

Earth's Future

RESEARCH ARTICLE

10.1029/2024EF005437

Exploring Global Water Scarcity Dynamics Through Causal Discovery and Structural Causal Modelling



Key Points:

- Causal modelling confirms that population growth significantly impacts water scarcity, surpassing climate factors
- Causal discovery of water-scarce regions faces challenges regarding data limitations, currently substituted by expert knowledge
- A future focus on localized data sets can enhance causal modelling to improve water management strategies in water-scarce regions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

M. Leijnse,
m.leijnse@uu.nl

Citation:

Leijnse, M., Bierkens, M. F. P., & Wanders, N. (2025). Exploring global water scarcity dynamics through causal discovery and structural causal modelling. *Earth's Future*, 13, e2024EF005437. <https://doi.org/10.1029/2024EF005437>

Received 2 OCT 2024

Accepted 25 JUL 2025

Author Contributions:

Conceptualization: Myrthe Leijnse, Marc F. P. Bierkens, Niko Wanders

Formal analysis: Myrthe Leijnse

Funding acquisition: Marc F. P. Bierkens, Niko Wanders

Writing – original draft: Myrthe Leijnse

Writing – review & editing: Marc F. P. Bierkens, Niko Wanders

Myrthe Leijnse¹ , Marc F. P. Bierkens^{1,2} , and Niko Wanders¹ 

¹Department of Physical Geography, Utrecht University, Utrecht, The Netherlands, ²Unit Subsurface and Groundwater Systems, Deltares, Utrecht, The Netherlands

Abstract Water scarcity represents a critical global challenge driven by diverse complex interactions between natural and anthropogenic factors. Long-term water scarcity often results in depletion of water resources in so-called water scarcity hotspots. To understand the interactions among social, ecological, and hydrological components within these water scarcity hotspots, we applied causal discovery to observational time series of socio-economic, meteorological, and ecological variables. This resulted in a network representing the causal relations between these variables and Terrestrial Water Storage (TWS). Recognizing the limitations of causal discovery, we supplemented the network with expert knowledge. From this we derived Structural Causal Models (SCMs) that simulate the causal mechanisms influencing TWS trends at the water scarcity hotspots. The resulting SCMs have a variable performance with a median r^2 of 0.67 compared to TWS observations. The SCMs allowed us to estimate the impact of anthropogenic and natural changes on TWS variability at water scarcity hotspots. Our analysis confirms population growth as the most significant cause of TWS change in hotspots. As such, this study demonstrates how causal discovery and SCMs can enhance modelling of human-water system dynamics affected by water scarcity, improving the understanding of these systems and potential impacts of future changes on water storage and availability. For future research, more detailed data on human-water use is needed to improve the robustness of these models. This is essential for developing effective water management strategies to mitigate water scarcity at hotspots.

Plain Language Summary Water scarcity is a critical issue, especially in regions with severely depleted water resources. This study shows how causal modelling techniques can be used to understand what drives water scarcity in these areas. The causal model connects different factors like rainfall, temperature, agriculture, population, and water storage. The resulting network shows how these elements interact. Our findings confirm that population growth impacts water storage the greatest, more so than changes in climate. This study also highlights the need for better data on water use and more robust methods to identify the causes of water scarcity. Understanding these causes is important for developing strategies to manage water scarcity that address the needs of each affected region.

1. Introduction

Freshwater resources are vital to humanity and ecosystems. However, at least half a billion people live in water scarce conditions all year round (Mekonnen & Hoekstra, 2016). Water scarcity and its driving mechanisms have been widely studied using modelling applications (Dolan et al., 2021; Felfelani et al., 2017; Greve et al., 2018; Haddeland et al., 2014; Hanasaki et al., 2018; Kim et al., 2016; Kummu et al., 2016; Van Vliet et al., 2017; van Vuuren et al., 2015; Wada et al., 2011). Independent observation-based analysis of water scarcity drivers is, however, in its infancy.

Global modelling efforts found that water resources are under pressure due to excessive human water use ($\sim 1700 \text{ km}^3 \text{ year}^{-1}$) (Qin et al., 2019). Human water use can be separated into different sectoral uses: municipal, industrial, and most importantly irrigation water use. Irrigated agriculture is by far the largest water consumer, accounting for 85%–90% of the total consumptive water use (Qin et al., 2019). About 52% of irrigated crops are irrigated from unsustainable resources, of which 15% is produced for the global trade market (Rosa et al., 2019). At the same time, urbanization, industrialization, and agricultural expansions have been deteriorating water quality of freshwater resources, exacerbating water scarcity (UNEP, 2016; Van Vliet et al., 2017, 2021). Also, the perception of water scarcity may be aggravated by water mismanagement, for example, inefficient water use or unequal water distribution. Moreover, the availability of renewable water resources is affected by hydroclimatic

© 2025. The Author(s).

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

change. At the same time, socio-economic factors, for example, population growth and economic development, are projected to increase future water use (Haddeland et al., 2014; Felfelani et al., 2017; Y. Pokhrel et al., 2021; Desbureaux et al., 2022). Thus far, these developments have reduced streamflow and water storage, leading to shifts towards drier conditions in many regions (Porkka et al., 2024). If these human and climate induced drying trends persist due to an imbalance between human water use and natural water availability, a water gap develops. A water gap is thus defined as water demand that cannot be met by renewable water resources. When the water gap of a region is large and persistent, human water use is often unsustainable, depleting surface and groundwater resources. Regions where such depletion occurs, are defined as the so-called water scarcity hotspots (Leijnse et al., 2024).

Water scarcity has various detrimental impacts on societies and ecosystems. Examples include reduced food production, human health deterioration, or damage to aquatic ecosystems (de Graaf et al., 2019; Dinar et al., 2019; Leijnse et al., 2024). To alleviate such negative impacts of water scarcity, implementation of effective water management is necessary through mitigation strategies or adaptation measures. For finding effective strategies, a detailed understanding of water system dynamics and its drivers at hotspots is crucial. Ways to evaluate drivers of water scarcity include global hydrological models, integrated assessment models, or data-driven methods. For all methodologies sufficient in situ or remote sensing measurements of variables driving water scarcity are essential.

Some of the drivers listed above are hard to measure and quantify on a global scale. For instance, limited observational records of water quantity and quality are available with a sufficiently high temporal resolution and a uniform global distribution (Desbureaux et al., 2022; Jasechko et al., 2024; Jones et al., 2024; Liu et al., 2017; Van Vliet et al., 2021). Observational data sets of water use at the global scale are even more limited. Reported data (e.g., AQUASTAT) is available on country level and 5-yearly scale, missing any interannual and spatial variability (Cárdenas Belleza et al., 2023). Moreover, many countries do not regulate reporting of water use, making the data that is reported uncertain, inconsistent, and incomplete (Huang et al., 2018). Lastly, socio-political drivers, for example, mismanagement of water resources, are even harder to quantify or monitor dynamically over time.

Satellite observations provide a full global coverage of water resource depletion, for example, the Gravity Recovery and Climate Experiment (GRACE) and GRACE-FO satellites (Landerer et al., 2020; Tapley et al., 2004). These satellites detect anomalies in Earth's gravitational field from which changes in Terrestrial Water Storage (TWS; water storage in ice, snow, surface water, soil water, and groundwater) are derived. Such methods have identified regions with long-term trends in surface water depletion, groundwater depletion, or hydrological droughts, and thus water scarcity hotspots in general (Richey et al., 2015; Rodell et al., 2018). Well-known hotspots include the Middle East, the Ganges-Brahmaputra basin, North China Plain, and California in the USA. However, not all important driving mechanisms of water scarcity, such as socio-political factors or human water use, can be measured with remote sensing. Also, the observational period of satellite products is limited to the recent decades.

Current modelling efforts predominantly consist of global hydrological models that assess global water scarcity from a physically based perspective (Greve et al., 2018; Haddeland et al., 2014; Hanasaki et al., 2018; Kummu et al., 2016; Wada et al., 2011) or Integrated Assessment Models that assess water scarcity derived from a socio-economical context (Awais et al., 2024; Dolan et al., 2021; Kim et al., 2016; van Vuuren et al., 2015). A disadvantage of both model types is, however, that they consist of complex relationships. This complexity makes the computation of system changes demanding and slow. Moreover, knowledge on some processes is lacking, and thus based on assumptions. Especially consistent theories on human-water interactions, such as water withdrawal, distribution, and consumption, are lacking (Döll et al., 2016; Kim et al., 2016; Bierkens, 2015; Y. N. Pokhrel et al., 2016; Nazemi & Wheeler, 2015).

Alternatively, data-driven methods, for example, machine learning and regression models, allow for assessing global water scarcity hotspots. Such methods have the advantage to be less computationally demanding, but come at the expense that they do not automatically include theories on system processes, feedbacks, and causal relationships. Even though data-driven methods are often seen as a “black box,” as information on system processes may be hidden within the data (Nearing et al., 2020). Methods that provide more context to data-driven models include causality (Pearl & Mackenzie, 2018), sensitivity analyses (Frey & Patil, 2002), or feature importance (König et al., 2020). Such methods are not commonly used to evaluate driving mechanisms of water scarcity.

Thus far, its potential has been underutilized, especially when focusing on observational data of water scarcity and its drivers.

Here, we focus on data-driven methods that include causality. This topic has been gaining interest in Earth and environmental sciences recently, as new algorithms are being developed to answer causal questions in these fields (Camps-Valls et al., 2023; Nichol et al., 2023; Runge et al., 2023). Three layers of causality exist that each build on the previous layer, starting with causal discovery. Causal discovery is the inference of a causal graph between observed variables, thereby finding causal links between system variables. From this follows the construction of a Structural Causal Model (SCM) that describes how each variable of the system is (in)directly affected by other variables of the system. This model usually comprises of a non-linear function representing the causal mechanism of a target variable and an exogenous noise term. Lastly, interventions (or perturbations) of input data can be applied to the SCM, thereby estimating the causal impact of a variable or counterfactual on the target variable (Runge et al., 2023).

Integrating causal thinking and causal impact into data-driven hydrological modelling offers a different perspective on understanding system dynamics affecting water resources in water-scarce regions, while allowing for a more robust data-driven model (Runge et al., 2023). Moreover, it allows for a dynamic framework in which system changes and feedback mechanisms affecting water scarcity are included. At the same time, such a framework has a low computational demand, allowing to readily assess which system changes impact water storage significantly. Therefore, the objective of this paper is to explore the possibilities to integrate causality in data-driven modelling of TWS changes at water scarcity hotspots, where a decline of TWS is considered as a sign of persistent water scarcity. With such a data-driven model we thus aim to assess the impact of main drivers causing unsustainable water use at water scarcity hotspots.

This study builds on work of Leijnse et al. (2024) that identified global water scarcity hotspots and analyzed the drivers, pressures, states, impacts, and responses of water scarcity through a literature review. Here, we extend this analysis by focusing exclusively on the physical causes of water scarcity, disregarding any occurrence of water scarcity perceptions due to water quality, inaccessibility, or socio-political issues. We employ a data-driven causal inference approach that has not previously been applied for such purpose. More specifically, we apply a causal discovery method (JPCMCI⁺) to observational time series of socio-economic variables (population), meteorological drivers, and a vegetation index with TWS as target variable to obtain a causal network representing human-water system interactions at 21 water scarcity hotspots. After conducting and evaluating the causal network at global scale, we construct a data-driven SCM for each water scarcity hotspot. These SCMs represent the causal relationships between each observed variable and changes in TWS as they are based on the identified global causal network. Lastly, we apply a causal (counterfactual) impact analysis to estimate the sensitivity of each SCM, and thus TWS, to changes in each model input variable.

2. Methods

2.1. Data

To find a network of causal links that relate to water scarcity at the hotspots, globally available observational data sets were used as input data for each hotspot (Table 1, Figure 1a). The location and extent of each hotspot has been predetermined by Leijnse et al. (2024). By using globally available data, we ensure that the methodology is consistent at all hotspots. The choice of which variables to include, was based on a priori knowledge obtained from the DPSIR (Driver, Pressure, State, Impact, Response) framework results, that have found the key drivers and pressures of water scarcity hotspots (Leijnse et al., 2024). We chose TWS measured by GRACE as the target variable representing “water scarcity”. The rationale to use TWS as target is that GRACE picks up the signal of persistent water scarcity (Rodell et al., 2018), which is over time resolved by hotspots through depleting surface water and groundwater resources (Leijnse et al., 2024). Another criterion of the input data was that it should at least cover June 2002 to December 2019 at a monthly resolution, as this aligns with the observational period and temporal resolution of GRACE. These criteria restricted which variables we could include in our analysis, as some important driving variables are not available at global scale or at this temporal resolution and period. Global observational data of water use is, for example, reported on 5-yearly basis or does not cover the required temporal period, for example, Huang et al. (2018). Water use was thus replaced by indirect (driving) variables. For any sectoral water use (domestic, industrial, or agricultural water use) we use population count as a proxy, as is also

Table 1
JPCMCI+ Input Variables Obtained From Global Observational Data Sets With Their Resolutions Before Preprocessing

Variable	Data set	Spatial resolution/cell size at equator	Temporal resolution
Terrestrial Water Storage (TWS)	GRACE	3°/200 km	Monthly
Temperature (t2m)	ERA5	0.5°/50 km	Monthly
Precipitation (pr)	ERA5	0.5°/50 km	Monthly
Population count (pop)	WorldPop	30 arcsec/1 km	Yearly
Discharge (Q)	PCR GLOBWB 2	5 arcmin/10 km	Monthly
Enhanced Vegetation Index (EVI)	MODIS	0.5°/50 km	Monthly
Irrigated area	Nagaraj et al. (2021)	5 arcmin/10 km	Yearly

Note. More details on data references and preprocessing are available in Table A1.

often the case in other modeling frameworks (Bijl et al., 2016; Davies & Simonovic, 2011; Hejazi et al., 2014; Wada et al., 2011). Additionally, we combined the Enhanced Vegetation Index (EVI) and irrigated area as an indirect measure of both land use change and irrigation intensity, thus affecting irrigation water use. A more detailed overview of data sets used as input can be found in Table A1.

Before applying the observational data to the JPCMCI+ algorithm (Section 2.2), we preprocessed the data. First, all globally gridded databases were clipped according to the boundaries of each individual water scarcity hotspot (Figure 1a) (Leijnse et al., 2024). The EVI grid cells were additionally clipped according to grid cells where irrigation occurs as estimated by Nagaraj et al. (2021). This limits the EVI to irrigated croplands. Next, the zonal average (or zonal sum for total precipitation and population count) over each hotspot was estimated for each variable, resulting in time series of 6 variables for each of the 21 hotspots. The GRACE data contains missing data and resulting data gaps. About 15% of the time series is missing, of which July 2017 to May 2018 is the largest gap. This gap will have consequences for the causal discovery and Structural Causal Model (SCM) training later on (Section 2.2 and 2.3). To increase the resolution of the population data from yearly to monthly resolution, we applied a linear interpolation between each known data point. We added randomly generated noise with a maximum standard deviation of 5% of the yearly values to the interpolated population values. This approach preserves the long-term trends while introducing potential sub-annual variability. The noise also portrays that we

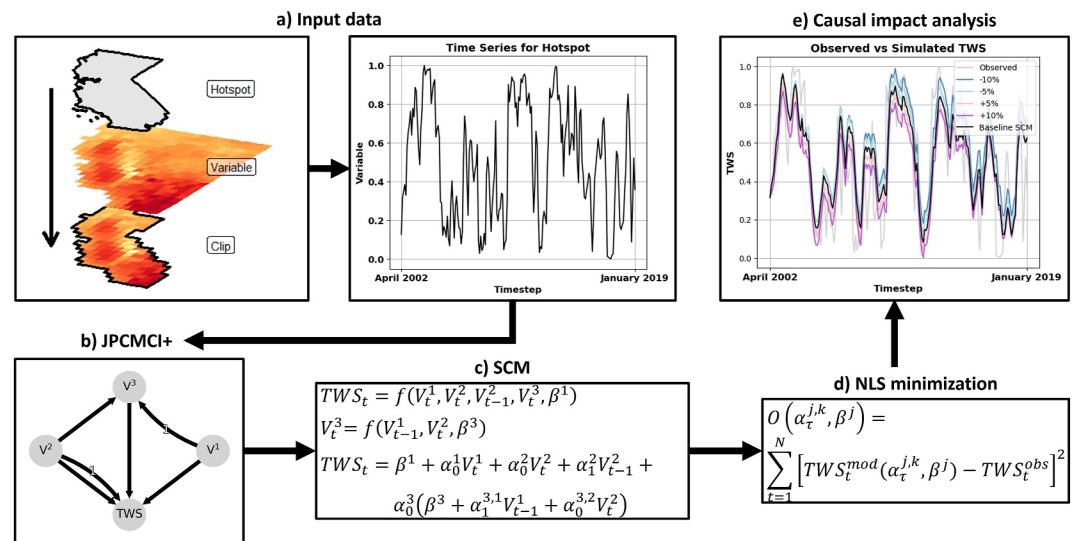


Figure 1. Generalized workflow of (a) clipping global input data sets per hotspot to obtain a time series per variable, (b) example causal network graph obtained from the JPCMCI+ causal discovery algorithm applied to Terrestrial Water Storage (TWS) and independent observational variables (V^1 , V^2 , V^3), (c) derivation of an example Structural Causal Model (SCM) based on the causal graph, (d) Nonlinear Least Squares minimization to obtain values for coefficients $\alpha_t^{j,k}$ and β^j , and (e) example of TWS impact analysis results where input variable V^j is changed with different percentages.

have no information about the actual variability within a year. This lacking knowledge about sub-annual variability reduces the explanatory power of population as a driver. Then, the long-term (2002–2019) monthly average and linear trend were subtracted from each time series to remove seasonality and trends, that would otherwise interfere with identifying causal relationships (Runge et al., 2023). Lastly, all time series were quantile normalized, ensuring that both magnitude as well as distribution of the time series become comparable.

2.2. Causal Discovery JPCMCI⁺

Causal discovery or causal learning is a conditional and statistical independence-based approach of obtaining a causal graph from observational data (Peters et al., 2017; Runge et al., 2023). It qualitatively reconstructs links of a causal graph towards a particular target variable, in our case TWS (Figure 1b). The approach comes with a list of assumptions, depending on which method used. J(oint)-PCMCI⁺, like any other Partial Correlation Momentary Conditional Independence algorithm, learns causal time series graphs, including time lags and bi-directed links for latent (hidden) contexts (Günther et al., 2023). It is a constraint-based approach, which assumes:

1. all data to be part of an acyclic time-dependent SCM, where exogenous noise variables are jointly independent;
2. the Markov condition: if a (conditional) dependency occurs between variables, there is causality between those variables, making the variable conditionally independent of any other variable;
3. faithfulness: a d-separation implies (conditional) independence, meaning that there is no causality between the two variables.

The JPCMCI⁺ algorithm starts with a fully connected graph between all variables. Then it iteratively applies independency tests between each input variable, while conditioning for a set of other input variables. The number of variables included in the conditional set increases for each iteration, until an independency is found. If no independency is found between two variables conditional to any of the input sets, a causal link between the two variables remains in the graph. This identifies a so-called “causal parent” of a variable, meaning that this variable directly influences another variable in the system. In this study we used the partial correlation test as conditional independency test, which finds linear partial correlations for data with Gaussian distributions. If during the test the p-value between two variables exceeds the alpha threshold, the two variables are assumed to be (conditionally) independent. Their causal link is thus removed. In this study, we apply three different alpha values (0.001, 0.01, 0.05) and choose the alpha resulting in a Directed Acyclic Graph (DAG) that is preferred for defining an SCM (Section 2.3). The test also takes lagged timesteps into account, so dependencies between variables at the current timestep and previous timesteps from another or the same variable (autocorrelation). To include system memory within 1 year in the causal graph, we considered a range from 1 to 11 months of time lags. For all links in the graph that are not rejected by the independency testing, the direction of each link is determined by several orientation rules. For example, the v-structure rule: if X and Y are independent, while X and Z , Y and Z are dependent, links are directed as: $X \rightarrow Z \leftarrow Y$. For time lags the orientation will always be forward in time (Runge et al., 2019).

As the sample size of each time series per hotspot and variable is small ($N = 213$) and contains data gaps (Section 2.1), learning a causal graph for each individual hotspot gives inconclusive results (Nichol et al., 2023; Runge et al., 2023). An advantage of the JPCMCI⁺ method is that it allows for pooled (joint) data sets as input, meaning that it can include time series from different spatial contexts. In this study we pooled the time series of each system variable at the hotspots together ($N = 213 \times 21 = 4,473$). By doing so, the JPCMCI⁺ algorithm learns a causal graph that represents the system at all hotspots.

2.3. Structural Causal Model (SCM)

We constructed the SCM according to the causal graph found by the JPCMCI⁺ causal discovery method (Section 2.2), that estimates (in)direct causes of TWS change including lagged timesteps and autocorrelations. However, causal network discovery often results in a Partially-Directed Acyclic Graph (PDAG) with undirected edges, whereas an SCM can only be defined from a DAG (Nichol et al., 2023). Therefore, we decided to supplement the causal graph with a priori expert knowledge, as we expect all input variables to have at least one causal link directed towards TWS (Section 2.1). In this study, we consider the following generalized SCM structure (Runge et al., 2023):

$$TWS_t := f[pd(V_{t-\tau, \tau \geq 0}), \eta_t] \quad (1)$$

where f is a (non-)linear function of the causal parents $\text{pa}(V_{t-\tau})$ of TWS_t and η_t is an exogenous noise term representing the non-explained part of the TWS_t signal. The causal parents $V_t = (V_t^1, V_t^2, \dots, V_t^J)$ are the system input variables from the underlying multivariate time series (t2m, pop, pr, EVI, Q , TWS; Table 1). The subscript t represents the timestep with τ as time lag, which in our case has a maximum value of 11 months.

As the causal discovery method tests for linear correlations (Section 2.2), the SCM of TWS_t is a multivariate function. In this function each parent variable is assumed to be linearly related to TWS. This results in the following general equation:

$$\text{TWS}_t = \beta^0 + \sum_{\tau=1}^{\tau_{\max}} (\alpha_{\tau}^0 \text{TWS}_{t-\tau}) + \sum_{j=1}^J \sum_{\tau=0}^{\tau_{\max}} (\alpha_{\tau}^{j,0} V_{t-\tau}^j) + \eta_t^0 \quad (2)$$

And the causal parent variables themselves are also related to each other as:

$$V_t^j = \beta^j + \sum_{\tau=1}^{\tau_{\max}} (\alpha_{\tau}^j V_{t-\tau}^j) + \sum_{\substack{k=1 \\ k \neq j}}^K \sum_{\tau=0}^{\tau_{\max}} (\alpha_{\tau}^{j,k} V_{t-\tau}^k) + \eta_t^j, j = 1, \dots, J \quad (3)$$

where α is the coefficient and β the intercept of each linear function that defines a causal parent of TWS. The j refers to the corresponding causal parent, k corresponds to indirect causal parents, and $\tau_{\max} = 11$, as explained above. The $\sum (\alpha_{\tau}^0 \text{TWS}_{t-\tau})$ and $\sum (\alpha_{\tau}^j V_{t-\tau}^j)$ refer to the autocorrelation of TWS_t and V_t^j , respectively. Note that if we substitute Equation 3 for any of the parent variables into Equation 2, we obtain an equation that is non-linear (polynomial). An example of such a substitution can be found in Figure 1c. In this example autocorrelation is excluded.

2.4. Non-Linear Least Squares Minimization

Parameters α and β are estimated by minimizing the sum of squares of the residuals from the modelled ($\text{TWS}_t^{\text{mod}}$) and observed ($\text{TWS}_t^{\text{obs}}$) TWS (Figure 1d):

$$O(\alpha_{\tau}^{j,k}, \beta^j) = \sum_{t=1}^N [\text{TWS}_t^{\text{mod}}(\alpha_{\tau}^{j,k}, \beta^j) - \text{TWS}_t^{\text{obs}}]^2 \quad (4)$$

with N as the length of the observed TWS time series and O as objective function. We used the ‘‘Levenberg-Marquardt’’ method to minimize the objective function. This method iteratively finds the optimal value for each parameter in the objective function within a trust-region shape (Moré, 2006).

We estimated a non-linear SCM separately for each hotspot. As a result, we obtain 21 SCMs that are all specifically trained on observational data from each hotspot area. Consequently, the data sets used across the different hotspot areas are consistent, while resulting SCMs vary. This variation arises from the local variability affecting the relationship between input variables and TWS. The full observational time series for each hotspot were used as input to train each model and find optimal parameters. Moreover, when an autocorrelation occurred between TWS_t and $\text{TWS}_{t-\tau}$, we used the observed TWS only as input for the initialization (first timestep; TWS_0) of each model. For the subsequent timesteps, the simulated $\text{TWS}_{t-\tau}$ was used as input. This approach better captures the propagation of simulation errors, instead of correcting the simulated model with observed data from the previous timestep.

The significance of the model parameters was estimated by bootstrapping: resampling a subset of 80% of input data 100 times to obtain the variance of each parameter, followed by a t -test for parameter significance. The resulting set of 100 TWS simulations were then used to estimate the 90% confidence interval of each SCM. Another randomly split test set (20%) was used to evaluate model performance. We calculated evaluation metrics to compare the modelled TWS to the test set of TWS observations for corresponding timesteps. The metrics calculated are r^2 , Mean Squared Error (MSE), and Mean Absolute Error (MAE). Additionally, we calculated the residuals (noise terms; η_t) of each SCM, their mean and standard deviation. We also estimated the lag-one

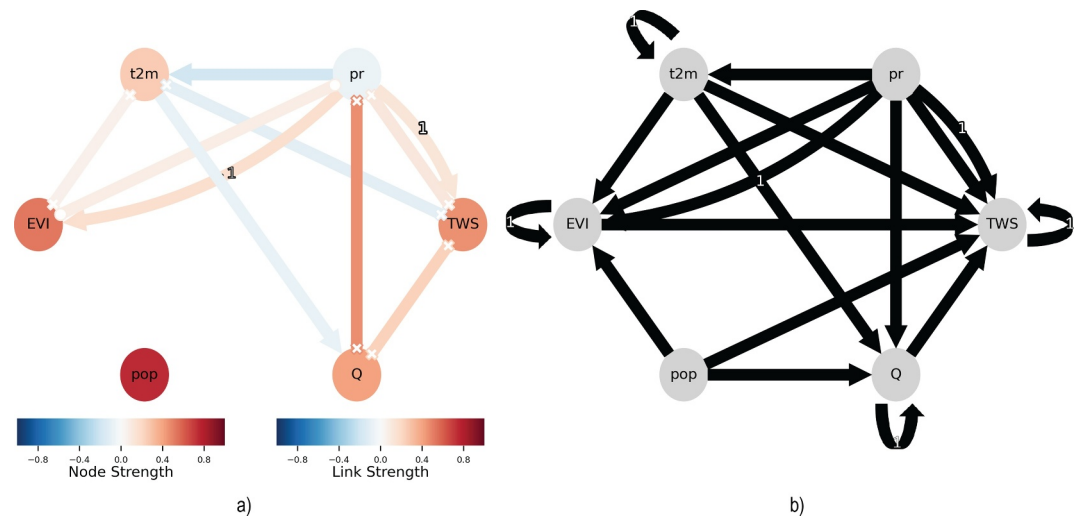


Figure 2. Causal graphs of target variable Terrestrial Water Storage (TWS). Each node (circle) represents a variable (full variable names are given in Table 1). The \rightarrow or $x-x$ lines are the links between the variables. A link with the number “1” indicates a time lag of 1 month. (a) Causal graph resulting from JPCMCI⁺. The link strength shows the result of the partial correlation test, with in red a positive dependency and in blue a negative dependency. The node strength shows the strength of the autocorrelation. The $x-x$ links indicate an undecided link direction due to conflicting orientation rules. (b) Expert graph with causal links assumed for the Structural Causal Model. A link from a node to itself with a “1” indicates a 1-month autocorrelation.

correlation metrics of the residuals per hotspot. Lastly, we applied the Shapiro-Wilk test to assess the normality of residuals (Shapiro & Wilk, 1965). These metrics of η_t indicate to what extent the SCMs simulate all causal mechanisms of TWS change or whether unidentified causal relationships or unobserved variable bias are still present in the residuals.

2.5. Causal Impact Analysis

To quantify the impact of each input variable on TWS, we applied perturbations to each input variable. For this we altered each observational input time series per hotspot by +10%, +5%, -5%, and -10% (for temperature: +1K, +0.5K, -0.5K, -1K). This resulted in 24 scenarios with slightly altered input time series. Then, we predicted the TWS for each hotspot per altered input set (Figure 1e). Subsequently, we calculated the Mean Bias Deviation (MBD) between the predicted TWS by the respective scenario and the simulated TWS by the baseline SCM. Any deviations between the scenario and the original simulation are presumed to reflect the causal impact of a perturbation from an input variable on TWS. The significance of each perturbation is estimated with the 90% confidence interval obtained from bootstrapping the baseline SCMs (Section 2.4). If a perturbation exceeds this confidence interval for 10% of the simulation period at a hotspot, the perturbation is considered to be significant for the corresponding hotspot.

3. Results

3.1. JPCMCI⁺ Causal Graph

Figure 2a shows the causal graph of target variable TWS resulting from the JPCMCI⁺ algorithm. The causal graph of Figure 2a shows the generalized system across all hotspots, relating potential causes of water scarcity to TWS. The graph nodes (circles) represent each input variable (Table 1), while the links (\rightarrow or $x-x$) indicate dependencies between variables as identified by JPCMCI⁺. The graph is a PDAG, where some link edges are undirected ($x-x$ link) due to conflicting orientation rules. Link colors represent the partial correlation effect size between variables. The node colors indicate autocorrelation strength. Red implies positive correlations, whereas blue suggests negative correlations.

The graph shows precipitation to be positively related to TWS with no time lag and an one month time lag. Undirected links with TWS are found for discharge and temperature, where discharge is positively related to TWS and temperature is negatively related. EVI is only indirectly connected to TWS, while population count shows no

direct or indirect links with TWS. Precipitation indirectly relates to temperature (negatively), discharge (positively), and EVI (positively including an one month time lag). Moreover, all variables have positive autocorrelations, except for precipitation.

3.2. Expert Causal Graph

The data-driven causal graph found by the JPCMCI⁺ algorithm (Figure 2a) was supplemented with additional expert knowledge to construct a directed, acyclic expert causal graph (Figure 2b) required for defining a Structural Causal Model (SCM). By expert knowledge we mean that the authors make physically substantiated choices that give direction to undirected links or add directed links to variables that are not connected in the graph.

In the expert graph all input variables (temperature, precipitation, EVI, population, and discharge) are directly linked to target variable TWS since these input variables were specifically selected assuming they are key drivers of water scarcity (Section 2.1). Moreover, we have provided the undirected link between temperature and EVI with a direction from temperature to EVI. This originally undirected link captures the complex interaction between irrigated cropland greenness and temperature (Yang, Jin, et al., 2023). However, we cannot build an SCM with bidirectional links. Therefore, we choose to direct temperature towards vegetation, as large-scale temperature changes (ERA5) are more likely to affect cropland greenness via evapotranspiration than vice versa (Yang, Roderick, et al., 2023). As EVI represents irrigated cropland greenness, we included a link between population and EVI to represent the effect of population (growth) on agriculture. We also added a link between population and discharge to include the impact of human activities on discharge (Gerten et al., 2008; Veldkamp et al., 2018). We did not add any additional lagged links in our expert graph. Overall the scientific evidence and our data preprocessing approach suggest that both population effects and vegetation impacts on TWS are predominantly immediate rather than lagged, so including lagged links adds little explanatory value (Xie et al. (2019); Section 2.1). Lastly, autocorrelations are included for all variables that have causal parents. These autocorrelations will become important for defining the SCM function that estimates those variables. For precipitation and population, we did not include the autocorrelation in the expert graph, as they do not have causal parents. Consequently, their autocorrelation will have no part in the SCM equations.

3.3. Structural Causal Model (SCM)

To translate the casual graph into a modelling framework that can reproduce TWS dynamics, we created a SCM (SCM) from the expert causal graph. This SCM consists of the following equations:

$$TWS_t = \beta^0 + \alpha_1^0 TWS_{t-1} + \alpha_0^1 t2m_t + \alpha_0^2 tp_t + \alpha_1^2 tp_{t-1} + \alpha_0^3 pop_t + \alpha_0^4 EVI_t + \alpha_0^5 Q_t + \eta_t^0 \quad (5)$$

With

$$t2m_t = \beta^1 + \alpha_1^1 t2m_{t-1} + \alpha_0^{1,2} tp_t + \eta_t^1 \quad (6)$$

$$EVI_t = \beta^4 + \alpha_1^4 EVI_{t-1} + \alpha_0^{4,1} t2m_t + \alpha_0^{4,2} tp_t + \alpha_1^{4,2} tp_{t-1} + \alpha_0^{4,3} pop_t + \eta_t^4 \quad (7)$$

$$Q_t = \beta^5 + \alpha_1^5 Q_{t-1} + \alpha_0^{5,1} t2m_t + \alpha_0^{5,2} tp_t + \alpha_0^{5,3} pop_t + \eta_t^5 \quad (8)$$

Substitution of $t2m_t$, EVI_t , and Q_t gives:

$$\begin{aligned} TWS_t = & \beta^0 + \alpha_1^0 TWS_{t-1} + \alpha_0^1 (\beta^1 + \alpha_1^1 t2m_{t-1} + \alpha_0^{1,2} tp_t) + \alpha_0^2 tp_t + \alpha_1^2 tp_{t-1} + \\ & \alpha_0^3 pop_t + \alpha_0^4 [\beta^4 + \alpha_1^4 EVI_{t-1} + \alpha_0^{4,1} (\beta^1 + \alpha_1^1 t2m_{t-1} + \alpha_0^{1,2} tp_t) + \alpha_0^{4,2} tp_t + \\ & \alpha_1^{4,2} tp_{t-1} + \alpha_0^{4,3} pop_t] + \alpha_0^5 [\beta^5 + \alpha_1^5 Q_{t-1} + \alpha_0^{5,1} (\beta^1 + \alpha_1^1 t2m_{t-1} + \alpha_0^{1,2} tp_t) + \\ & \alpha_0^{5,2} tp_t + \alpha_0^{5,3} pop_t] + \eta_t \end{aligned} \quad (9)$$

here, $t2m$, pop , pr , EVI , Q , and TWS refer to the observational input time series (Table 1). η_t is the exogenous noise term, and $\alpha_t^{j,k}$ and β^j are the parameters to be fitted, where β^j is the intercept of each linear equation. These parameters were obtained by fitting the baseline SCMs to TWS observations at each hotspot using NLS

Table 2
Performance Metrics of Each Baseline SCM Per Hotspot Compared to the TWS Observational Time Series With Three Model Evaluation Metrics: r^2 , Mean Absolute Error (MAE), and Mean Squared Error (MSE)

Hotspot	r^2	MAE	MSE
Arabian Peninsula	0.97	0.04	0.00
California	0.73	0.11	0.02
Central Chile	0.47	0.18	0.04
Coastal Peru	0.63	0.15	0.03
Ganges	0.82	0.11	0.02
Greece	0.63	0.14	0.03
Indus	0.84	0.09	0.01
Iran	0.89	0.08	0.01
Italy	0.64	0.14	0.03
Japan	0.18	0.21	0.06
Java	0.77	0.14	0.02
Mexico	0.60	0.16	0.04
Murray-Darling	0.63	0.15	0.03
Nile Delta	0.49	0.18	0.04
North China Plain	0.76	0.12	0.02
Spain	0.67	0.13	0.03
Thailand	0.63	0.15	0.03
Türkiye	0.80	0.10	0.02
U.S. High Plains	0.72	0.12	0.02
Vietnam	0.40	0.19	0.05
White Nile	0.86	0.09	0.01

minimization. Their resulting values and significance can be found in Tables B1 and B2. Note that by forward simulating TWS, we use observed values for each causal parent that has no causal links from other variables (*pop* and *tp*) and simulated values for those that have (indirect) causal links (*t2m*, *EVI*, and *Q*) to preserve the causal structure. For the lag-one autoregressive term of TWS_{t-1} (except for $t = 0$), we also use the simulated TWS values to assure the simulation to be open loop.

The performance of each baseline SCM for individual hotspots is described in Table 2 and shows a varying performance of the SCMs. This becomes evident looking at the r^2 in Table 2 that varies between 0.97 (Arabian Peninsula) and 0.18 (Japan) with a median of 0.67. This difference in SCM performance is also shown in Figure 3. The figure indicates that the 90% confidence interval of the Arabian Peninsula is much smaller than for Japan. The normalized MAE and MSE are on average 0.13 and 0.03, respectively. Furthermore, statistical analysis of the residuals (η_t) of each SCM shows that the mean is close to zero. Together with the Shapiro-Wilk test results these findings suggest a normal distribution of the residuals for some hotspots (Table B3). The lag correlation, however, indicates a strong signal of autocorrelation of the residuals, thus the residuals are not white noise. This result suggests that our SCMs contain unobserved (parent) variable bias, which is to be expected when simulating regional water storage variability with global data sets.

3.4. Causal Impact Analysis

Each altered SCM input set affects TWS differently, depending on the input variable and hotspot location. Figure 4 shows the impact of altering each input variable of the SCMs on the simulated TWS and their significance at all hotspots. In this figure we compare each scenario with each corresponding baseline SCM through calculating the MBD. Altering the initial value of TWS (TWS_0) does not affect TWS much at any hotspot as MBD values are close to zero and the perturbations are insignificant at all hotspots. This indicates that

the error propagation from the initial value of TWS is low. The perturbation of discharge input values does not affect the TWS much either: it is insignificant at all hotspots and the average MBD is close to zero. In contrast, changing the values of temperature, precipitation, or EVI give larger changes in TWS. The average MBD ranges from -0.08 to 0.07 for precipitation, while for temperature and EVI the average MBD is again almost zero.

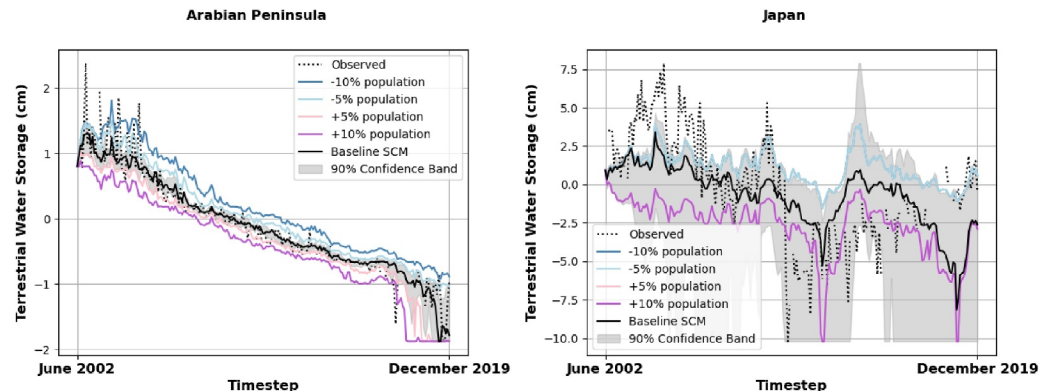


Figure 3. Example of causal impact analysis results showing the impact of a population perturbation of $\pm 5\%$ to 10% on Terrestrial Water Storage (TWS). The TWS perturbations are compared to the TWS simulated by the baseline Structural Causal Model (SCM) and TWS from the observations at two different hotspots: the Arabian Peninsula (left) and Japan (right). The 90% confidence band of the baseline SCM is displayed in gray. In Japan the $\pm 5\%$ and $\pm 10\%$ scenarios are overlapping. Seasonal trends are removed in both graphs for improved visualization of the long-term TWS trends.

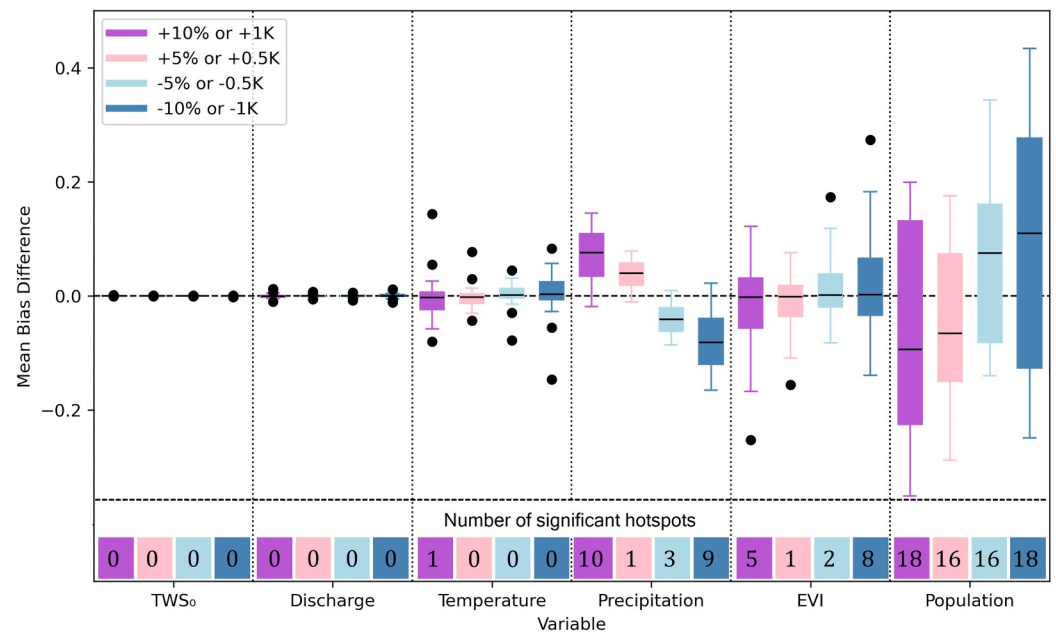


Figure 4. Boxplots of Mean Bias Difference (MBD) between Terrestrial Water Storage (TWS) simulated by the Structural Causal Models (SCMs) for each perturbation scenario and the TWS simulated by the baseline SCMs at each hotspot. The numbers at the bottom show the number of hotspots where the impact of the perturbation on TWS is significant.

Another difference between these three variables is that the temperature perturbation is insignificant at almost all hotspots, whereas for precipitation and EVI the significance varies per hotspot. Population changes have a relatively large effect on TWS (average MBD ranges from -0.08 to 0.10) and are significant at most hotspots. An example of resulting TWS changes caused by a population perturbation for the Arabian Peninsula and Japan compared to the observed and simulated TWS is displayed in Figure 3. Other per hotspot results from the impact analysis can be found in the Supporting Information S1.

Overall, the scenarios of $\pm 10\%$ or $\pm 1K$ have a larger effect on TWS than the smaller changes of $\pm 5\%$ or $\pm 0.5K$. This may not always be the case for every hotspot and variable. Figure 3 shows that for Japan the resulting changes in TWS due to $\pm 5\%$ and $\pm 10\%$ population impact TWS the same. This occurs when a small change in the input variable, in this case population, results in a perturbation that exceeds any observed input value. This is because the applied quantile normalization to the perturbation follows the distribution of the observed input data set. Consequently, both scenarios of 5% and 10% change have the same input value.

The MBD boxplots of Figure 4 indicate whether perturbations increase or decrease TWS compared to the baseline model. A positive MBD shows a TWS increase and negative MBD a TWS decrease. The MBD results show that increasing or decreasing temperature or EVI can both impact TWS positively and negatively depending on the hotspot. Precipitation, however, shows a positive relationship with TWS, as increased precipitation increases TWS, while decreased precipitation decreases TWS. The reverse applies to population. Increased population decreases TWS, while decreased population increases TWS. At a few hotspots, however, increased population also increases TWS.

4. Discussion

4.1. Impact of Water Scarcity Causes on Terrestrial Water Storage

Changes in Terrestrial Water Storage (TWS) can be attributed to multiple different causes, such as natural spatio-temporal variability or anthropogenic interventions (Felfelani et al., 2017; Haddeland et al., 2014; Rodell et al., 2018). In our study, results show that changes in population, potentially serving as a proxy for anthropogenic interventions, have a larger impact on TWS change than climate change factors (in/decreased precipitation or increased temperature). Other modelling studies have also evaluated that human interventions (population growth, water withdrawals, consumption, and management) have the largest impact on TWS change,

especially for regions with significant irrigation (Haddeland et al., 2014; Kummu et al., 2016; Porkka et al., 2016; van Vuuren et al., 2015; Wada et al., 2011). Methods in such studies differ significantly from our data-driven causal modelling approach. To our knowledge, data-driven approaches on this topic are limited, making direct comparisons challenging. However, two case studies identified population and agricultural water use as key drivers of TWS change in the North China Plain using feature importance analysis from machine learning models (Jing et al., 2019; Ou et al., 2023). Although these methods do not include causality, the results agree with our analysis.

When looking at hotspot specific results, an increasing population decreases TWS most severely in Türkiye, Iran, and Thailand. When comparing these results to the literature review from Leijnse et al. (2024), population growth is estimated as a driver of water scarcity at these three hotspots, although it is not identified as a more pronounced driver of water scarcity in comparison to other hotspots. It is, however, difficult to explain why in certain hotspots (Central Chile, Coastal Peru, Greece, Java, Vietnam, and the White Nile) population growth increases TWS significantly. The observational period of GRACE from 2002 to 2019 could be an explanation. This period does not capture a (strong) decreasing trend of TWS at these hotspots, due to, for example, a progression from a dry to a wet period in these areas (Rodell et al., 2018). Therefore, these SCMs consider the population increase to be positively related to TWS over the model training period.

The impact of climate change or variability (precipitation and temperature) is not identified as most dominant, but still has a profound effect on the TWS in some hotspots. This finding agrees with other analyses, for example, Huo et al. (2024) and Haddeland et al. (2014). Increased precipitation variability has the largest effect on TWS changes on the Arabian Peninsula, in the Nile Delta and Iran. When comparing this to the literature review of Leijnse et al. (2024), hydroclimatic change has been identified as a driver of water scarcity in these hotspots. Haddeland et al. (2014) estimated that climate change has a larger impact on runoff than human impacts in the Mediterranean region, Central Chile, and the Murray-Darling basin in Australia. Although using a different methodology and approach, our findings confirm that the impact of precipitation variability is larger or is similar to the impact of population changes on TWS for these specific regions found by Haddeland et al. (2014).

The response of TWS was mixed for EVI and temperature perturbations. These mixed responses may be explained by multiple complex interactions and feedbacks between vegetation and water availability. Our JPCMCI⁺ causal graph already identified the complexity, as it was unable to find a directional causal relationship. Depending on the regional context, increased vegetation cover and temperature can lead to both water loss due to enhanced evapotranspiration and water supply due to evaporation-recycling (Yang, Jin, et al., 2023). Moreover, these interactions have different small scale effects (between soil and vegetation) than at a larger scale (between land surface and atmosphere). On a large scale, the irrigation-precipitation feedbacks have a large variability in spatial extent (Theeuwes et al., 2023). In some hotspots moisture recycling is mostly local, for example, the U.S. High Plains (Deangelis et al., 2010), while in other hotspots precipitation from irrigation does not return within the hotspot area, for example the Arabian Peninsula or the North China Plain (Song et al., 2024; Zampieri et al., 2024). These complex positive and negative feedback mechanisms across multiple scales may also explain why temperature and EVI perturbations do not significantly impact TWS at more than half of the hotspots, as effects may counteract. To gain a deeper understanding of the interaction between EVI, temperature, and water storage change, further research is required on small and large scale interactions at each hotspot.

Surface water is part of TWS measured by GRACE. As TWS should fluctuate with the magnitude of hydrological fluxes (Huo et al., 2024), we expected the discharge perturbations to have a more pronounced effect on TWS changes. However, our results suggest that discharge is a small hydrological flux in water scarcity hotspots relative to other fluxes that alter TWS: groundwater depletion, soil drying, and snowmelt. Thus, the decreasing TWS over the past decades at hotspots is unlikely the result of change in discharge and more likely the result of unsustainable groundwater use, as has been suggested by other studies (Felfelani et al., 2017; Rodell et al., 2018).

4.2. The Practicality of Causal Discovery for Water-Scarce Systems

Water scarcity is part of a complex temporal system, where water deficiencies develop and propagate slowly through the system (Schwalm et al., 2017). Only a few causal discovery methods are temporally-aware and include time lags, of which (J)PCMCI is one. The open source availability of the algorithm makes this method readily applicable. However, the method comes with a list of requirements and assumptions, reducing a successful performance when applied to such complex systems with limited independent observations available.

First, achieving the required sample size of each time series for an acceptable performance of (J)PCMCI ($N > 1000$) is challenging (Nichol et al., 2023). In our case we initially intended to discover separate causal graphs for each hotspot, so that we could compare the causal networks across hotspots. However, the maximum sample size we could obtain from global data sets was only 213 monthly timesteps, which was too low for stable causal graph results. Therefore, we adopted a global approach to obtain the JPCMCI⁺ causal graph, which reflects the general causal relationships for all hotspots combined. As a consequence, we lost the specificity in the causal graph and could not identify differences in local water scarcity causal dynamics. These are essential for finding local solutions, as variability in hydrological response is large across different regions (Chávez García Silva et al., 2024). However, the current data availability limits this more localized approach.

Additionally, the partial correlation test has a low computational demand, allowing for quick computation of system changes for the causal impact analysis. A disadvantage of the partial correlation test, however, is that it only identifies linear-Gaussian systems. Not all observed data sets in Earth sciences reflect such linear-Gaussian structures, which is also the case for water-scarce systems. Therefore, the performance of (J)PCMCI may be suboptimal for our analysis with complex spatio-temporal dynamics (Nichol et al., 2023). Accordingly, any direct links between the input variables that are non-linear remain unknown in our analysis. Other conditional dependency testing methods that could identify such links are available, but require even larger sample sizes. With the current temporal resolutions of global data sets it is infeasible to apply such methods to our analysis, thus limiting potential further causal assessment of water scarcity at the global scale.

The accuracy of a causal graph depends on the quality of the input data sets. Uncertainties across data sets may infer spurious links in the resulting causal graph. Moreover (J)PCMCI requires data inputs to be observed variables from data sets without controls or planned interventions (Nichol et al., 2023). For our analysis we have not strictly used only observed variables, as observed time series of variables affecting water scarcity are limited on a global scale. In this study, temperature and precipitation are products of reanalysis (a combination of observations and model estimations), while discharge is obtained from a global hydrological model. We chose these modelled products because they have a high temporal and spatial coverage. For reanalysis the purpose of the model estimation is to improve the quality and availability of the observations (Hersbach et al., 2020). So, the model has minimal control over the observations. Modelled discharge, however, is a controlled product that depends on pre-identified physical relations (Sutanudjaja et al., 2018). As a consequence, the causal links of discharge in the JPCMCI⁺ graph should be interpreted with caution. These links may also be a result of physical relationships that are described in the governing modelling equations. Similar input variables as the global hydrological model are, however, avoided in our input data set. Still, we decided to include discharge in our input data, as no alternative exists and it reflects the (un)importance of surface water changes in TWS variability.

An alternative method for causal discovery of time-dependent systems is transfer entropy. This method filters out the autoregressive component of the target variable to assess the information flow from a conditional set of input variables to the target variable. An advantage of this method is that it allows for identifying non-linear relations (Sun & Bollt, 2014). When applying TEFS (Transfer Entropy Feature Selection; Bonetti et al. (2024)) to our input data, different causal links between TWS and the input data result compared to the JPCMCI⁺ method. Links with population and discharge were found by TEFS, while JPCMCI⁺ found direct links with precipitation, discharge, and temperature. Thus, population may be regarded as a causal parent of TWS, which justifies our additional direct causal link from population to TWS in the expert graph (Figure 2b). However, results of TEFS and JPCMCI⁺ are not directly comparable as autoregression is disregarded in TEFS. In our framework, the JPCMCI⁺ causal graph (Figure 2a) and the autoregressive coefficient α_1^0 (0.73 on average; Table B2) both indicate that the autocorrelation of TWS is large. Moreover, results from transfer entropy do not indicate how variables mathematically relate to TWS. Thus, it does not allow to generate an SCM from the TEFS results alone.

4.3. Relevance of the SCM

Ideally, an SCM will follow directly from a Directed Acyclic Graph (DAG) that is obtained through causal discovery. However, the JPCMCI⁺ causal graph in this study has resulted in a Partially DAG (PDAG). Therefore, the PDAG has been supplemented with expert knowledge to transform it into a DAG, allowing the definition of a closed set of equations for the SCM. Benchmarking such a causal graph is hard, as the ground-truth can only be hypothesized. From Nichol et al. (2023) follows that (J)PCMCI⁺ performance improves when autocorrelation coefficients are large, which is the case for our causal graph (average α_1^0 of 0.73, Table B2), increasing its reliability.

To what extent JPCMCI⁺ can find all causal relationships in complex water-scarce systems on a global scale contains some uncertainty, as discussed in Section 4.2. However, substituting potentially missed links with expert knowledge also has limitations. It introduces a degree of subjectivity to the causal framework. Although the added links are supported by scientific evidence and process knowledge from the literature, they deviate from the data-driven causal relationships identified by the algorithm. Consequently, the extent of the model to remain strictly causal is uncertain, potentially bringing subjectivity to its results.

Evaluation of each baseline SCM per hotspot shows much variation in SCM performance per hotspot (r^2 values; Table 2). Also, the residuals of some SCMs have a strong autocorrelation (mean lag corr η_t of 0.63; Table B3). These values indicate that the generalized global causal graph and SCM structure do not capture all causal relationships at those hotspots. The lower baseline SCM performance at specific hotspots has implications for the reliability and quality of impact analysis results. The generalized patterns resulting from the impact analysis, however, remain relevant, as the median performance of the SCMs is reasonable ($r^2 = 0.67$).

Hotspots with a lower baseline SCM performance or with a large residual autocorrelation, for example, Japan, Vietnam, or Central Chile, would benefit from a hotspot specific SCM with more local input data. Additional variables, such as local policy implementations or sectoral water use measurements, are not captured in global data sets but are relevant to local water scarcity dynamics. These variables may be available in regional data sets and could enhance causal networks and SCM performance. Currently, human water use is represented in our graph by the population variable. The socio-economic and political state of a region, such as prosperity and living standards, are, however, also crucial for the amount of water used (Hoekstra & Chapagain, 2007; Kummur et al., 2016). Additionally, water resource management, particularly irrigation practices, has a large impact on TWS change (Haddeland et al., 2014). While data sets on human-water interactions may be available on local scale, global data sets are essential for cross-hotspot comparisons. Therefore, more comprehensive global observational data sets on human water interactions and management are necessary. A monthly temporal resolution and an observational period of over 20 years of this data are essential for improved causal modelling. Key missing variables at such resolution on the global scale thus include sectoral water use and historical records of water policy implementations. If such products become available in the future, a next step would be to add these to the causal modelling framework, to better investigate human water use and management impacts on water availability.

Future collaboration with local authorities to acquire hotspot-specific input data offers the opportunity to develop locally trained SCMs. Such SCMs could then be compared to our current SCM results for model validation and assessment of causal impact consistency. Additionally, locally trained SCMs have broader applications beyond improving the understanding of water scarcity drivers. Such a causal framework has potential to evaluate the effectiveness of water policies, while accounting for the complex interplay of factors contributing to TWS changes in a region, for example the effect of a water treaty on streamflow (Penny et al., 2020).

TWS change is measured at a coarse spatial resolution and consists of five water storages: surface water, soil moisture, groundwater, snow and ice (Felfelani et al., 2017). To find causes of local variability and water scarcity, each storage component could be isolated by replacing TWS with local measurements of, for example, groundwater head or snow water equivalent. This would require much more data inputs from different local sources, which is currently strongly limited by restricted data accessibility or availability per region. More readily available data on groundwater, reservoir levels, or water use, also with high-temporal resolution, would help to better understand the causal relationships between groundwater depletion, natural variability in the climate system, and the impact of human-water interactions.

Regardless of data limitations, combining causal discovery with the SCM framework, successfully provides more context to data-driven modelling of water scarcity causes. It highlights where causal factors still remain unidentified. This approach has potential to be applied to more localized data sets. As such, it could reveal previously undetected drivers of water scarcity or allow for relating more specified causes to depletion of particular freshwater resources.

5. Conclusion

Physical water scarcity remains a pressing issue, especially at the 21 water scarcity hotspots where severe water depletion is prevalent. While water quality, socio-political factors, and accessibility challenges may exacerbate

the perception of water scarcity, physical causes of water depletion are attributed to a complex interplay of hydroclimatic and socio-economic change. In this study, we focus on the latter and explored the potential of causal discovery and structural causal modelling to identify and assess the impacts of these dynamics driving physical water scarcity at these hotspots.

We applied the JPCMCI⁺ algorithm to global data sets to find causal links between changes in Terrestrial Water Storage (TWS) and various natural and anthropogenic variables: precipitation, temperature, irrigated cropland Enhanced Vegetation Index (EVI), discharge, and population count. However, obtaining a robust and stable Directed Acyclic Graph (DAG) rather than a Partially DAG (PDAG) from JPCMCI⁺ proved to be challenging due to the algorithm's requirement for a high number of independent input data sets with sufficient temporal resolution. Our current analysis allowed for the discovery of a “global” causal network of water-scarce systems, although the causal graph contained undirected links (PDAG), showing some of the limitations of this approach for the problem at hand. Therefore, we supplemented the JPCMCI⁺ causal graph with expert knowledge to direct the undirected links to obtain a DAG. From this expert graph we derived a corresponding Structural Causal Model (SCM). This SCM framework was then solved through non-linear least squares minimization based on TWS observations from each hotspot to establish 21 hotspot-specific baseline SCMs.

The results indicate that the performance of each baseline SCM varies across hotspots, with a median r^2 of 0.67. This r^2 variability and the remaining autocorrelation in the error term η_t suggest that some hotspots are missing important causal parents in their causal network and a corresponding SCM that explains TWS variability. Impact analysis, conducted by altering each input data set by $\pm 5\%$ to 10% (or ± 0.5 to 1K for temperature), revealed that changes in population have the largest impact on TWS change. This finding implies that population dynamics (human factors) have a larger impact on TWS change than climate change factors (changes in precipitation and temperature). While increased population generally causes a decrease in TWS (negative relation), increases in precipitation are positively related to TWS. In contrast, the effects of temperature and EVI on TWS change are more variable, showing both positive and negative impacts depending on the hotspot. This results suggests that different vegetation-temperature interactions related to evapotranspiration and moisture recycling occur per hotspot. Changes in discharge have the smallest impact on TWS change, indicating and confirming findings of previous studies that other hydrological fluxes, such as snowmelt, soil drying, and groundwater depletion, are more dominant in decreasing water storage at the water scarcity hotspots.

The causal discovery and SCM framework provide opportunity to give improved context to data-driven modelling of water scarcity dynamics. It identifies whether all causal mechanisms driving TWS variability are captured by the models and allows for further exploration on unidentified causes. Future research should focus on improving the robustness of causal discovery and expanding the availability of global and uniform local data sets with a high temporal resolution, particularly on human water use, to better understand causes of water scarcity at hotspots. Expansion of this approach towards more localized data could provide valuable insights into a diversity of causal networks of water scarcity, ultimately aiming to give guidance to more targeted and effective water management strategies to mitigate water scarcity.

Appendix A: Input Data

Table A1

Details on Resources of JPCMCI⁺ Target, Input, and Additional Variables From Global Data Sets

Variable	Data source and authors
	DOI
	Abbreviation in causal graph
	Temporal range and resolution
	Spatial resolution/cell size at equator
	Unit
	Description
	Additional preprocessing

Table A1

Continued

Target variable	
Terrestrial water storage	<p>GRACE and GRACE-FO; Wiese et al. (2019)</p> <p>10.5067/TEMSC-3JC62</p> <p>TWS</p> <p>2002-2023; monthly</p> <p>3°/200 km</p> <p>cm</p> <p>Terrestrial Water Storage (TWS) observed by the Gravity Recovery and Climate Experiment (GRACE) and its follow up satellites. The satellites measure anomalies in Earth's gravitational field representing changes in TWS in groundwater, soil moisture, surface water, canopy water, ice and snow storages. We used the JPL RL06 Mascon solution that filters out noise from the observations and separates land observations from the ocean</p> <p>Mean of all grid cell values within each hotspot region. Subtraction of long-term average per month for seasonality removal. Linear interpolation between each monthly observation to fill data gaps. Quantile normalization to harmonize time series magnitude and distribution</p>
Input variables	
Population count	<p>WorldPop</p> <p>10.5258/SOTON/WP00647</p> <p>pop</p> <p>2000-2020; yearly</p> <p>30 arcsec/1 km</p> <p>Population count (–)</p> <p>Estimated total number of people per grid-cell. For the compilation of the data set a top-down modelling methodology is used: a global database of population count per administrative unit is disaggregated using more detailed geospatial data sets</p> <p>Sum of all grid cell values within each hotspot region. Linear interpolation between each yearly observation to obtain a monthly resolution. Quantile normalization to harmonize time series magnitude and distribution</p>
Temperature	<p>ERA5; Hersbach et al. (2023)</p> <p>10.24381/cds.f17050d7</p> <p>t2m</p> <p>1950-2022; monthly</p> <p>0.5°/50 km</p> <p>°K</p> <p>ECMWF reanalysis of air temperature 2 m above the surface. The data set is compiled from a combination of model data with observations on global scale</p> <p>Mean temperature for all grid cell values within each hotspot region. Subtraction of long-term monthly average for seasonality removal. Quantile normalization to harmonize time series magnitude and distribution</p>
Precipitation	<p>ERA5; Hersbach et al. (2023)</p> <p>10.24381/cds.f17050d7</p> <p>pr</p> <p>1950-2022; monthly</p> <p>0.5°/50 km</p> <p>m</p> <p>ECMWF reanalysis of total precipitation, including rain and snow. The total precipitation is accumulated over a daily period. The unit in meters is the depth of water assuming all water spreads equally over a grid cell. The data set is compiled from a combination of model data with observations on global scale</p>

Table A1 <i>Continued</i>	
	Sum of all grid cell values within each hotspot region. Subtraction of long-term monthly average for seasonality removal. Quantile normalization to harmonize time series magnitude and distribution
Discharge	<p>PCR GLOBWB 2; Sutanudjaja et al. (2018)</p> <p>10.5194/gmd-11-2429-2018</p> <p>Q</p> <p>1980-2019; monthly</p> <p>5 arcmin/10 km</p> <p>m³/s</p> <p>Modelled discharge estimated through routing methods along the river network. The modelled discharge is in fair agreement with GRDC discharge observation point data</p> <p>Mean discharge for all grid cell values within each hotspot region. Subtraction of long-term monthly average for seasonality removal. Quantile normalization to harmonize time series magnitude and distribution</p>
EVI	<p>MODIS MYD13A2 EVI; Didan. (2015)</p> <p>10.5067/MODIS/MYD13A2.006</p> <p>EVI</p> <p>2000-2022; monthly</p> <p>0.5°/50 km</p> <p>–</p> <p>The Enhanced Vegetation Index (EVI) is a metric for vegetation greenness. It minimizes canopy-soil variations and performs well in conditions with high vegetation density. Index values range between –0.2 (no greenness) and 10 (high greenness)</p> <p>Clipping over irrigated area (see additional data). For years 2016–2019 irrigated area is missing and the static area observed in 2015 is assumed for those years. Mean EVI for all grid cell values within the irrigated area of a hotspot. Subtraction of long-term monthly average for seasonality removal. Quantile normalization to harmonize time series magnitude and distribution</p>
Additional variable	
Irrigated area	<p>Global irrigation areas; Nagaraj et al. (2021)</p> <p>https://zenodo.org/record/4659476</p> <p>–</p> <p>2001-2015; yearly</p> <p>5 arcmin/10 km</p> <p>0: No irrigation, 1: Low to medium irrigation, 2: High irrigation</p> <p>Global maps of irrigated area. The maps are predicted from machine learning models that are trained on climate, soil, vegetation and land cover data</p> <p>Clipping according to hotspot area. Subsetting cells with value larger than 1 to obtain irrigated areas of both low and high intensity</p>

Appendix B: Baseline SCM Intercepts, Coefficients, and Residuals Statistics

Table B1

Values of Intercepts β^j for the Baseline Structural Causal Models per Hotspot Estimated by the NLS Minimization. The Values in Superscript (j) Refer to the Corresponding Variable (0: TWS, 1: t2m, 4: EVI, 5: Q)

Hotspot	β^0	β^1	β^4	β^5
Arabian Peninsula	0.16	1.40	-2.44	0.91
California	0.00‡	0.70	0.79	0.56
Central Chile	-0.21	-1.26	0.69	0.43
Coastal Peru	-0.57	-0.53	-1.54	3.24
Ganges	0.21	0.27	0.38	0.27
Greece	-0.11‡	0.30	-0.29	0.62
Indus	0.05‡	0.44	0.41	0.40
Iran	0.13	0.75	0.41	0.69
Italy	-0.09‡	0.04‡	0.22	0.23
Japan	0.33	0.54	0.93	0.72
Java	-0.15	0.47	0.17	0.52
Mexico	2.01	-1.11	1.28	1.09
Murray-Darling	-0.92	-1.76	2.30	1.61
Nile Delta	-0.01‡	0.25	0.25	0.30
North China Plain	0.15	0.36	1.32	1.10
Spain	-0.16	0.24	0.70	0.52
Thailand	-0.03‡	0.17†	0.07‡	0.16
Türkiye	-0.22	0.27	0.10‡	0.31
U.S. High Plains	1.15	-1.48	0.28	0.50
Vietnam	-0.36	0.36	0.62	0.62
White Nile	-0.07‡	0.38	0.25	0.38

Note. All values are significant at the 1% level except values with the †, which refers to a significance at the 5% level, and the ‡ that refers to insignificance.

Table B2
Values of Coefficient $\alpha_c^{j,k}$ for the Baseline Structural Causal Models Per Hotspot Estimated by the NLS Minimization. The Values in Superscript (j, k) Refer to the Corresponding Variable (0: TWS, 1: 12m, 2: tp, 3: Pop, 4: EVI, 5: Q), the Value in Subscript (τ) Refers to the Corresponding Time Lag

Hotspot	α_1^0	α_0^0	α_1^1	$\alpha_0^{1,2}$	α_0^2	α_1^2	α_0^3	α_0^4	α_1^4	$\alpha_0^{4,1}$	$\alpha_0^{4,2}$	$\alpha_1^{4,2}$	$\alpha_0^{4,3}$	α_0^5	α_1^5	$\alpha_0^{5,1}$	$\alpha_0^{5,2}$	$\alpha_0^{5,3}$
Arabian Peninsula	0.46	-0.15	0.02‡	3.59	0.82	0.21	-0.52	-0.23	0.04‡	0.38	0.54	0.97	0.19‡	0.03‡	1.47	1.18	0.67	1.15
California	0.76	0.01‡	0.51	1.04	0.10‡	0.16	-0.36	0.00‡	0.83	-0.25	1.13	0.77	0.96	0.18	0.42	-0.51	0.55	1.20
Central Chile	0.85	0.34	0.00‡	0.29	-0.60	-0.24	-0.15	0.59	-0.10	0.31	0.41	0.57	0.01‡	0.33	0.26	-0.49	1.20	0.68
Coastal Peru	0.42	-1.10	0.04‡	2.10	2.03	-0.08	0.18	0.07	1.15	2.64	-3.05	2.28	0.65	0.06	0.90	1.34	1.30	2.38
Ganges	0.83	-0.68	0.04‡	0.87	0.49	0.20	-0.09‡	0.09	0.53	0.62	-0.38	-1.41	-0.66	0.06	-0.80	-0.84	4.29	-0.64
Greece	0.65	-0.06‡	0.54	1.07	0.31	0.23	0.17	0.08	-0.07‡	-0.33‡	-0.35	0.59	-0.43	0.08	0.79	-0.54	0.95	-0.24‡
Indus	0.65	-0.22	0.05‡	0.30	-0.30	-0.05	-0.47	0.23	0.23	0.79	1.72	0.74	0.57	0.05	1.14	-0.03‡	1.40	-0.29
Iran	0.79	0.24	0.15‡	1.27	0.16	0.17	-0.06‡	-0.35	0.14‡	0.33	0.62	0.37	0.35	0.06	0.70	-0.15	-0.51	0.32
Italy	0.72	0.51	-0.20	0.34	-0.16	-0.03‡	-0.26	0.28	-0.01‡	-0.59	0.23‡	0.58	1.47	0.22	0.15	-0.25	0.81	-1.04
Japan	0.87	0.56	0.39	1.13	-0.53	0.29	0.26	-0.45	0.20	0.86	0.10‡	0.93	0.85	-0.32	0.35	-1.71	0.56	-0.24‡
Java	0.66	-0.68	0.48	-0.64	-0.15‡	-1.46	-0.36	0.86	0.04‡	0.76	0.23‡	1.80	0.69	0.01‡	0.94	-0.68	-0.15‡	-0.79
Mexico	0.84	1.69	0.00‡	0.54	-0.68	0.13	0.01‡	-0.03‡	1.11	-1.58	0.77	1.47	0.66	-0.06	-0.77	0.51	0.48	1.77
Murray-Darling	0.94	-0.52	0.01‡	4.28	2.43	0.13	0.06‡	-0.02‡	-0.21	-0.91	2.42	3.77	1.51	-0.01‡	1.55	1.87	1.18	1.04
Nile Delta	0.77	0.56	-0.08	-0.76	0.47	-0.16	-0.16	0.20	0.07‡	1.53	0.46	0.22	0.25	0.12	-0.43	-0.08‡	0.32	0.27‡
North China Plain	0.85	0.92	0.03‡	2.26	-1.61	0.12	0.12	-0.02‡	1.50	0.87	0.94	0.88	2.21	-0.46	0.08‡	-0.52	1.87	0.34
Spain	0.66	-0.04‡	0.04‡	0.27‡	-0.18	0.12	-0.06‡	0.00‡	0.50	0.54	0.57	0.69	1.33	0.20	1.06	1.40	1.05	-0.11‡
Thailand	0.82	0.20	0.05‡	-1.00	0.24	0.02‡	-0.54	-0.11	0.06‡	-0.06‡	-0.54	-0.18‡	-0.01‡	0.29	0.05‡	0.73	0.72	1.37
Türkiye	0.78	0.32	0.07	-0.23	0.30	0.48	-0.37	0.58	-0.06‡	0.27	-0.24	-0.54	0.38	0.10	0.68	0.12‡	0.83	-0.31
U.S. High Plains	0.89	0.83	0.00‡	-3.51	2.10	0.55	-0.46	0.48	-0.06‡	1.15	2.25	-0.92	0.72	0.27	0.09‡	-1.46	1.60	0.34
Vietnam	0.71	0.33	-0.02‡	-0.63	0.44	0.13	-0.20	0.18	0.51	1.29	0.54	0.25‡	1.88	-0.09	0.87	-0.54	-0.42	0.49
White Nile	0.49	0.00‡	0.29	1.31	-0.03‡	-0.16	0.65	0.14	0.65	-0.38	0.89	1.38	-1.65	0.04‡	0.33	1.11	-0.48	-0.28

Note. All values are significant at the 1% level except values with the †, which refers to a significance at the 5% level, and the ‡ that refers to insignificance.

Table B3

SCM Residuals (η_t) Metrics, With the Mean, Standard Deviation (Stdev), Autocorrelation Value (Lag Corr) of η_t , and the Shapiro-Wilk Test Statistic (Stat) and P-Value. If the Test Statistic is Close to One It Suggests a Normal Distribution and If the P-Value is Larger Than 0.05 It Also Suggests a Normal Distribution

Hotspot	Mean η_t	Stdev η_t	Lag corr η_t	Shapiro	
				Stat	p-value
Arabian Peninsula	0.00	0.05	0.32	0.97	0.00
California	-0.04	0.13	0.63	0.99	0.26
Central Chile	0.05	0.21	0.73	0.98	0.00
Coastal Peru	-0.02	0.19	0.42	0.99	0.19
Ganges	0.00	0.14	0.73	0.99	0.21
Greece	0.00	0.19	0.47	1.00	0.88
Indus	0.01	0.13	0.46	1.00	0.97
Iran	0.00	0.10	0.71	0.98	0.01
Italy	0.00	0.20	0.63	0.99	0.18
Japan	0.04	0.24	0.82	0.98	0.01
Java	0.00	0.16	0.58	0.99	0.28
Mexico	-0.02	0.19	0.76	0.98	0.00
Murray-Darling	-0.01	0.18	0.89	0.97	0.00
Nile Delta	-0.01	0.20	0.67	0.99	0.09
North China Plain	-0.02	0.13	0.74	0.99	0.07
Spain	-0.01	0.17	0.53	0.98	0.03
Thailand	-0.01	0.16	0.76	0.99	0.30
Türkiye	0.02	0.16	0.64	0.99	0.66
U.S. High Plains	0.02	0.15	0.73	0.98	0.00
Vietnam	0.02	0.23	0.64	0.99	0.06
White Nile	0.01	0.14	0.43	0.99	0.45

Data Availability Statement

Data from GRACE and GRACE-FO (Wiese et al., 2019), WorldPop (2018), ERA5 (Hersbach et al., 2023), MODIS (Didan, 2015), PCR GLOBWB 2 (Sutanudjaja et al., 2017) and global irrigated areas (Nagaraj et al., 2021) were used in creation of this manuscript. Numpy (v1.23.5), pandas (v2.1.1), matplotlib (v3.8.0), scipy (v1.11.3), scikit-learn (v1.3.1) and tigramite (v5.2.3.1) were used to process data and create figures. Processed data and code used to generate figures are available in Leijnse et al. (2025).

Acknowledgments

We would like to thank Bram Droppers for proofreading our manuscript. This research was supported by the National Geographic Society, as part of the World Water Map and Freshwater Initiative.

References

- Awais, M., Vinca, A., Byers, E., Frank, S., Fricko, O., Boere, E., et al. (2024). MESSAGEix-GLOBIOM nexus module: Integrating water sector and climate impacts. *Geoscientific Model Development*, 17(6), 2447–2469. <https://doi.org/10.5194/gmd-17-2447-2024>
- Bierkens, M. F. P. (2015). Global hydrology 2015: State, trends, and directions. *Water Resources Research*, 51(7), 4923–4947. <https://doi.org/10.1002/2015WR017173>
- Bijl, D. L., Bogaart, P. W., Kram, T., de Vries, B. J., & van Vuuren, D. P. (2016). Long-term water demand for electricity, industry and households. *Environmental Science & Policy*, 55, 75–86. <https://doi.org/10.1016/j.envsci.2015.09.005>
- Bonetti, P., Metelli, A. M., & Restelli, M. (2024). Causal feature selection via transfer entropy. In *2024 international joint conference on neural networks (ijcnn)* (pp. 1–10). Institute of Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/ijcnn60899.2024.10651028>
- Camps-Valls, G., Gerhardus, A., Ninad, U., Varando, G., Martius, G., Balaguer-Ballester, E., et al. (2023). Discovering causal relations and equations from data. *Physics Reports*, 1044(1044), 1–68. <https://doi.org/10.1016/j.physrep.2023.10.005>
- Cárdenas Belleza, G. A., Bierkens, M. F., & van Vliet, M. T. (2023). Sectoral water use responses to droughts and heatwaves: Analyses from local to global scales for 1990–2019. *Environmental Research Letters*, 18(10), 104008. <https://doi.org/10.1088/1748-9326/acf82e>
- Chávez García Silva, R., Reinecke, R., Coptly, N. K., Barry, D. A., Heggy, E., Labat, D., et al. (2024). Multi-decadal groundwater observations reveal surprisingly stable levels in southwestern Europe. *Communications Earth & Environment*, 5(1), 387. <https://doi.org/10.1038/s43247-024-01554-w>

- Davies, E. G., & Simonovic, S. P. (2011). Global water resources modeling with an integrated model of the social-economic-environmental system. *Advances in Water Resources*, 34(6), 684–700. <https://doi.org/10.1016/j.advwatres.2011.02.010>
- Deangelis, A., Dominguez, F., Fan, Y., Robock, A., Kustu, M. D., & Robinson, D. (2010). Evidence of enhanced precipitation due to irrigation over the Great Plains of the United States. *Journal of Geophysical Research*, 115(15), D15115. <https://doi.org/10.1029/2010JD013892>
- de Graaf, I. E., Gleeson, T., van Rens Beek, L. P., Sutanudjaja, E. H., & Bierkens, M. F. (2019). Environmental flow limits to global groundwater pumping. *Nature*, 574(7776), 90–94. <https://doi.org/10.1038/s41586-019-1594-4>
- Desbureaux, S., Mortier, F., Zaveri, E., van Vliet, M. T., Russ, J., Rodella, A. S., & Damania, R. (2022). Mapping global hotspots and trends of water quality (1992-2010): A data driven approach. *Environmental Research Letters*, 17(11), 114048. <https://doi.org/10.1088/1748-9326/ac9cf6>
- Didan, K. (2015). MYD13A2 MODIS/aqua vegetation Indices 16-Day L3 global 1km SIN grid V006 [Dataset]. *NASA EOSDIS Land Processes Distributed Active Archive Center*. <https://doi.org/10.5067/MODIS/MYD13A2.006>
- Dinar, A., Tieu, A., & Huynh, H. (2019). Water scarcity impacts on global food production. *Global Food Security*, 23, 212–226. <https://doi.org/10.1016/j.gfs.2019.07.007>
- Dolan, F., Lamontagne, J., Link, R., Hejazi, M., Reed, P., & Edmonds, J. (2021). Evaluating the economic impact of water scarcity in a changing world. *Nature Communications*, 12(1), 1915. <https://doi.org/10.1038/s41467-021-22194-0>
- Döll, P., Douville, H., Güntner, A., Müller Schmied, H., & Wada, Y. (2016). Modelling freshwater resources at the global scale: Challenges and Prospects. *Surveys in Geophysics*, 37(2), 195–221. <https://doi.org/10.1007/s10712-015-9343-1>
- Felfelani, F., Wada, Y., Longuevergne, L., & Pokhrel, Y. N. (2017). Natural and human-induced terrestrial water storage change: A global analysis using hydrological models and GRACE. *Journal of Hydrology*, 553, 105–118. <https://doi.org/10.1016/j.jhydrol.2017.07.048>
- Frey, H. C., & Patil, S. R. (2002). Identification and review of sensitivity analysis methods. *Risk Analysis*, 22(3), 553–578. <https://doi.org/10.1111/0272-4332.00039>
- Gerten, D., Rost, S., von Bloh, W., & Lucht, W. (2008). Causes of change in 20th century global river discharge. *Geophysical Research Letters*, 35(20), L20405. <https://doi.org/10.1029/2008GL035258>
- Greve, P., Kahil, T., Mochizuki, J., Schinko, T., Satoh, Y., Burek, P., et al. (2018). Global assessment of water challenges under uncertainty in water scarcity projections. *Nature Sustainability*, 1(9), 486–494. <https://doi.org/10.1038/S41893-018-0134-9>
- Günther, W., Ninad, U., & Runge, J. (2023). Causal Discovery for time series from multiple datasets with latent contexts. In *Proceedings of machine learning research* (pp. 766–776).
- Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., et al. (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences*, 111(9), 3251–3256. <https://doi.org/10.1073/pnas.1222475110>
- Hanasaki, N., Yoshikawa, S., Pokhrel, Y., & Kanae, S. (2018). A global hydrological simulation to specify the sources of water used by humans. *Hydrology and Earth System Sciences*, 22(1), 789–817. <https://doi.org/10.5194/hess-22-789-2018>
- Hejazi, M., Edmonds, J., Clarke, L., Kyle, P., Davies, E., Chaturvedi, V., et al. (2014). Long-term global water projections using six socio-economic scenarios in an integrated assessment modeling framework. *Technological Forecasting and Social Change*, 81(1), 205–226. <https://doi.org/10.1016/j.techfore.2013.05.006>
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023). ERA5 monthly averaged data on single levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. <https://doi.org/10.24381/cds.f17050d7>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Hoekstra, A. Y., & Chapagain, A. K. (2007). Water footprints of nations: Water use by people as a function of their consumption pattern. *Water Resources Management*, 21(1), 35–48. <https://doi.org/10.1007/s11269-006-9039-x>
- Huang, Z., Hejazi, M., Li, X., Tang, Q., Vernon, C., Leng, G., et al. (2018). Reconstruction of global gridded monthly sectoral water withdrawals for 1971–2010 and analysis of their spatiotemporal patterns. *Hydrology and Earth System Sciences*, 22(4), 2117–2133. <https://doi.org/10.5194/hess-22-2117-2018>
- Huo, F., Xu, L., Li, Z., Li, Y., Famiglietti, J. S., & Chandanpurkar, H. A. (2024). Can climate change signals be detected from the terrestrial water storage at daily timescale? *npj Climate and Atmospheric Science*, 7(1), 158. <https://doi.org/10.1038/s41612-024-00646-w>
- Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsudduha, M., Taylor, R. G., et al. (2024). Rapid groundwater decline and some cases of recovery in aquifers globally. *Nature*, 625(7996), 715–721. <https://doi.org/10.1038/s41586-023-06879-8>
- Jing, W., Yao, L., Zhao, X., Zhang, P., Liu, Y., Xia, X., et al. (2019). Understanding terrestrial water storage declining trends in the yellow river basin. *Journal of Geophysical Research: Atmospheres*, 124(23), 12963–12984. <https://doi.org/10.1029/2019JD031432>
- Jones, E. R., Graham, D. J., van Griensven, A., Flörke, M., & van Vliet, M. T. (2024). Blind spots in global water quality monitoring. *Environmental Research Letters*, 19(9), 091001. <https://doi.org/10.1088/1748-9326/ad6919>
- Kim, S. H., Hejazi, M., Liu, L., Calvin, K., Clarke, L., Edmonds, J., et al. (2016). Balancing global water availability and use at basin scale in an integrated assessment model. *Climatic Change*, 136(2), 217–231. <https://doi.org/10.1007/s10584-016-1604-6>
- König, G., Molnar, C., Bischl, B., & Grosse-Wentrup, M. (2020). Relative feature importance. In *Proceedings - international conference on pattern recognition* (pp. 623–630). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICPR48806.2021.9413090>
- Kummu, M., Guillaume, J. H., De Moel, H., Eisner, S., Flörke, M., Porkka, M., et al. (2016). The world's road to water scarcity: Shortage and stress in the 20th century and pathways towards sustainability. *Scientific Reports*, 6(1), 38495. <https://doi.org/10.1038/srep38495>
- Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., et al. (2020). Extending the global Mass change data record: GRACE follow-on Instrument and science data performance. *Geophysical Research Letters*, 47(12), e2020GL088306. <https://doi.org/10.1029/2020GL088306>
- Leijnse, M., F P Bierkens, M., H M Gommans, K., Lin, D., Tait, A., & Wanders, N. (2024). Key drivers and pressures of global water scarcity hotspots. *Environmental Research Letters*, 19(5), 054035. <https://doi.org/10.1088/1748-9326/ad3c54>
- Leijnse, M., F P Bierkens, M., & Wanders, N. (2025). Causality water scarcity hotspots. [Software]. *Zenodo*. <https://doi.org/10.5281/zenodo.15050439>
- Liu, J., Yang, H., Gosling, S. N., Kummu, M., Flörke, M., Pfister, S., et al. (2017). Water scarcity assessments in the past, present, and future. *Earth's Future*, 5(6), 545–559. <https://doi.org/10.1002/2016EF000518>
- Mekonnen, M. M., & Hoekstra, A. Y. (2016). Sustainability: Four billion people facing severe water scarcity. *Science Advances*, 2(2), e1500323. <https://doi.org/10.1126/sciadv.1500323>
- Moré, J. J. (2006). The Levenberg-marquardt algorithm: Implementation and theory. In *Numerical analysis: Proceedings of the biennial conference held at dundee* (pp. 105–116). Springer.
- Nagaraj, D., Proust, E., Todeschini, A., Rulli, M. C., & Odorico, P. D. (2021). A new dataset of global irrigation areas from 2001 to 2015 [Dataset]. *Zenodo*, 152, 103910. <https://doi.org/10.1016/j.advwatres.2021.103910>

- Nazemi, A., & Wheeler, H. S. (2015). On inclusion of water resource management in Earth system models-Part 2: Representation of water supply and allocation and opportunities for improved modeling. *Hydrology and Earth System Sciences*, 19(1), 63–90. <https://doi.org/10.5194/hess-19-63-2015>
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., et al. (2020). What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57(3), e2020WR028091. <https://doi.org/10.1029/2020WR028091>
- Nichol, J. J., Weylandt, M., Smith, M., & Swiler, L. P. (2023). Benchmarking the PCMC1 causal discovery algorithm for spatiotemporal systems. (*Tech. Rep.*). Retrieved from <https://classic.ntis.gov/help/order-methods>
- Ou, Z., He, F., Zhu, Y., Lu, P., & Wang, L. (2023). Analysis of driving factors of water demand based on explainable artificial intelligence. *Journal of Hydrology: Regional Studies*, 47, 101396. <https://doi.org/10.1016/j.ejrh.2023.101396>
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.
- Penny, G., Mondal, M. S., Biswas, S., Bolster, D., Tank, J. L., & Müller, M. F. (2020). Using natural experiments and counterfactuals for causal assessment: River salinity and the Ganges water agreement. *Water Resources Research*, 56(4), e2019WR026166. <https://doi.org/10.1029/2019WR026166>
- Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms*. The MIT press.
- Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., et al. (2021). Global terrestrial water storage and drought severity under climate change. *Nature Climate Change*, 11(3), 226–233. <https://doi.org/10.1038/s41558-020-00972-w>
- Pokhrel, Y. N., Hanasaki, N., Wada, Y., & Kim, H. (2016). Recent progresses in incorporating human land–water management into global land surface models toward their integration into Earth system models. *Wiley Interdisciplinary Reviews: Water*, 3(4), 548–574. <https://doi.org/10.1002/wat2.1150>
- Porkka, M., Gerten, D., Schaphoff, S., Siebert, S., & Kumm, M. (2016). Causes and trends of water scarcity in food production. *Environmental Research Letters*, 11(1), 015001. <https://doi.org/10.1088/1748-9326/11/1/015001>
- Porkka, M., Virkki, V., Wang-Erlandsson, L., Gerten, D., Gleeson, T., Mohan, C., et al. (2024). Notable shifts beyond pre-industrial streamflow and soil moisture conditions transgress the planetary boundary for freshwater change. *Nature Water*, 2(3), 262–273. <https://doi.org/10.1038/s44221-024-00208-7>
- Qin, Y., Mueller, N. D., Siebert, S., Jackson, R. B., AghaKouchak, A., Zimmerman, J. B., et al. (2019). Flexibility and intensity of global water use. *Nature Sustainability*, 2(6), 515–523. <https://doi.org/10.1038/s41893-019-0294-2>
- Richey, A. S., Thomas, B. F., Lo, M.-H., Reager, J. T., Famiglietti, J. S., Voss, K., et al. (2015). Quantifying renewable groundwater stress with GRACE. *Water Resources Research*, 51(7), 5217–5238. <https://doi.org/10.1002/2015WR017349>
- Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoin, H. K., Landerer, F. W., & Lo, M.-H. (2018). Emerging trends in global freshwater availability. *Nature*, 557(7707), 651–659. <https://doi.org/10.1038/s41586-018-0123-1>
- Rosa, L., Chiarelli, D. D., Tu, C., Rulli, M. C., & D'Odorico, P. (2019). Global unsustainable virtual water flows in agricultural trade. *Environmental Research Letters*, 14(11), 114001. <https://doi.org/10.1088/1748-9326/ab4bfc>
- Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., et al. (2019). Inferring causation from time series in Earth system sciences. *Nature Communications*, 10(1), 2553. <https://doi.org/10.1038/s41467-019-10105-3>
- Runge, J., Gerhardus, A., Varando, G., Eyring, V., & Camps-Valls, G. (2023). Causal inference for time series. *Nature Reviews Earth and Environment*, 4(7), 1–13. <https://doi.org/10.1038/s43017-023-00431-y>
- Schwalm, C. R., Anderegg, W. R., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., et al. (2017). Global patterns of drought recovery. *Nature*, 548(7666), 202–205. <https://doi.org/10.1038/nature23021>
- Shapiro, S. S., & Wilk, M. B. (1965). Biometrika trust an analysis of variance test for normality (complete samples). *Biometrika*, 52(3), 591–611. <https://doi.org/10.2307/2333709>
- Song, Y., Wei, J., Cheng, H., & Zan, B. (2024). Irrigation in the North China plain regulates the diurnal cycle of precipitation and regional water cycle. *Climate Dynamics*. <https://doi.org/10.1007/s00382-024-07208-z>
- Sun, J., & Bollt, E. M. (2014). Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings. *Physica D: Nonlinear Phenomena*, 267, 49–57. <https://doi.org/10.1016/j.physd.2013.07.001>
- Sutanudjaja, E., van Beek, R., Wanders, N., Wada, Y., C Bosmans, J. H., Drost, N., et al. (2017). PCR-GLOBWB model: PCR-GLOBWB version v2.1.0 beta 1. [Software]. *Zenodo*. <https://doi.org/10.5281/zenodo.247139>
- Sutanudjaja, E., van Beek, R., Wanders, N., Wada, Y., C Bosmans, J. H., Drost, N., et al. (2018). PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, 11(6), 2429–2453. <https://doi.org/10.5194/gmd-11-2429-2018>
- Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, 305(5683), 503–505. <https://doi.org/10.1126/science.1099192>
- Theeuwens, J. J., Staal, A., Tuinenburg, O. A., Hamelers, B. V., & Dekker, S. C. (2023). Local moisture recycling across the globe. *Hydrology and Earth System Sciences*, 27(7), 1457–1476. <https://doi.org/10.5194/hess-27-1457-2023>
- UNEP. (2016). A Snapshot of the World's water quality: Towards a global assessment (Tech. Rep.). Retrieved from www.unep.org
- Van Vliet, M. T. H., Florke, M., & Wada, Y. (2017). Quality matters for water scarcity. *Nature Geoscience*, 10(11), 800–802. <https://doi.org/10.1038/NGEO3047>
- Van Vliet, M. T. H., Jones, E. R., Flörke, M., Franssen, W. H. P., Hanasaki, N., Wada, Y., & Yearsly, J. R. (2021). Global water scarcity including surface water quality and expansions of clean water technologies. *Environmental Research Letters*, 16(2), 024020. <https://doi.org/10.1088/1748-9326/abbfc3>
- van Vuuren, D. P., Kok, M., Lucas, P. L., Prins, A. G., Alkemade, R., van den Berg, M., et al. (2015). Pathways to achieve a set of ambitious global sustainability objectives by 2050: Explorations using the IMAGE integrated assessment model. *Technological Forecasting and Social Change*, 98, 303–323. <https://doi.org/10.1016/j.techfore.2015.03.005>
- Veldkamp, T. I., Zhao, F., Ward, P. J., De Moel, H., Aerts, J. C., Schmied, H. M., et al. (2018). Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: A multi-model validation study. *Environmental Research Letters*, 13(5), 055008. <https://doi.org/10.1088/1748-9326/aab96f>
- Wada, Y., Van Beek, L. P. H., & Bierkens, M. F. P. (2011). Modelling global water stress of the recent past: On the relative importance of trends in water demand and climate variability. *Hydrology and Earth System Sciences*, 15(12), 3785–3808. <https://doi.org/10.5194/hess-15-3785-2011>
- Wiese, D., Yuan, D.-N., Boening, C., Landerer, F., & Watkins, M. (2019). JPL GRACE and GRACE-FO mascon ocean, ice, and hydrology equivalent water height CRI filtered [Dataset]. *PO.DAAC*. <https://doi.org/10.5067/TEMSC-3JC62>
- WorldPop. (2018). Global high resolution population denominators project [Dataset]. *Funded by The Bill Melinda Gates Foundation*. <https://doi.org/10.5258/SOTON/WP00647>
- Xie, X., He, B., Guo, L., Miao, C., & Zhang, Y. (2019). Detecting hotspots of interactions between vegetation greenness and terrestrial water storage using satellite observations. *Remote Sensing of Environment*, 231, 111259. <https://doi.org/10.1016/j.rse.2019.111259>

- Yang, Y., Jin, Z., Mueller, N. D., Driscoll, A. W., Hernandez, R. R., Grodsky, S. M., et al. (2023). Sustainable irrigation and climate feedbacks. *Nature Food*, 4(8), 654–663. <https://doi.org/10.1038/s43016-023-00821-x>
- Yang, Y., Roderick, M. L., Guo, H., Miralles, D. G., Zhang, L., Fatichi, S., et al. (2023). Evapotranspiration on a greening Earth. *Nature Reviews Earth & Environment*, 4(9), 626–641. <https://doi.org/10.1038/s43017-023-00464-3>
- Zampieri, M., Luong, T. M., Ashok, K., Dasari, H. P., Pistocchi, A., & Hoteit, I. (2024). Leveraging atmospheric moisture recycling in Saudi Arabia and neighboring countries for irrigation and afforestation planning. *Regional Environmental Change*, 24(3), 124. <https://doi.org/10.1007/s10113-024-02284-7>