entirely independent discovery emerging from the data. In addition, the fitted relationships represent statistical associations rather than

identified causal effects. Important omitted variables, including crop composition, irrigation practices, and regulatory restrictions, could influence both historical withdrawal patterns and their apparent association with climate.

Technical corrections

- **Section 2.3.1:** The text mentions five cross-validation experiments but describes only four configurations.
- **Section 4.1.2:** The descriptions of undercoverage and overcoverage contain repeated experiment labels that should be verified.
- **Section 4.2:** The period 2000–2022 includes 23 years, rather than 22.
- **Appendix A, Equation (A2):** The penalized least-squares objective should be checked for consistency with the log-transformed response and the squared-error formulation.
- **Section 4.4.3:** The notation for the regional trend and IWW indicators appears to be interchanged in the discussion of the Mediterranean region.
- **Throughout:** Several minor grammatical and typographical inconsistencies should be corrected, particularly concerning plural forms, mathematical notation, and terminology.

Final assessment

While the scientific problem addressed in the manuscript is relevant, I do not find the methodological advancement sufficiently established. The approach primarily applies an existing statistical model to a historical reconstruction problem, without demonstrating clear advantages over established alternatives. Furthermore, the main scientific conclusions depend on model-generated trends whose reliability and statistical interpretation remain insufficiently supported.