Agricultural Water Management 333 (2026) 110582

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat

Assessing structural uncertainty in water–human systems: A multi-model ensemble approach for agricultural water management

Héctor González-López [a,b,*], Arthur Hrast Essenfelder [c], Francesco Sapino [a], David Rivas-Tabares [a], José María Bodoque [d], C. Dionisio Pérez-Blanco [a,e]

[a] IMDEA Water Institute, Avda. Punto Com 2, 28805 Alcalá de Henares, Madrid, Spain
[b] University of Alcalá (UAH), Plaza de la Victoria, 2, 28802, Alcalá de Henares, Madrid, Spain
[c] European Commission, Joint Research Centre (JRC), Ispra, Italy
[d] Faculty of Environmental Sciences and Biochemistry, Universidad de Castilla-La Mancha, Spain
[e] Euro-Mediterranean Center on Climate Change, Via della Libertà, 12, Venezia, VE 30121, Italy

A R T I C L E I N F O

Handling Editor -Rodney Thompson

Keywords:
SWAT
HEC-HMS
Hydro-economic modeling
Positive Mathematical Programming
Decision-making
Sustainable water allocation

A B S T R A C T

Coupled hydro-economic models are increasingly used to support agricultural water management, yet their results can be highly sensitive to structural (model-form) choices in both human and water systems. We analyze how structural uncertainty propagates across a modular coupled human–water system model by building a hydrological multi-model ensemble for an irrigated catchment in Spain's Douro basin (Tormes). The ensemble combines alternative hydrological models (Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) and two Soil & Water Assessment Tool (SWAT) configurations) with four mathematical-programming representations of irrigator behavior (two Positive Mathematical Programming models and two Multi-attribute models) through a standardized coupling protocol that translates crop choices and irrigation demand into water-system inputs. We evaluate coupled performance of daily reservoir storage and harmonized diagnostics under two drought scenarios and four volumetric water charge levels based on standardized anomalies (z-scores). Results show that model structural uncertainty cascades through the ensemble and concentrates around irrigation onset, producing storage spreads of approximately 150–202 million m^3 across coupled configurations, equivalent to 31–41% of Santa Teresa's maximum permitted storage and 95–127% of annual downstream irrigation demand. Hydrological model choice explains most of this spread: family median differences reach 1.88–2.12 z-score units, meaning that the central tendency of standardized monthly storage anomalies shifts by almost two standard deviations depending on the hydrological structure selected. Economic-model structure mainly modulates within-family variability through heterogeneous water-demand signals. These findings imply that single-model decision support can underestimate the range of plausible futures and bias policy appraisal. We argue that systematically assessing model structural uncertainty through multi-model ensembles can strengthen robust decision-making on agricultural water management under drought and water charge interventions.

1. Introduction

Water scarcity constitutes one of the most pressing global challenges, with profound implications for ecosystems, economies, and societies (Dolan et al., 2021; Jones et al., 2024). Small environmental or socio-economic changes can create ripples, where human decisions and hydrological processes co-evolve and lead to emergent phenomena that are difficult to predict (Gil-García et al., 2023; Rachunok and Fletcher, 2023). Such coupled human-water dynamics are inherently complex and nonlinear, and their adequate representation necessitates modeling frameworks that integrate relevant biophysical and economic processes as well as their linkages (Knowling et al., 2020; Ortiz-Partida et al., 2023). In this context, the choice of models and their coupling design is a sensitive issue with nontrivial repercussions on model outputs and forecasts, potentially propagating into decision risk through structural uncertainty.

Structural uncertainty captures the variability in system forecasts that stems from alternative yet plausible model formulations, including

* Corresponding author at: IMDEA Water Institute, Avda. Punto Com 2, 28805 Alcalá de Henares, Madrid, Spain
E-mail address: hector.gonzalez@agua.imdea.org (H. González-López).

https://doi.org/10.1016/j.agwat.2026.110582
Received 26 February 2026; Received in revised form 20 May 2026; Accepted 26 June 2026
Available online 29 June 2026

the design of human and water system models as well as their couplings (Gupta and Govindaraju, 2019; Moges et al., 2020). While multi-model ensemble approaches that compare forecasts across alternative models have been used to assess structural uncertainty within human or water system models independently (Hasan et al., 2025; Noh et al., 2024), the propagation of structural uncertainty across coupled human-water systems remains largely unexplored (Dwivedi et al., 2025; Zhou et al., 2025). Neglecting this aspect risks systematically underestimating the range of possible futures and fostering unwarranted confidence in management interventions, particularly in contexts where feedbacks between human and water systems are relevant (Hino and Hall, 2017).

This study addresses this research gap by assembling a cross-system ensemble of widely used human and water system models representing diverse behavioral paradigms and hydrological complexities, and evaluating their coupled performance under drought and water charge scenarios. The framework couples alternative hydrological models (Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) and Soil & Water Assessment Tool (SWAT)) with a set of microeconomic models (two Positive Mathematical Programming (PMP) models (Howitt, 1995; Júdez et al., 2002), and two multi-attribute models: one Positive Multi-Attribute Utility Programming (PMAUP) model (Gutierrez-Martin and Gomez, 2011) and one Weighted Goals Programming (WGP) model (Sumpsi et al., 1997)) via a protocolized coupling, and compares outcomes using standardized diagnostics (z-score anomalies) to analyze structure-driven effects that emerge both from within each model and due to their coupling. We show that changing model architecture yields different and plausible hydro-economic trajectories, including shifts in where and when impacts concentrate. This implies that relying on a single model can mask plausible outcomes and bias uncertainty appraisal. On this basis, we advocate acknowledging structural uncertainty through ensemble analysis and provide an operational approach—models, coupling protocol, and harmonized metrics—that can be reproduced to assess modeling structural uncertainty and inform policy making under uncertainty.

2. Materials and methods

2.1. Study zone description

The Tormes Catchment, located in the southwestern Douro River Basin of Spain near the Portuguese border, encompasses 7107 km^2, of which 45 731 ha are irrigated. Agriculture represents the catchment's principal water use, with an annual demand of 286.78 million m^3, 78.15% of which are supplied by surface water and 21.85% by groundwater. Total water availability amounts to approximately 1300 million m^3 per year and is regulated by two major reservoirs: Almendra, with a capacity of 2586 million m^3, situated at the catchment outlet and primarily serving hydropower while controlling releases to the Douro River, and Santa Teresa, with a capacity of 496 million m^3, which supplies irrigation water across the catchment.

Agricultural water users are divided into seventeen Agricultural Water Demand Units (AWDUs), clusters of farmers sharing water sources, geographical context, administrative arrangements, and hydrological characteristics (Fig. 1). Some upstream AWDUs have historically been more vulnerable to drought due to weaker access to local regulation infrastructure, with intensifying severity in recent years. Downstream AWDUs, supplied by the Santa Teresa reservoir, experience growing water stress driven by climate change. For the downstream AWDUs supplied by Santa Teresa Reservoir, the annual irrigation demand used as a reference in the results is approximately 158.47 million m^3 yr^{-1} . Despite increasing water scarcity concerns, an irrigation expansion of 26 000 ha demanding 10–20 million m3 has been planned for the catchment.

2.2. Hydro-economic multimodel ensemble configuration

A protocol based modular framework is developed where a set of rules or protocols are used to couple hydrological models of the catchment (water system module) and microeconomic models of irrigators' behavior (human system module) into a cross-system multi-model ensemble (Fig. 2), using a protocolized coupling applied consistently across models. The human module is populated with a multi-model

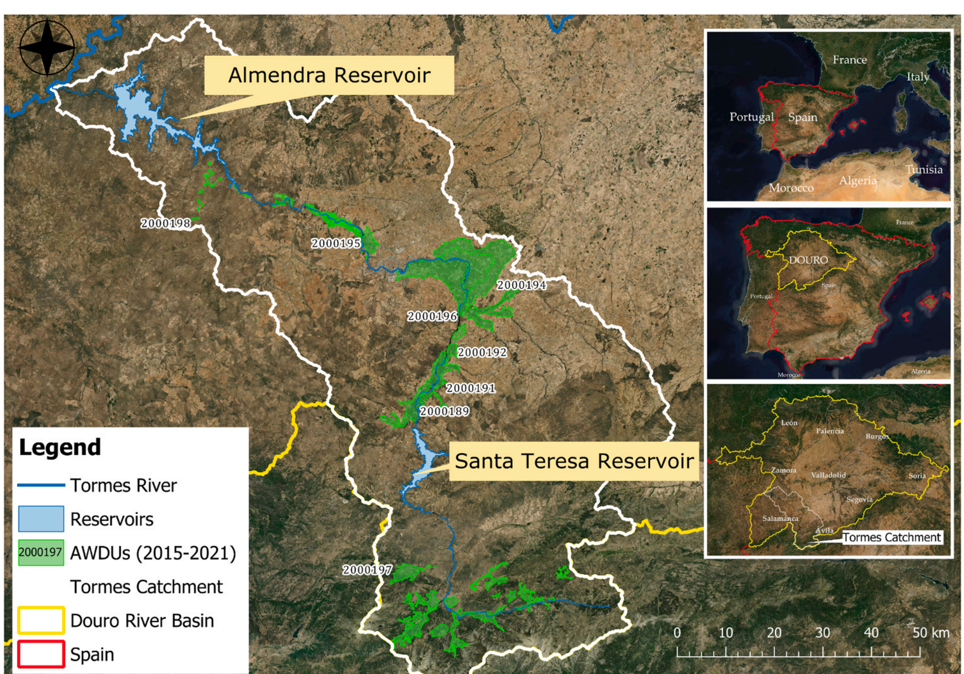

Fig. 1. Location of the Tormes Catchment in the Douro River Basin and its Agricultural Water Demand Units (AWDUs). Numeric codes identify the 17 AWDUs used in the hydro-economic modeling framework.

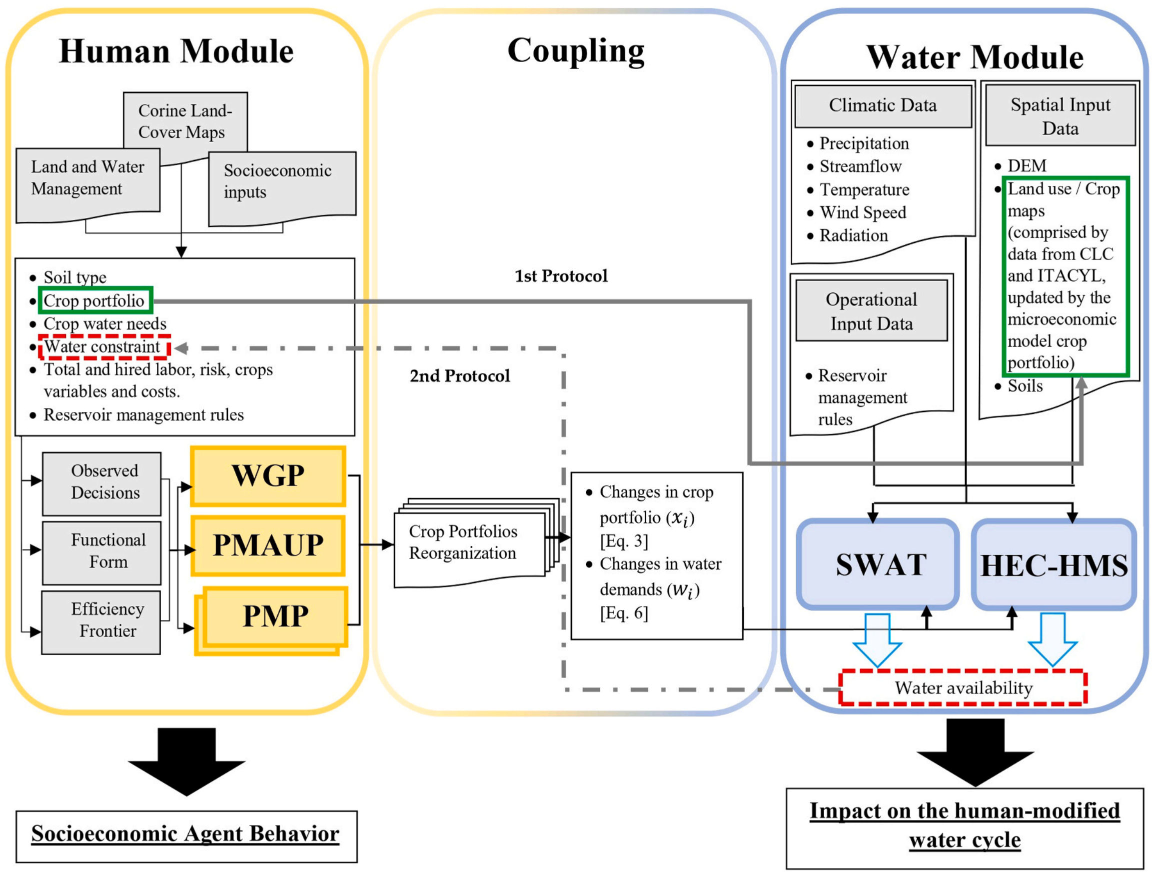

Fig. 2. Workflow of the protocol-based modular framework. The diagram shows the data flow between human and water modules through the two coupling protocols. Solid arrows indicate the protocol triggered in the simulations, whereas dashed arrows indicate the feedback protocol, which is defined in the framework but not triggered since water availability does not become binding in this study.

ensemble of Mathematical Programming microeconomic models, namely two Positive Mathematical Programming models (PMP1 (Howitt, 1995) and PMP2 (Júdez et al., 2002)), one Positive Multi-Attribute Utility Programming Model (PMAUP) (Gutierrez-Martin and Gomez, 2011), and one Weighted Goal Programming model (WGP) (Sumpsi et al., 1997); while the water system is populated with two hydrological models, namely Soil & Water Assessment Tool (SWAT) (Arnold et al., 1998), developing two different structural setups, and Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) (Hydrologic Engineering Center, 2018). Two protocols are defined to interconnect the two modules. The first protocol translates crop portfolio choices and related irrigation demand generated by the human module into land-use and water-demand inputs for the water module. The second protocol can transmit water-availability information generated by the water-system models back to the human-system models by updating water-availability constraints. In the simulations presented here, the feedback protocol is defined but not triggered because water availability does not become binding for the economic module.

2.2.1. Human module

Mathematical programming models represent irrigators as constrained decision-makers who allocate land among alternative crops to reproduce observed behavior and predict responses to exogenous shocks. In this framework, irrigators choose a crop portfolio by optimizing an objective or utility function subject to agronomic, economic, land, and water constraints. Different mathematical programming formulations imply different assumptions about farmers' objectives, calibration procedures, and behavioral responses, which is why alternative model structures are included in the ensemble. This satisfaction is represented through a linear or nonlinear utility or objective function U (**X**) that considers one or more single- or multi-attribute utility-relevant criteria, such as profit or risk avoidance:

$$Max\ U(\boldsymbol{X}) = f(z_1(\boldsymbol{X}), \ldots, z_m(\boldsymbol{X})) \tag{7}$$

Subject to:

$$x_i \geq 0 \tag{8}$$

$$\sum\nolimits_{i=1}^{n} x_i = 1 \tag{9}$$

$$\boldsymbol{X} \in F \tag{10}$$

$$\boldsymbol{X} \in \mathbb{R}^n \tag{11}$$

$$z_1(\boldsymbol{X}), \ldots, z_m(\boldsymbol{X}) = \boldsymbol{Z}(\boldsymbol{X}) \in \mathbb{R}^m \tag{12}$$

where the decision variable is the crop portfolio ($\boldsymbol{X}$), containing the share of land used by each crop x_i. Each attribute $z(x_i)$ depends on the land allocated to the different crops. The set of constraints includes the water availability constraint, described as follows:

$$\sum_{i=1}^{n} \left(\frac{w_i}{eff} x_i \right) \leq W_g \tag{13}$$

where w_i represents water consumption by crop i; eff is a parameter capturing technical irrigation efficiency in the irrigation system per agent; $\frac{w_i}{eff}$ is the water requirement per crop (i.e. including water consumption and return flows); $\sum_{i=1}^{n} \left(\frac{w_i}{eff} x_i \right)$ indicates the quantity of water required by the crop portfolio selected by the economic agent; while W_g represents the water allocated to each agent in the observed situation.

The ensemble is populated with four mathematical programming models with alternative setups for the equations above, namely: two nonlinear single-attribute PMP models (PMP1 and PMP2) that aim at profit maximization and (PMP1, which follows the average cost calibration method proposed by Howitt 1995, and PMP2 which follows the calibration method of Júdez et al. 2002); one linear multi-attribute WGP (Sumpsi et al., 1997); and one non-linear multi-attribute PMAUP model (Gutierrez-Martin and Gomez, 2011). Detailed descriptions of calibration procedures and results are provided in Annex I and Annex II of the supplementary material. These models span different objective functions, attributes, calibration approaches, and levels of complexity, enabling the exploration of heterogeneity in decision rules, behavior, and responses to changes in water availability.

2.2.2. Water module

Hydrological simulations were conducted using two complementary modeling platforms, each offering distinct advantages for representing catchment dynamics. HEC-HMS is a semi-distributed, open-access rainfall-runoff model widely employed in hydraulic studies (Hydrologic Engineering Center, 2018). While originally designed for engineering applications, HEC-HMS includes a continuous simulation module incorporating soil moisture accounting, enabling long-term analysis of flows at daily or hourly time steps. Reservoir operations, baseflow contributions, and irrigation abstractions can be represented, although the system is constrained in its capacity to capture fine-scale agricultural dynamics or crop-specific water use. To address this limitation, a tailored ecological component was integrated (Pérez-Blanco, González-López, et al., 2021). This component aggregates crop portfolios at sub-basin scale using area-weighted crop coefficients, implemented through the HEC-HMS canopy crop coefficient controlling evapotranspiration rather than through curve-number parameters. In HEC-HMS, the crop coefficient scales potential evapotranspiration in the canopy–surface–soil system, while irrigation is represented through monthly river-diversion accounting, soil-moisture adjustments, percolation to aquifers, and crop-water uptake removal from the system. Reservoir release schedules are incorporated using historical daily discharge patterns while ensuring compliance with environmental flow requirements and downstream water demands. The delineation of the watershed and sub-basin configuration was performed with HEC-GeoHMS, drawing on land cover datasets, crop maps, and administrative boundaries to provide a coherent spatial framework.

In parallel, the Soil and Water Assessment Tool – SWAT (Arnold et al., 1998) was employed to complement the analysis. SWAT is a semi-distributed eco-hydrological model that integrates land use, soil properties, topography, and management practices within Hydrological Response Units – HRUs, providing fine-scale resolution of agricultural and ecological processes (Mbungu and Kashaigili, 2017). The subdivision of sub-basins into HRUs enhances computational efficiency when scaling to larger catchments while maintaining the capacity to resolve heterogeneous land management and crop dynamics. Crop databases and irrigation routines allow explicit representation of water consumption and evapotranspiration, linking vegetation dynamics, agricultural productivity, and water availability. Reservoir operations are incorporated directly, enabling flexible management of storage and releases across diverse regulatory infrastructures. Two alternative reservoir configurations were implemented in SWAT. The first is a fixed-discharge approach (SWAT-FD), which follows historical daily release patterns but allows adjustments to ensure environmental-flow requirements and downstream irrigation demands. The second is a monthly-target setup (SWAT-MT), which uses monthly storage targets and allows daily releases to adjust dynamically in response to inflows, storage conditions, environmental-flow requirements, and irrigation-demand accounting. In parallel, HEC-HMS reservoir releases also follow historical daily discharge patterns, slightly adapted to ensure compliance with environmental-flow requirements and downstream AWDU demands. Therefore, SWAT-MT is used here as an alternative reservoir-operation structure, not as an endogenous economic feedback mechanism: monthly storage targets do not modify water charges or trigger within-year crop-allocation changes. These configurations illustrate the influence of operational choices on hydrological trajectories and the potential propagation of uncertainty within the system.

Together, HEC-HMS and SWAT offer complementary perspectives on catchment hydrology. HEC-HMS provides a computationally efficient framework for rapid setup and scenario testing, particularly where detailed crop dynamics are less critical, whereas SWAT allows integration of detailed land use, management, and ecological processes, capturing the complex interactions that drive agricultural water use and system dynamics. The combined use of these models supports robust evaluation of hydrological responses under different configurations and provides a flexible basis for testing coupling strategies with decision-making components.

Hydrological model calibration and validation were conducted before coupling the water models with the economic module. The SWAT hydrological component, shared by both SWAT-FD and SWAT-MT, was calibrated for 2005–2013 and validated for 2013–2014 at the upstream flow points draining into Santa Teresa Reservoir. Because SWAT-FD and SWAT-MT share the same watershed, soil, land-use, crop, and runoff-generation parameterization, the calibration statistics reported for SWAT apply to both configurations; their differences arise from the reservoir-operation rule used during scenario simulation, not from separate hydrological calibrations. HEC-HMS was calibrated for 1996–2013 and validated for 2013–2014 at the corresponding gauging locations. To ensure direct comparability across hydrological models, Table 1 reports NSE and R^2 for calibration and validation periods. Additional model-specific performance metrics, including KGE for SWAT and log NSE and RMSE for HEC-HMS, are provided in Annex III of the Supplementary Material.

The calibration and validation statistics indicate moderate but consistent performance, suitable for comparative scenario analysis and structural-uncertainty assessment in a regulated and human-managed catchment. The lower NSE values obtained for HEC-HMS (0.41–0.50 in calibration and 0.41–0.48 in validation) reflect the model's more aggregated representation of catchment processes, particularly at the daily time step in a strongly regulated basin where observed flows reflect reservoir-operation decisions in addition to natural runoff. For the purposes of this study, calibration adequacy is evaluated in terms of plausibility of the simulated hydrological response under controlled forcing, rather than benchmark predictive skill at gauged points: the analysis exploits between-model differences in regulated-system response under common scenarios and harmonized diagnostics, not absolute predictive accuracy. Within this scope, the calibration and validation statistics confirm that all three configurations (HEC-HMS, SWAT-FD, SWAT-MT) provide a consistent and physically plausible basis for structural-uncertainty assessment.

2.2.3. Human–water coupling process

The information exchange between the human and water modules is organized through a protocol-based coupling designed to transmit land-

Table 1
Hydrological model calibration and validation performance.

Model	Period	Calibration point	NSE	R^2
SWAT	Calibration 2005–2013	FLOW_OUT_10	0.68	0.69
SWAT	Calibration 2005–2013	FLOW_OUT_11	0.72	0.74
SWAT	Validation 2013–2014	FLOW_OUT_10	0.65	0.66
SWAT	Validation 2013–2014	FLOW_OUT_11	0.65	0.66
HEC-HMS	Calibration 1996–2013	W1060	0.42	0.44
HEC-HMS	Calibration 1996–2013	J130	0.50	0.59
HEC-HMS	Validation 2013–2014	W1060	0.41	0.42
HEC-HMS	Validation 2013–2014	J130	0.48	0.57

use, irrigation-demand, and water-availability information in a structured and reproducible manner (Essenfelder et al., 2018; Fernández et al., 2016). In the simulations presented here, the economic module operates at the irrigation-campaign scale. For each combination of drought condition and water charge scenario, farmers are represented as choosing a seasonal crop portfolio before the irrigation campaign, based on the relevant agronomic, economic, and water-availability conditions. Once this decision is made, the crop portfolio and associated irrigation-demand signal remain fixed during the simulated hydrological year. Therefore, water charge scenarios do not generate daily land-use changes. Instead, they generate seasonal crop-area and irrigation-demand changes, whose daily hydrological consequences are simulated by the water models through meteorological forcing, soil-water dynamics, evapotranspiration, runoff generation, routing, reservoir inflows, and reservoir releases.

The translation of the economic signal differs across hydrological models because of their different spatial and process structures. In HEC-HMS, crop portfolios are aggregated at the sub-basin level through area-weighted crop coefficients, which modify evapotranspiration through the canopy crop-coefficient parameter rather than through curve-number parameters. Irrigation demand is represented through a monthly water-accounting procedure: the absolute irrigation demand of each downstream irrigated sub-basin is withdrawn from the river at the sub-basin inlet, introduced into the sub-basin as an initial soil-moisture adjustment, and then reduced at the outlet by the estimated crop-water uptake. The remaining non-consumed fraction can return to the hydrological system through percolation, groundwater storage, baseflow, and routing. Reservoir operations are therefore adapted to release the total irrigation volume required by downstream AWDUs while maintaining compliance with environmental-flow requirements. In SWAT, where crops and irrigation operations are simulated explicitly at HRU scale, crop-portfolio changes are directly translated into land-use and management changes for each individual crop, while less significant crops are aggregated under general codes (see Annex III of the Supplementary Material). For downstream AWDUs, irrigation requirements are specified at HRU level together with the corresponding water source. When the river reach is defined as the source, SWAT simulates the river withdrawal, irrigation application, infiltration, crop uptake, and delayed return flows. Area reallocations are operationalized through targeted edits of HRU files, conserving sub-basin crop surface and maintaining pedological consistency, while ensuring that SWAT withdrawals align with AWDU demands.

In both hydrological models, irrigated AWDUs located upstream of Santa Teresa Reservoir were represented as being supplied at the Santa Teresa reservoir node, so that their irrigation demand was explicitly accounted for within the reservoir-centered water balance. This should be interpreted as a modeling and accounting representation adopted to ensure consistency between upstream abstractions, return flows, AWDU demands, reservoir releases, and the regulated-system water balance, rather than as a literal description of every local diversion infrastructure. In contrast, irrigated AWDUs located downstream of the reservoir were represented as withdrawing water from the river network after reservoir releases had incorporated their irrigation demand. Consequently, reservoir inflow diagnostics should be interpreted as effective net upstream hydrological responses under the adopted coupling structure, rather than as purely natural upstream runoff.

The framework also includes a second protocol through which water-availability constraints can be updated from the water module to the human module. In the simulations presented here, this protocol is defined but not triggered because water availability does not become binding for the economic module: under the tested water charge scenarios, simulated water demand remains below the available allocation. Therefore, crop portfolios are not recalculated within the hydrological year. Thus, although the framework is capable of two-way information exchange, the simulated scenario set results in a one-way realization of the coupling: the human-to-water protocol is active, whereas the water-to-human feedback protocol is not triggered because the water-availability constraint remains non-binding. Consequently, the reported results should be interpreted as diagnosing how seasonal crop-allocation and irrigation-demand signals are transmitted, amplified, or attenuated by alternative hydrological structures and reservoir-operation assumptions under contrasting drought forcing, rather than as simulations of dynamic within-year feedbacks between reservoir storage and farmer decisions. This scenario-contingent one-way realization is therefore consistent with the diagnostic objective of the study: to assess model structural uncertainty under common exogenous drought and water charge forcings, rather than to forecast realized system behavior under adaptive within-year management.

2.2.4. Scenario design

In this section, we elaborate on the drought and water pricing scenarios briefly referenced earlier in Section 2.2.3. The scenario design follows a controlled factorial logic intended to expose model structural uncertainty across the coupled ensemble rather than to forecast realized system behavior. Two drought conditions are crossed with four water charge levels and applied to twelve coupled model configurations, defined by three hydrological structures and four economic model formulations. This generates 96 coupled simulations. Drought and water charges are treated as exogenous forcings held constant across all model combinations within each scenario cell, so that divergence in simulated outputs can be attributed to model-structural differences rather than to changes in scenario specification. The choice of two contrasting drought conditions and a graduated set of water charge perturbations is intended to span a meaningful but tractable portion of the policy-relevant input space, while keeping the experiment within the computational and interpretive limits required for systematic structural-uncertainty diagnosis.

Economic scenarios are represented through alternative administrative water charge levels. In the Douro Basin, irrigation water payments are not market-clearing prices emerging from supply–demand interactions, but public charges set by the basin authority to recover part of the financial, resource, and environmental costs of water use (European Parliament and Council, 2000). These charges remain fixed until modified through administrative procedures and therefore do not respond endogenously to within-year reservoir storage, release conditions, or short-term scarcity. Consequently, water charges are treated here as exogenous policy instruments rather than endogenous scarcity prices, and within-year reservoir fluctuations do not feed back into the economic module through price adjustments.

We introduce three incremental volumetric water charge scenarios that increase the baseline charge by €0.01, €0.02, and €0.03/m^3 . Three water charge increments were selected to balance two competing requirements of the factorial design: spanning a meaningful range of policy-plausible perturbations and keeping the total number of coupled simulations tractable for systematic diagnostic analysis. A denser charge grid would multiply computational cost without proportionally enriching the structural-uncertainty diagnostic, because the analysis is concerned with the gradient of structural divergence across charge levels rather than with the precise reconstruction of the charge–response curve. These increments were selected to represent modest, policy-plausible adjustments to irrigation water charges in the Tormes Catchment, within the range of cost-recovery reforms compatible with Article 9 of the Water Framework Directive for the Douro Basin (DRBA, 2015). They are therefore policy-plausible perturbations rather than stress-test values, and their selection reflects the policy-appraisal framing of the study. These values are modest yet realistic (Sapino et al., 2022) and allow us to explore how small administrative charge adjustments cascade through crop-portfolio decisions, irrigation demand, and hydrological outcomes. For each charge level, the economic module generates a seasonal crop-allocation and irrigation-demand response, which is subsequently imposed on the hydrological models. The hydrological models then simulate the daily consequences of those seasonal decisions under the corresponding drought forcing.

Climate scenarios are represented through two complete hydrological years, defined from 1 October to 30 September, selected to span contrasting drought conditions in the Tormes–Santa Teresa system. The intense drought scenario is based on the 1980–1981 hydrological year, while the intermediate drought scenario is based on the 2001–2002 hydrological year. Daily meteorological forcing for each scenario was taken from the SWAT Global Weather Data database, which is based on the Climate Forecast System Reanalysis (CFSR) from the National Centers for Environmental Prediction (NCEP). Therefore, drought conditions affect the entire hydrological year rather than a specific crop-growth stage. Water charge scenarios, in contrast, affect the irrigation season only, from April to September, through seasonal crop-allocation and irrigation-demand changes.

The 2001–2002 hydrological year was selected as an intermediate drought benchmark because it represents a dry but non-extreme condition, allowing comparison with the more severe 1980–1981 drought benchmark. These scenarios are not intended to reproduce historical socioeconomic conditions exactly; instead, they use historical daily climate forcing to examine how the coupled models respond to contrasting drought intensities under present-day land use, water use, and socioeconomic conditions. The two scenarios are anchored to standardized drought intensity using the 6-months Standardized Precipitation–Evapotranspiration Index (SPEI−6), retrieved from the CSIC SPEI Global Drought Monitor (Beguería et al., 2014; Vicente-Serrano et al., 2010) for the grid cells covering the Tormes Catchment. SPEI−6 is preferred over precipitation-only indices because it accounts for the role of atmospheric evaporative demand in modulating water availability, which is particularly relevant for irrigation, particularly in semi-arid conditions such as those of the southwestern Douro Basin. The intense drought scenario (1980–1981) corresponds to a catchment-averaged SPEI−6 of −2.03 (severe to extreme drought) by December 1981, while the intermediate drought scenario (2001–2002) corresponds to a catchment-averaged SPEI−6 of −1.48 (moderate to severe drought). These values place the two scenarios at distinct points along the climatic-drought continuum, ensuring a meaningful contrast in hydrological-year forcing while keeping both within the historical record of the basin.

These scenarios do not reproduce the historical years exactly; instead, they distill their essential hydrological characteristics to examine how the coupled models respond to variations in drought intensity under present-day land and water use and socioeconomic conditions. In this study, protocol 2 is available within the coupling framework but is not triggered under the simulated scenarios because water availability remains non-binding for the economic module. Drought scenarios therefore stress the hydrological response of the coupled system under a realized human-to-water coupling pathway, rather than activating dynamic water-to-human feedbacks.Both drought scenarios are applied under the same reservoir-centered representation of upstream AWDU withdrawals adopted throughout the study. The interaction between the imposed climatic forcing and this representation is part of the scenario-design choice, and the diagnostic characterizes the response of the coupled system as represented, not of a notional unrepresented natural system.

The drought and water charge scenarios differ in their conceptual nature and in the temporal scale at which they enter the coupled system. Drought scenarios act through daily meteorological forcing fields that drive evapotranspiration, runoff generation, and routing within the hydrological models throughout the simulated year. Water charge scenarios act through seasonal economic decisions taken at the planning stage by the human module, which modify the crop portfolio and irrigation-demand signal used to configure the hydrological year. This asymmetry is intentional: it preserves the temporal logic of each module (e.g. daily for hydrology, seasonal for cropping decisions) and avoids imposing artificial sub-seasonal economic dynamics that would lie outside the standard formulation of the economic models used here.

2.2.5. *Structural uncertainty analysis*

Model structural uncertainty is assessed across the modular ensemble using standardized metrics that enable comparison across magnitudes, models, and scales. Here, model structural uncertainty refers to differences arising from alternative hydrological-model structures, economic-model formulations, and their coupling configurations. The analytical backbone is provided by *z*-scores (S_i), which transform deviations into a unitless measure, facilitating direct inter-model comparison. The standard *z*-score is computed as

$$S_i = \frac{Y_i - \bar{Y}}{S_Y} \tag{14}$$

where Y_i is the value of interest (e.g., irrigation water demand, reservoir storage, release volumes), while $\bar{Y}$ and S_Y are the relevant mean and standard deviation for the comparison set. For each drought condition and water charge scenario, z-scores are computed using the ensemble distribution of monthly reservoir-storage values across the coupled model configurations. This ensures that standardized anomalies compare model-structure responses under a common exogenous scenario rather than mixing scenario-driven and structure-driven variability. Widely used in environmental and hydro-climatic studies for anomaly detection (Botero-Acosta et al., 2024; Van Der Schalie et al., 2022), *z*-scores allow us to position model behavior relative to a shared statistical frame, thereby enabling harmonized cross-model diagnosis of structural divergence. Whisker plots and related visualizations are then employed to expose systematic contrasts and structural variability across the ensemble, making uncertainty patterns transparent and interpretable (Urgilez-Clavijo et al., 2024).

The outputs of the coupled hydro-economic system are first presented as an uncertainty cascade of daily reservoir-storage trajectories, which diagnose how dispersion emerges across hydrological and socio-economic model combinations for each drought and water charge scenario. To explicitly characterize the drivers of storage dynamics, we additionally evaluate reservoir inflows and outflows simultaneously. Daily inflow and outflow series are compared across hydrological models, economic models, and water charge scenarios, and scenario effects are quantified relative to the baseline as:

$$\Delta Q_{in,s}(t) = Q_{in,s}(t) - Q_{in,baseline}(t) \tag{15}$$

$$\Delta Q_{out,s}(t) = Q_{out,s}(t) - Q_{out,baseline}(t) \tag{16}$$

where *s* denotes each water charge scenario. Daily anomalies are converted into cumulative annual volume differences as:

$$\Delta V_{in,s} = \sum_t \Delta Q_{in,s}(t) \cdot 86400 \Big/ 10^6 \tag{17}$$

$$\Delta V_{out,s} = \sum_t \Delta Q_{out,s}(t) \cdot 86400 \Big/ 10^6 \tag{18}$$

with volumes expressed in hm^3. The net contribution of inflow and outflow changes to reservoir storage forcing is summarized as:

$$\Delta V_{net,s} = \Delta V_{in,s} - \Delta V_{out,s} \tag{19}$$

This diagnostic separates upstream hydrological responses from reservoir release responses and allows us to quantify whether storage differences are primarily associated with inflow changes, outflow changes, or their combined effect. It is not intended to close the complete reservoir mass balance, which may also include evaporation, seepage, spill, and changes in storage depending on the hydrological model. Rather, it isolates the contribution of the two fluxes explicitly requested by the reviewer—reservoir inflow and outflow—to storage forcing relative to the baseline scenario.

Next, structural uncertainty is analysed using monthly reservoir-storage z-scores represented by whisker plots for Santa Teresa Reservoir. This enables a consistent evaluation of heterogeneous responses across model families—defined as each hydrological model coupled with the full set of microeconomic models (e.g., HEC-HMS/PMP1, HEC-HMS/PMP2, HEC-HMS/PMAUP, and HEC-HMS/WGP, and analogously for SWAT-FD and SWAT-MT)—and within model families, i.e., across alternative microeconomic models for a fixed hydrological model. In doing so, we assess uncertainty associated with the coupled hydro-economic system and disentangle it from the component attributable solely to hydrological model structure.

3. Results

3.1. Reservoir storage, inflow, and outflow responses

In this section, we present the results in two steps. First, daily reservoir-storage trajectories are used to identify when ensemble dispersion emerges and how it evolves over the simulated hydrological year. Reservoir inflows and outflows are then assessed simultaneously to quantify the contribution of upstream hydrological responses and reservoir releases to storage dynamics. Second, standardized z-score diagnostics are used to distinguish between hydrological-model family effects and within-family economic-model effects, and the main numerical diagnostics are synthesized to identify the dominant model-structure uncertainty pathways. Because water availability does not become binding for the economic module in the simulations, the feedback protocol is not triggered. The reported results therefore diagnose the propagation of seasonal crop-allocation and irrigation-demand signals from the human module to the water module, rather than dynamic feedbacks from simulated reservoir conditions back to farmer decisions. Fig. 3 displays daily storage trajectories for the intense and intermediate drought scenarios under the baseline and three water charge scenarios.

The storage trajectories show that model structural uncertainty is temporally concentrated around the irrigation season. Under intense drought, dispersion increases sharply around M7–M9: SWAT-FD produces the widest storage envelope, HEC-HMS follows a narrower and smoother trajectory, and SWAT-MT shows a more constrained response associated with its monthly target-storage rule. Around M9 in the intense-drought baseline panel, mean monthly storage spans approximately 66–216 million m^3 across coupled configurations, corresponding to a range of ~150 million m^3, ~31% of Santa Teresa's maximum permitted storage, and ~95% of the aggregated annual downstream irrigation demand. Under intermediate drought, the timing of divergence is similar but the magnitude is larger: around M9, mean monthly storage ranges from ~226 million m^3 for SWAT-FD + PMP1 to ~428 million m^3 for SWAT-MT + PMAUP, corresponding to a ~202 million

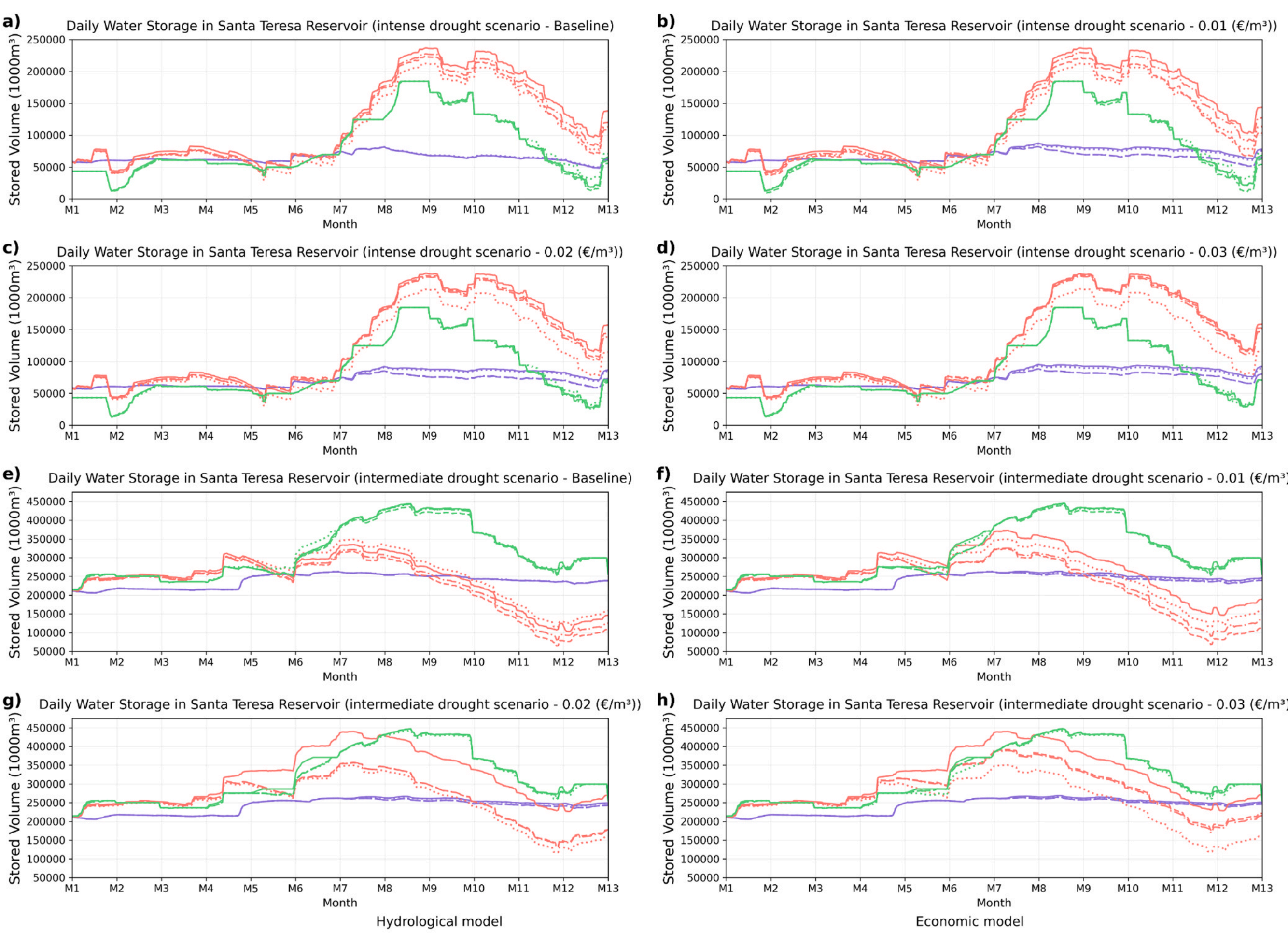

Fig. 3. Daily storage in the Santa Teresa reservoir under intense (panels a–d) and intermediate (panels e–h) drought scenarios. For each scenario, results are shown by water charge scenario (baseline, €0.01/m^3, €0.02/m^3, and €0.03/m^3). Each trace represents one coupled hydro-economic model configuration, combining one hydrological model with one microeconomic model. Line color identifies the hydrological model: HEC-HMS (Hydrologic Engineering Center's Hydrologic Modeling System) in blue, SWAT-FD (Soil & Water Assessment Tool - Fixed Discharge) in red, and SWAT-MT (Soil & Water Assessment Tool - Monthly Target) in green. Line style identifies the microeconomic model: PMAUP, PMP1, PMP2, or WGP. Both drought scenarios correspond to complete hydrological-year forcings, from month 0 (October) to month 12 (September).

m^3 spread, ~41% of reservoir capacity, ~127% of annual downstream irrigation demand, and ~120% of the annual surface-water supplied volume. Across water charge scenarios, the timing of divergence remains broadly consistent, whereas the magnitude of within-family spread differs among hydrological models. This is most visible for SWAT-FD late in the season, where economic formulations generate storage differences on the order of ~48 million m^3 . These storage patterns already indicate that structural uncertainty operates through two related pathways. First, hydrological structure determines the overall storage envelope by shaping how runoff generation, irrigation demand, routing, and reservoir releases are represented. Second, economic-model differences become visible only where the hydrological structure allows crop-portfolio and irrigation-demand changes to propagate into the water balance. This explains why SWAT-FD shows the largest late-season within-family spread, whereas HEC-HMS and SWAT-MT remain comparatively more constrained.

To disentangle the drivers of reservoir-storage dynamics, we compared reservoir inflows and outflows across hydrological models, economic models, drought periods, and water charge scenarios. Fig. 4 shows a representative case for the PMP1 economic model, while the equivalent inflow–outflow hydrographs for the remaining economic models (PMAUP, PMP2, and WGP) are provided in the Supplementary Material. The raw hydrographs show that reservoir inflow variability is largely dominated by the seasonal hydrological signal, whereas outflow variability is more clearly shaped by reservoir-operation rules and irrigation-demand propagation through the coupled system. This distinction is particularly evident when comparing HEC-HMS, SWAT-FD, and SWAT-MT. HEC-HMS produces smoother inflow and outflow responses, whereas SWAT-FD and SWAT-MT produce more event-driven inflow patterns. These differences reflect how each hydrological structure transmits the same seasonal economic signal. In SWAT, crop-area reallocations and irrigation operations are represented explicitly at HRU scale. For upstream irrigated AWDUs, Santa Teresa Reservoir accounts for their irrigation demand; therefore, when higher water charges reduce irrigated area and irrigation demand, less irrigation water is applied to upstream HRUs. This can reduce abstractions, but it can also reduce irrigation return flows, percolation, groundwater recharge, and delayed contributions to the river network. As a result, reductions in irrigation demand do not necessarily translate into higher reservoir inflow. In HEC-HMS, by contrast, crop portfolios are aggregated at sub-basin scale through area-weighted crop coefficients, and irrigation demand is represented through monthly soil-moisture and diversion adjustments. This attenuation should not be interpreted as a lack of response to the economic signal. Rather, it is the structural consequence of representing crop portfolios through area-weighted sub-basin crop coefficients and irrigation demand through monthly water-accounting adjustments, which smooths the spatial and temporal expression of charge-induced crop and irrigation changes.

Across water charge scenarios, changes in reservoir inflow are generally modest relative to the magnitude of the full hydrograph. In contrast, outflow differences are more visible, particularly in SWAT-MT, where reservoir releases respond to the monthly target-storage logic. The comparison therefore indicates that water-charge-induced economic signals are detectable in reservoir inflows, but their effect is small compared with drought-driven seasonal variability. The outflow response is generally stronger because it directly reflects the interaction between irrigation-demand signals and reservoir-operation assumptions. The contrast between SWAT-FD and SWAT-MT should therefore be interpreted as uncertainty associated with reservoir-operation structure, not as evidence of dynamic feedback between reservoir storage, water charges, and within-year farmer decisions. In the simulations reported here, the feedback protocol is not activated, so changes in reservoir storage or releases do not modify water charges or trigger additional crop reallocation within the simulated year. The comparison between SWAT-FD and SWAT-MT isolates the effect of alternative reservoir-operation assumptions, whereas differences among water

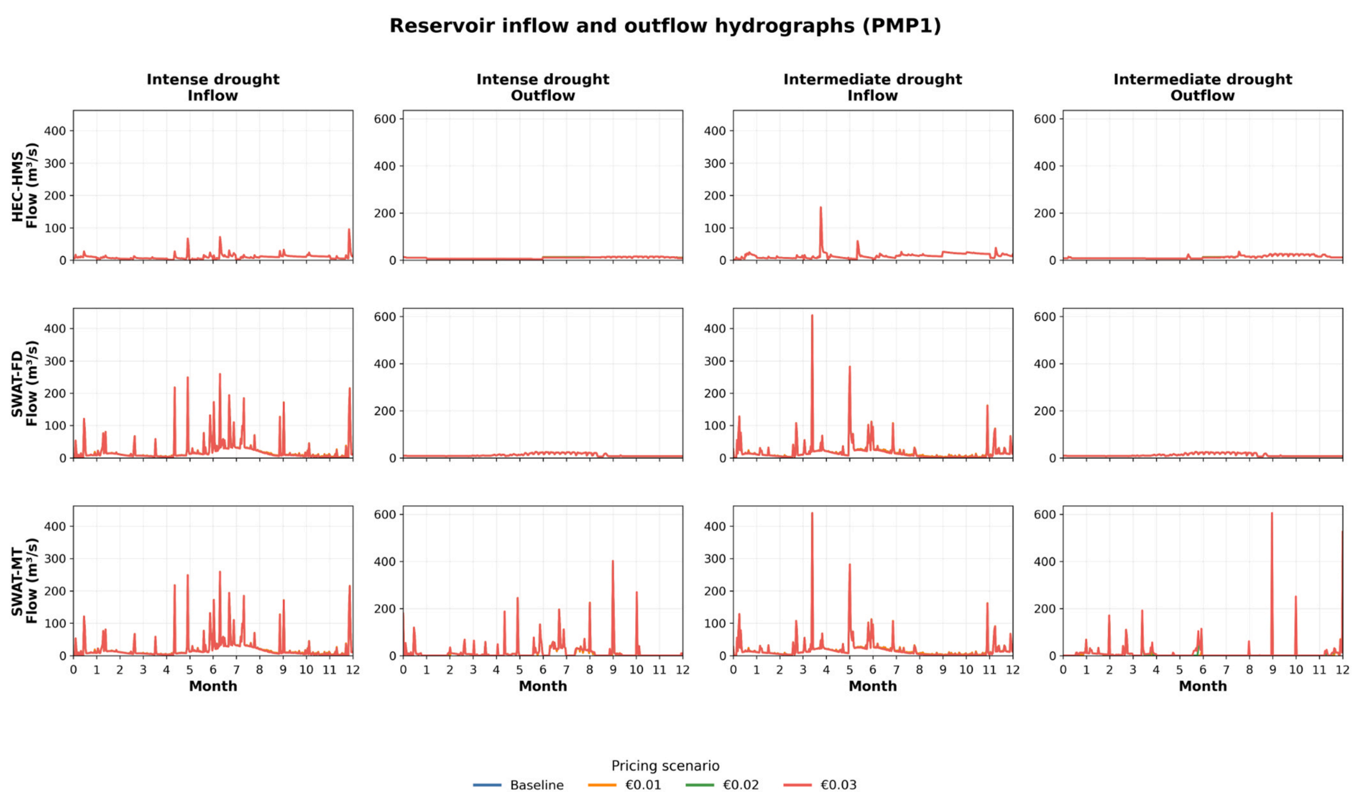

Fig. 4. Daily reservoir inflows and outflows for the PMP1 economic model under intense and intermediate drought conditions. Rows correspond to hydrological models and columns to inflow/outflow responses. The x-axis represents the simulated annual period from month 0 to month 12. Colors indicate the baseline and water charge scenarios. Equivalent hydrographs for the remaining economic models are provided in the Supplementary Material.

charge scenarios isolate the propagation of exogenous seasonal economic signals through the one-way coupling.

To quantify the combined influence of inflows and outflows on reservoir-storage dynamics, we computed annual cumulative changes relative to the baseline scenario. For each water charge scenario, cumulative inflow and outflow anomalies were calculated as $\Delta V_{in}$and ΔV_{out}, respectively, and their combined contribution was summarized using the net inflow–outflow balance defined in Eq. (19). Positive values indicate a net inflow–outflow balance that tends to favor higher storage relative to baseline, whereas negative values indicate a net tendency toward lower storage. Fig. 5 summarizes this diagnostic for all hydrological models, economic models, water charge scenarios, and drought conditions. Across the full ensemble, annual ΔV_{net} ranges from -63.9 to $+30.7$ million m^3 yr^{-1} under intense drought and from -162.7 to $+13.0$ million m^3 yr^{-1} under intermediate drought, showing that the net inflow–outflow balance varies substantially across hydrological structures and reservoir-operation configurations. The complete numerical values underlying this diagnostic, including ΔV_{in}, ΔV_{out}, and ΔV_{net} for all hydrological–economic model combinations, are provided in Table A IV.1 of the Supplementary Material. The heatmap shows that the net inflow–outflow balance differs substantially across hydrological structures. HEC-HMS generally produces positive net balances under water charge scenarios, whereas SWAT-FD and especially SWAT-MT tend to produce negative balances in several model combinations, particularly under intermediate drought. This contrast reflects the different ways in which hydrological structures transmit or attenuate seasonal water-charge-induced demand changes. In HEC-HMS, the aggregation of crop portfolios at sub-basin scale and the monthly representation of irrigation demand smooth the inflow response. In SWAT, explicit HRU-scale irrigation operations allow changes in irrigated area and irrigation demand to affect abstractions, return flows, percolation, groundwater recharge, and delayed streamflow contributions. Because upstream AWDUs are represented within the Santa Teresa reservoir-centered water balance, annual inflow anomalies should be interpreted as net upstream hydrological responses of the coupled representation. They integrate crop-area reallocations, irrigation-demand changes, abstractions, irrigation return flows, groundwater contributions, and routing effects. They should therefore not be interpreted as isolated responses of individual upstream AWDUs, nor as clean changes in natural runoff. Reservoir-storage divergence is consequently not explained by inflow or outflow alone, but by their combined balance under each hydrological structure and reservoir-operation configuration.

Overall, the inflow–outflow diagnostics show that reservoir-storage differences are associated with both upstream inflow responses and reservoir-release adjustments. Having quantified this balance, we next evaluate how these differences translate into standardized model-structure divergence across the coupled ensemble.

3.2. Standardized model-structure uncertainty diagnostics

We then apply the z-score assessment to the fully coupled ensemble outcomes (Fig. 6). The figure displays distributions of monthly reservoir-storage anomalies (z-scores) for each drought scenario (a): Intense; and b): Intermediate), shown separately again by water charge level (Baseline, €0.01/m^3, €0.02/m^3, €0.03/m^3). Boxes are organized first across hydrological model families (HEC-HMS, SWAT-FD, SWAT-MT) and then within each family across the four coupled economic baselines, corresponding to the calibrated no-policy configuration of each economic model. Between-family spread therefore reflects hydrological structural differences, whereas within-family spread reflects how alternative economic formulations propagate from their calibrated baseline under a fixed hydrological structure. This layout indicates how much dispersion is attributable to between-family hydrological differences and how

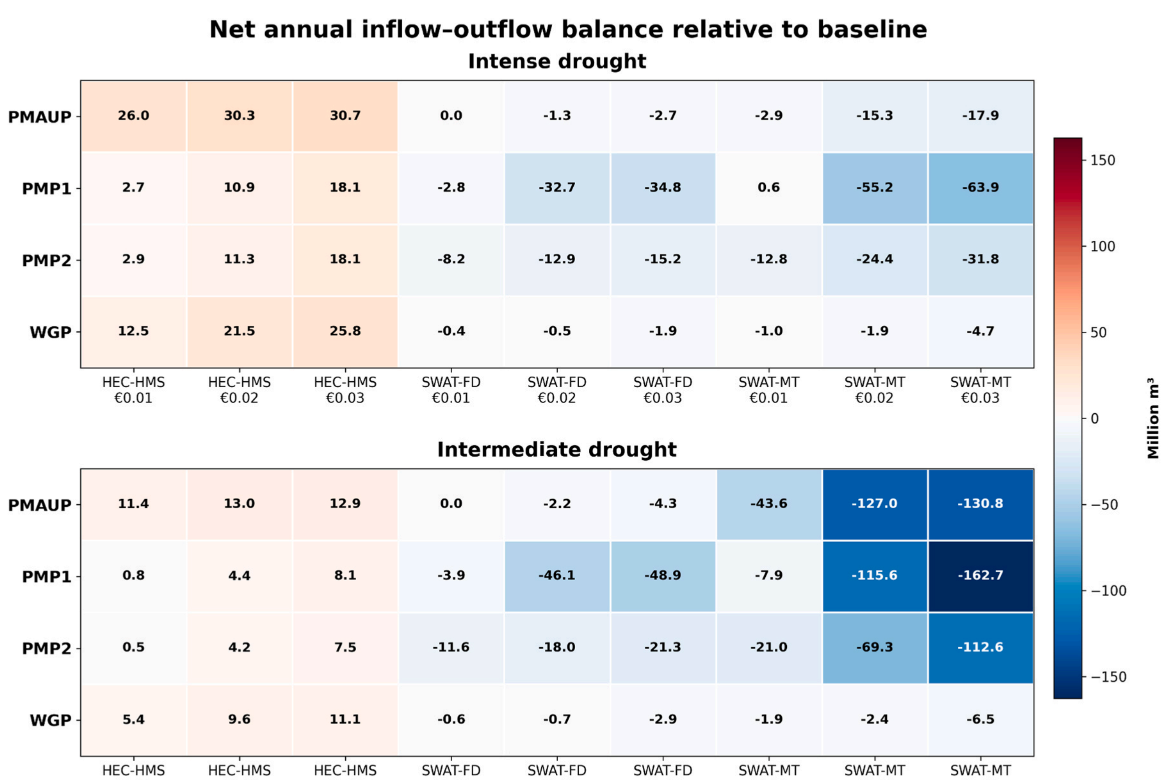

Fig. 5. Annual net inflow–outflow balance, computed as …, for each drought condition, hydrological model, economic model, and water charge scenario. Values are expressed in million m^3 . Positive values indicate a net tendency toward higher storage relative to baseline, while negative values indicate a net tendency toward lower storage. Inflow values should be interpreted as effective reservoir-node inflows under the reservoir-centered coupling representation, not as natural runoff alone.

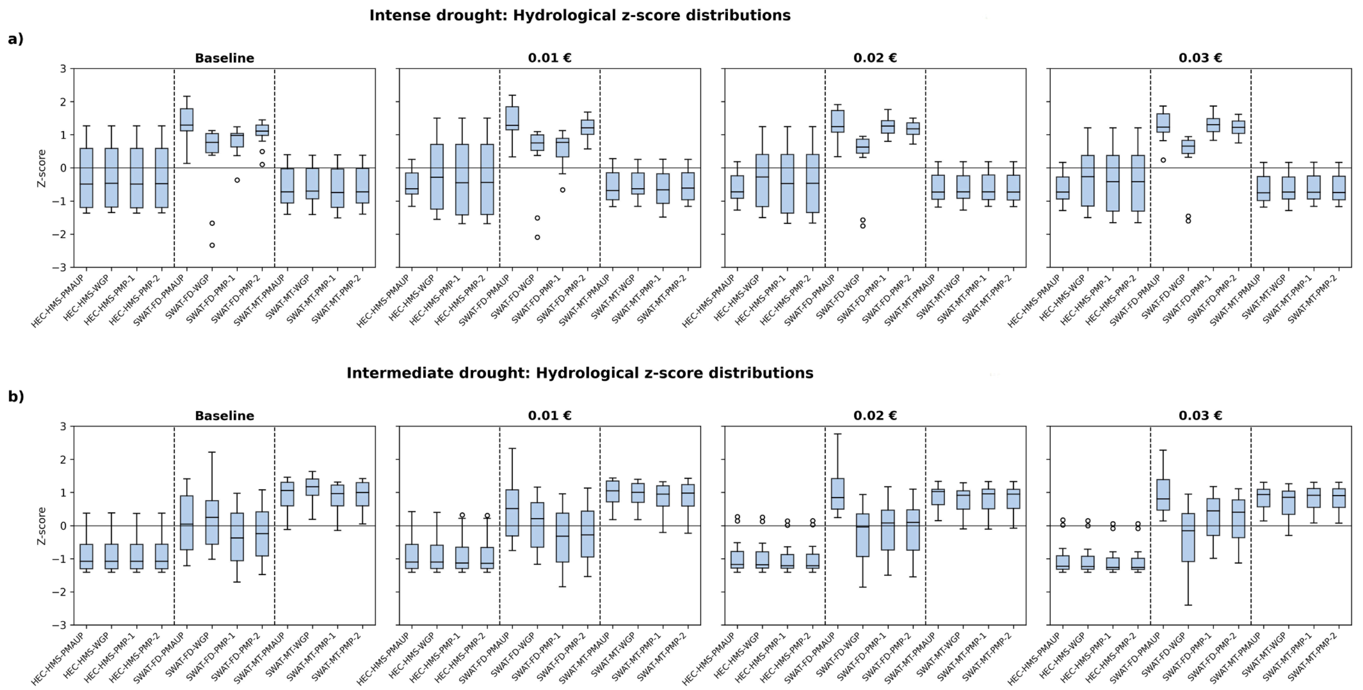

Fig. 6. Distribution of standardized anomalies (z-scores) across the multi-model ensemble for the two drought scenarios (a) Intense and (b) Intermediate. Boxplots represent the monthly reservoir-storage anomalies (z-scores) generated by the coupled hydro-economic model combinations, comprising microeconomic models PMP1, PMP2 (Positive Mathematical Programming), PMAUP (Positive Multi-Attribute Utility Programming), and WGP (Weighted Goal Programming), and hydrological models HEC-HMS (Hydrologic Engineering Center's Hydrologic Modeling System), SWAT-FD (Soil & Water Assessment Tool -Fixed Discharge), and SWAT-MT (Soil & Water Assessment Tool -Monthly Target). Results are grouped by water charge scenario (baseline, €0.01, €0.02, and €0.03).

much remains as within-family variability associated with alternative economic formulations.

The whisker plots show that monthly reservoir-storage anomalies are primarily driven by hydrological model family, while economic-model effects appear mainly as within-family variability. Because z-scores are computed within each water charge level, each panel isolates structural divergence under the same exogenous water charge input. The max–min range of coupling median z-scores reaches 1.99 under intense drought and 2.21 under intermediate drought, indicating that model-structure choice shifts the central tendency of monthly anomalies by approximately two standardized units. At the hydrological-family level, median separation reaches 1.88 and 2.12 z-score units under intense and intermediate drought, respectively, showing that hydrological structure explains most of the between-coupling spread. Economic-model effects are still visible, but mainly within hydrological families: SWAT-FD shows the largest within-family spread, reaching 0.57 and 0.77 z-score units under intense and intermediate drought, whereas HEC-HMS and SWAT-MT remain comparatively tighter. Aggregating across configurations, SWAT-FD tends to produce higher storage anomalies, HEC-HMS tends to produce lower storage anomalies, and SWAT-MT provides intermediate, monthly target-damped responses. The main numerical diagnostics are synthesized in Table 2.

These standardized diagnostics confirm the mechanism identified in the storage and inflow–outflow results. The large between-family separation indicates that hydrological structure controls the main position of each model family within the ensemble. The larger within-family spread in SWAT-FD shows that this configuration propagates economic-model differences most strongly, because crop-area and irrigation-demand changes are expressed through explicit HRU-scale crop and irrigation processes. By contrast, the tighter HEC-HMS distributions reflect the smoothing effect of sub-basin aggregation and monthly irrigation accounting, while the tighter SWAT-MT distributions reflect the damping effect of the monthly target-storage operation on storage trajectories. Taken together, the storage trajectories, inflow–outflow diagnostics, and standardized z-score distributions show that model structural uncertainty is not uniformly distributed across the coupled ensemble. Table 2 synthesizes the main numerical diagnostics, identifying when divergence is largest, which model components dominate the spread, and how these differences translate into management-relevant reservoir-storage uncertainty.

The synthesis shows that the largest divergence consistently occurs around irrigation onset, when irrigation demand, reservoir releases, routing, and crop-water processes interact. Hydrological model structure dominates the between-model spread, but economic-model effects become visible within hydrological families, especially for SWAT-FD. The magnitude of the storage spread is management-relevant, reaching values comparable to a large share, or more than the full amount, of annual downstream irrigation demand. These results provide the empirical basis for the broader implications discussed below.

4. Discussion

Hydro-economic decision support often relies on point predictions generated by a single model, or by a single combination of hydrological and economic models. Our results show that this practice can provide a false sense of precision in regulated agricultural basins. Under the same drought and water charge assumptions, alternative model structures generate storage spreads of 150–202 million m^3 around irrigation onset, equivalent to 31–41% of Santa Teresa's maximum permitted storage and 95–127% of annual downstream irrigation demand. These differences are large enough to affect reservoir-operation, drought-management, and water-charge decisions. Therefore, the relevant policy message is not that one model is "right" and another is "wrong", but that single-model appraisals may exclude plausible futures that matter for management.

The first implication is that hydro-economic policy appraisal should incorporate multi-model ensembles whenever structural choices are likely to affect management-relevant outcomes. This is especially

Table 2
Synthesis of model structural uncertainty diagnostics.

Diagnostic	Intense drought	Intermediate drought	Main interpretation
Main divergence window	M7–M9	M7–M9	The largest divergence appears around irrigation onset, when irrigation demand, reservoir releases, routing, and crop-water processes interact.
Maximum storage spread across coupled models	~150 million m^3	~202 million m^3	The ensemble spans management-relevant storage states that a single-model setup could miss.
Storage spread relative to reservoir capacity	~31%	~41%	Structural uncertainty is large relative to Santa Teresa's maximum permitted storage.
Storage spread relative to annual downstream irrigation demand	~95%	~127%	The uncertainty range is comparable to, or larger than, annual downstream irrigation demand.
Max–min range of coupling median z-scores	1.99	2.21	Model choice shifts monthly storage anomalies by around two standardized units.
Couplings defining the largest z-score range	SWAT-FD + PMAUP vs SWAT-MT + PMP1	SWAT-MT + PMAUP vs HEC-HMS + PMP2	The extreme responses depend on both hydrological structure and economic coupling.
Hydrological-family median separation	1.88	2.12	Most between-model divergence is explained by hydrological model structure.
Within-family economic spread: HEC-HMS	0.28 z-score units	0.02 z-score units	Economic-model effects are dampened by sub-basin crop aggregation and monthly irrigation accounting.
Within-family economic spread: SWAT-FD	0.57 z-score units	0.77 z-score units	Explicit HRU-scale crop and irrigation processes allow economic-model differences to propagate most strongly.
Within-family economic spread: SWAT-MT	0.03 z-score units	0.08 z-score units	Monthly target operation dampens within-family economic variability.
Annual net inflow–outflow balance range, ΔV_{net}	−63.9 to + 30.7 million m^3 yr^{-1}	−162.7 to + 13.0 million m^3 yr^{-1}	Storage response is controlled by the combined inflow–outflow balance; inflow anomalies represent net upstream coupled responses, not natural runoff alone.
Dominant uncertainty pathway	Hydrological structure, with SWAT-FD showing strongest economic coupling	Hydrological structure, with stronger SWAT-MT operational effect	Structural uncertainty is conditional on both model structure and scenario context.
Storage spread relative to annual surface-water supplied volume	~89%	~120%	The storage spread is comparable to the annual volume supplied by surface water in the regulated system.
Late-season SWAT-FD storage spread among economic formulations	~48 million m^3	~48 million m^3	Economic-model effects become most visible within SWAT-FD toward the end of the irrigation season.

important in systems where reservoir operation, irrigation demand, crop-water processes, and return flows interact during critical seasonal transitions. In such contexts, model choice becomes part of the uncertainty faced by decision makers. Ensembles allow analysts to bound plausible storage, inflow, and outflow responses, rather than presenting a single trajectory as if it were fully supported by the available evidence.

The second implication is that modeling uncertainty should be quantified and attributed, not only acknowledged qualitatively. The standardized z-score diagnostics used here provide a simple way to compare heterogeneous model outputs and identify whether uncertainty is mainly driven by hydrological model structure, economic-model formulation, or their coupling. In our case, hydrological model structure explains most of the between-model spread, while economic-model effects are expressed mainly within hydrological families, especially for SWAT-FD. This type of diagnostic helps target future modeling effort: where hydrological structure dominates, adding behavioral detail alone will not sufficiently reduce uncertainty unless the water-system representation is also tested.

The third implication is that uncertainty should be communicated as a management-relevant range, not only as model performance statistics or individual scenario outcomes. Expressing storage spread relative to reservoir capacity, annual irrigation demand, and supplied surface-water volume makes the uncertainty directly interpretable for planners. This supports more transparent communication with basin authorities and stakeholders, particularly when evaluating water charge reforms or drought-management strategies.

These findings also support robust decision-making approaches. Rather than optimizing policies against a single predicted future, water managers can use multi-model ensembles to identify strategies that perform satisfactorily across a wide range of plausible model structures and drought conditions. In practice, this means evaluating whether proposed water charges, reservoir-operation rules, or irrigation allocations remain acceptable when tested against alternative hydrological structures, economic formulations, and coupling assumptions. The diagnostic framework used here—storage trajectories, inflow–outflow balance, standardized z-scores, and synthesis tables—can be transferred to other regulated agricultural basins where alternative model structures are available.

The transferability of these findings depends on the presence of comparable system features rather than on the specific characteristics of the Tormes Catchment alone. The uncertainty patterns identified here are most relevant for regulated agricultural basins where irrigation supply is reservoir-mediated, multiple plausible reservoir-operation rules exist, farmer behavior can be represented through alternative economic formulations, and the irrigation season creates a critical interaction between crop-water demand and release scheduling. These conditions are common in Mediterranean and semi-arid basins, including other regulated sub-basins of the Douro and comparable systems such as the Ebro, Guadiana, or Po. In such contexts, the proposed diagnostic framework can support basin-scale climate services and policy appraisal by identifying when and where model structural uncertainty becomes decision-relevant.

Several limitations define priorities for future work. First, the present ensemble focuses on model structural uncertainty but does not provide a full global sensitivity analysis of input data and parameter uncertainty. Such analysis would be valuable for separating structural effects from uncertainty in climatic inputs, hydrological parameters, crop-water parameters, economic coefficients, and behavioral calibration, but it is computationally demanding when process-based hydrological models are coupled with economic models. Second, the drought scenarios used here are based on historical hydrological years rather than climate projections. This choice is consistent with the diagnostic objective of isolating model-structural behavior under observed drought conditions of contrasting severity; extending the framework to projected droughts under climate change would be a natural next step, but would require separating structural uncertainty from climate-projection uncertainty. Third, future applications could expand the ensemble to include additional system components, such as agronomic models, macroeconomic feedbacks, endogenous climate-risk modules, or alternative reservoir-operation formulations. Fourth, metamodels, emulators, reduced-order models, AI-based surrogates, or targeted experimental designs could help address computational cost and enable broader exploration of structural, parametric, and scenario uncertainty. Finally, future work should test cases where water availability becomes binding for the economic module, allowing the feedback protocol to be assessed explicitly under conditions where it is expected to affect crop-allocation decisions.

5. Conclusions

Model structural uncertainty shapes coupled human–water outcomes and should be treated as a first-order component of hydro-economic policy appraisal. Across both drought scenarios, model-structure choice shifts monthly reservoir-storage anomalies by about two standardized units, with max–min ranges of coupling median z-scores of 1.99 under intense drought and 2.21 under intermediate drought; most of this separation is explained by hydrological structure, with family median separations of 1.88 and 2.12 z-score units, respectively. In volumetric terms, the largest divergence occurs around irrigation onset and reaches ~150–202 million m^3 across coupled configurations, equivalent to 31–41% of Santa Teresa's maximum permitted storage and 95–127% of annual downstream irrigation demand. Economic formulation contributes mainly through within-family variability and is most evident when coupled with SWAT-FD, whereas HEC-HMS and SWAT-MT remain comparatively tighter. The additional inflow–outflow diagnostics show that storage divergence emerges from the joint contribution of upstream hydrological responses and reservoir-release adjustments: water-charge-induced crop-area and irrigation-demand changes produce modest but detectable inflow anomalies, while outflow responses are more pronounced because they directly reflect reservoir-operation rules and irrigation releases. These findings indicate that single-model assessments can materially underestimate the range of plausible reservoir-storage trajectories and may bias drought-management or water charge appraisals. Multi-model hydro-economic ensembles therefore provide a practical way to bound structural uncertainty, identify dominant uncertainty pathways, and support more robust decision-making in regulated agricultural basins. Future work should test cases where water availability becomes binding for the economic module and expand the structural ensemble using efficient strategies such as emulators, reduced-order models, or targeted experimental designs.

CRediT authorship contribution statement

C. Dionisio Pérez-Blanco: Writing – review & editing, Visualization, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **David Rivas-Tabares:** Writing – review & editing, Visualization. **José María Bodoque:** Writing – review & editing. **Arthur Hrast Essenfelder:** Writing – review & editing, Visualization, Validation, Methodology. **Francesco Sapino:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Data curation. **Héctor González-López:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research leading to these results has been developed with the support of the Horizon Europe's TRANSCEND (Transformational and Robust Adaptation to Water Scarcity and Climate Change Under Deep Uncertainty) Project, PRIMA Foundation's TALANOA-WATER Project (Talanoa Water Dialogue for Transformational Adaptation to Water Scarcity under Climate Change), national NATURA (Nature-based Solutions for Resilient Adaptation to Water Scarcity and Hydroclimatic Extremes) project, and the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101136799 (MARCLAIMED).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.agwat.2026.110582.

Data availability

Data will be made available on request.

References

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment Part I. Model. Dev. J. Am. Water Resour. Assoc. 34, 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x.

Beguería, S., Vicente-Serrano, S.M., Reig, F., Latorre, B., 2014. Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. Int. J. Climatol. 34, 3001–3023. https://doi.org/10.1002/joc.3887.

Botero-Acosta, A., Cox, A., Sagan, V., Demir, I., Muste, M., Boyd, P., Pathak, C., 2024. Multivariate Analysis and Anomaly Detection of U.S. Reservoir Sedimentation Dataset. https://doi.org/10.31223/X54H3Z.

Dolan, F., Lamontagne, J., Link, R., Hejazi, M., Reed, P., Edmonds, J., 2021. Evaluating the economic impact of water scarcity in a changing world. Nat. Commun. 12, 1915. https://doi.org/10.1038/s41467-021-22194-0.

DRBA, 2015. Plan Especial de Sequía. Plan Hidrológico de la parte española de la demarcación hidrográfica del Duero. 2015-2021 (Report). Douro River Basin Auth. ⟨https://www.chduero.es/documents/20126/89007/PHD15-1310_00_PESeq-v03_01.pdf⟩.

Dwivedi, D., Poeppl, R.E., Wohl, E., 2025. Hydrological connectivity: a review and emerging strategies for integrating measurement, modeling, and management. Front. Water 7, 1496199. https://doi.org/10.3389/frwa.2025.1496199.

Essenfelder, A.H., Pérez-Blanco, C.D., Mayer, A.S., 2018. Rationalizing systems analysis for the evaluation of adaptation strategies in complex human-water systems. Earths Future 6, 1181–1206. https://doi.org/10.1029/2018EF000826.

European Parliament and Council, 2000. Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy.

Fernández, F.J., Ponce, R.D., Blanco, M., Rivera, D., Vásquez, F., 2016. Water variability and the economic impacts on small-scale farmers. A farm risk-based integrated modelling approach. Water Resour. Manag. 30, 1357–1373. https://doi.org/10.1007/s11269-016-1227-8.

Gil-García, L., González-López, H., Pérez-Blanco, C.D., 2023. To dam or not to dam? Actionable socio-hydrology modeling to inform robust adaptation to water scarcity and water extremes. Environ. Sci. Policy 144, 74–87. https://doi.org/10.1016/j.envsci.2023.03.012.

Gupta, A., Govindaraju, R.S., 2019. Propagation of structural uncertainty in watershed hydrologic models. J. Hydrol. 575, 66–81. https://doi.org/10.1016/j.jhydrol.2019.05.026.

Gutierrez-Martin, C., Gomez, C.M., 2011. Assessing irrigation efficiency improvements by using a preference revelation model. Span. J. Agric. Res. 9, 1009. https://doi.org/10.5424/sjar/20110904-514-10.

Hasan, H., Hashmi, M.Z.U.R., Ahmed, S.I., Anees, M., 2025. Assessing climate sensitivity of the Upper Indus Basin using fully distributed, physically-based hydrologic modeling and multi-model climate ensemble approach. Sci. Rep. 15, 4109. https://doi.org/10.1038/s41598-024-84975-z.

Hino, M., Hall, J.W., 2017. Real options analysis of adaptation to changing flood risk: structural and nonstructural measures. ASCE-ASME J. Risk Uncertain. Eng. Syst. Part. Civ. Eng. 3, 04017005. https://doi.org/10.1061/AJRUA6.0000905.

Howitt, R.E., 1995. Positive mathematical programming. Am. J. Agric. Econ. 77, 329–342. https://doi.org/10.2307/1243543.

Hydrologic Engineering Center, 2018. HEC-HMS User's Manual version 4.3. ⟨https://www.hec.usace.army.mil/software/hec-hms/documentation/HEC-HMS_Users_Manual_4.3.pdf⟩ [software]. 640.

Jones, E.R., Bierkens, M.F.P., Van Vliet, M.T.H., 2024. Current and future global water scarcity intensifies when accounting for surface water quality. Nat. Clim. Change 14, 629–635. https://doi.org/10.1038/s41558-024-02007-0.

Júdez, L., de Miguel, J.M., Mas, J., Bru, R., 2002. Modeling crop regional production using positive mathematical programming. Math. Comput. Model. 35, 77–86. https://doi.org/10.1016/S0895-7177(01)00150-9.

Knowling, M.J., White, J.T., McDonald, G.W., Kim, J.-H., Moore, C.R., Hemmings, B., 2020. Disentangling environmental and economic contributions to hydro-economic model output uncertainty: an example in the context of land-use change impact assessment. Environ. Model. Softw. 127, 104653. https://doi.org/10.1016/j.envsoft.2020.104653.

Mbungu, W.B., Kashaigili, J.J., 2017. Assessing the hydrology of a data-scarce tropical watershed using the soil and water assessment tool: case of the little Ruaha river watershed in Iringa, Tanzania. Open. J. Mod. Hydrol. 07, 65–89. https://doi.org/10.4236/ojmh.2017.72004.

Moges, E., Demissie, Y., Larsen, L., Yassin, F., 2020. Review: sources of hydrological model uncertainties and advances in their analysis. Water 13, 28. https://doi.org/10.3390/w13010028.

Noh, S.J., Lee, G., Kim, B., Lee, S., Jo, J., Woo, D.K., 2024. Climate change impact assessment on water resources management using a combined multi-model approach in South Korea. J. Hydrol. Reg. Stud. 53, 101842. https://doi.org/10.1016/j.ejrh.2024.101842.

Ortiz-Partida, J.P., Fernandez-Bou, A.S., Maskey, M., Rodríguez-Flores, J.M., Medellín-Azuara, J., Sandoval-Solis, S., Ermolieva, T., Kanavas, Z., Sahu, R.K., Wada, Y., Kahil, T., 2023. Hydro-economic modeling of water resources management challenges: current applications and future directions. Water Econ. Policy 09, 2340003. https://doi.org/10.1142/S2382624X23400039.

Pérez-Blanco, C.D., González-López, H., Hrast-Essenfelder, A., 2021. Beyond piecewise methods: Modular integrated hydroeconomic modeling to assess the impacts of adaptation policies in irrigated agriculture. Environ. Model. Softw. 136, 104943. https://doi.org/10.1016/j.envsoft.2020.104943.

Rachunok, B., Fletcher, S., 2023. Socio-hydrological drought impacts on urban water affordability. Nat. Water 1, 83–94. https://doi.org/10.1038/s44221-022-00009-w.

Sapino, F., Pérez-Blanco, C.D., Saiz-Santiago, P., 2022. A hydro-economic model to calculate the resource costs of agricultural water use and the economic and environmental impacts of their recovery. Water Econ. Policy 08, 2240012. https://doi.org/10.1142/S2382624X22400124.

Sumpsi, J., Amador, F., Romero, C., 1997. On farmers' objectives: a multi-criteria approach. Eur. J. Oper. Res. 96, 64–71. https://doi.org/10.1016/0377-2217(95)00338-X.

Urgilez-Clavijo, A., Rivas-Tabares, D., Gobin, A., De La Riva, J., 2024. Comprehensive framework for analysing the intensity of land use and land cover change in continental ecuadorian biosphere reserves. Sustainability 16, 1566. https://doi.org/10.3390/su16041566.

Van Der Schalie, R., Van Der Vliet, M., Albergel, C., Dorigo, W., Wolski, P., De Jeu, R., 2022. Characterizing natural variability in complex hydrological systems using passive microwave-based climate data records: a case study for the Okavango Delta. Hydrol. Earth Syst. Sci. 26, 3611–3627. https://doi.org/10.5194/hess-26-3611-2022.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23, 1696–1718. https://doi.org/10.1175/2009JCLI2909.1.

Zhou, X., Sheng, Z., Manevski, K., Zhao, R., Zhang, Q., Yang, Yanmin, Han, S., Liu, J., Yang, Yonghui, 2025. Can system dynamics explain long-term hydrological behaviors? The role of endogenous linking structure. Hydrol. Earth Syst. Sci. 29, 159–177. https://doi.org/10.5194/hess-29-159-2025.