https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

Novel analogue-based and geostatistical approaches for space-time prediction of environmental variables

Loïc Gerber[1], Mathieu Gravey[2], and Grégoire Mariéthoz[1]

[1]Faculty of Geosciences and Environment, University of Lausanne, Switzerland
[2]Institute for Interdisciplinary Mountain Research, Austrian Academy of Sciences, Innsbruck, Austria

Correspondence: Loïc Gerber (loic.gerber.2@unil.ch)

Abstract. In many environmental applications, it is important to obtain accurate and coherent reconstructions of environmental maps. These can be for example spatial representations of hydrological variables or satellite-based remote sensing products. To do this, analogue-based methods provide a fast and interpretable approach that is based on replicating patterns coming from historical observations and predictor similarity. However, their classical weighted-average aggregation tends to smooth spatial variability and attenuate extremes. This study evaluates whether analogue-based reconstruction of MODIS evapotranspiration (ET) over the Ebro watershed can be improved through local analogue selection and stochastic pattern-based aggregation. Several strategies are compared, including domain-wise and tile-wise analogue selection, inverse-distance weighted averaging, and two Multiple-Point Statistics simulation techniques, *chessQS* and a new method named Anchor Sampling. Weighted averaging provides the best pixel-wise accuracy, image-structure agreement, and computational efficiency, while tile-wise analogue selection yields only marginal improvements over domain-wise selection. Anchor Sampling offers the most balanced stochastic option, improving variogram agreement, timestep water-balance error, and tail-value statistics while remaining closer to the deterministic baseline than *chessQS*. In contrast, *chessQS* increases spatial flexibility but reduces reconstruction accuracy and image-structure agreement in the present application. The results indicate that stochastic aggregation can help preserve aspects of spatial variability and distributions, but that simple analogue averaging remains a strong baseline for ET map reconstruction, unless the realism of the spatial structure or uncertainty representation are specifically required.

1 Introduction

Spatio-temporal reconstruction is central to the use of environmental datasets, particularly those derived from remote sensing. Many applications, such as fully distributed hydrological modelling, require spatially and temporally continuous fields to represent land-atmosphere interactions and surface fluxes across heterogeneous landscapes (AghaKouchak et al., 2015; Hrachowitz and Clark, 2017; De Lannoy et al., 2022). However, remote sensing products are frequently affected by missing values due to cloud cover, sensor limitations, or irregular acquisition conditions (Wang et al., 2023, 2022; Siabi et al., 2022). Beyond filling existing gaps, reconstruction methods can also support the temporal extension of environmental map archives. Producing complete and realistic spatio-temporal fields therefore remains a key methodological challenge. In this study, we address this

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

broader reconstruction problem through a controlled cross-validation setting, in which complete observed maps are withheld and reconstructed from historical analogues and predictor information.

A wide range of approaches has been developed to address this problem, including temporal, spatial, geostatistical, and machine learning methods (Mo et al., 2023; Wang et al., 2023; Wu et al., 2024). While often effective, simpler methods rely on assumptions such as temporal continuity, spatial stationarity, or the availability of informative covariates. Geostatistical approaches such as kriging provide optimal linear estimates under specified covariance assumptions, but are not designed to reproduce complex structures, particularly in heterogeneous landscapes or over large domains (Goovaerts, 1997; Chilès and Delfiner, 2012; Zhang et al., 2022; Chinasho et al., 2021). Machine learning and deep learning models capture non-linear relationships but typically require large datasets, are usually expensive to train, and may produce spatial patterns that are physically inconsistent or difficult to interpret (Wang and Zhang, 2024; Zhang et al., 2025; Miller et al., 2024). These limitations highlight the need for approaches that better preserve realistic spatial structures while remaining interpretable and physically plausible.

Analogue-based methods offer a flexible and physically grounded alternative by leveraging historical observations. Originally introduced by Lorenz (1969), they rely on the assumption that similar atmospheric conditions lead to similar system responses. Widely used in climate science for reconstruction, forecasting, and impact studies (Yiou et al., 2013; Yiou, 2014; Yiou and Jézéquel, 2020; Vautard and Yiou, 2009; Noyelle et al., 2023; Yule et al., 2024), analogue approaches have more recently been applied to remote sensing (Zakeri and Mariethoz, 2024) and the reconstruction of evapotranspiration maps (Gerber et al., 2026a). In these frameworks, similarity is quantified in a predictor space, and the closest historical states are used to estimate the target variable. However, classical implementations rely on deterministic aggregation, typically through (weighted) averaging of the best candidates, which tends to smooth spatial variability and attenuate extreme values. Moreover, analogue selection is often performed at the domain scale, implicitly assuming that a single set of situations represents the entire region. This can lead to suboptimal local matches and restrict the reconstruction of spatial configurations that have not been observed as complete historical maps but may arise under new predictor combinations.

These limitations suggest two complementary directions. The first is to refine analogue selection locally, so that similarity is evaluated over regions that are more consistent with local predictor-response relationships. In this study, this is achieved by dividing the domain into overlapping tiles and selecting analogues independently for each tile, using predictor fields restricted to the same spatial support. This tile-wise strategy avoids the assumption that a single historical date must provide the best analogue for the entire domain.

The second direction is to replace deterministic averaging with stochastic, pattern-based reconstruction. Multiple-Point Statistics (MPS) simulation methods directly address this objective by focusing on the reproduction of spatial patterns. Unlike two-point geostatistical methods, MPS learns higher-order spatial structures from training images and generates stochastic realizations that preserve complex structure (Strebelle, 2002; Journel and Zhang, 2006; Mariethoz et al., 2010; Mariethoz and Caers, 2014). This makes MPS well suited for applications requiring spatial realism, while its stochastic nature enables the generation of multiple plausible realizations and provides a basis for uncertainty quantification. MPS has been applied in hydrogeology (Kessler et al., 2013; Pirot et al., 2014), weather generation (Jha et al., 2015; Oriani et al., 2017), and remote

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

sensing reconstruction (Yin et al., 2016; Gravey et al., 2019). Among available algorithms, QuickSampling (QS) offers good performance and improved computational efficiency relative to other more demanding MPS implementations (Gravey and Mariethoz, 2020; Hadjipetrou et al., 2023).

A major limitation of MPS is its computational cost for large or high-dimensional datasets. To address this, strategies have been proposed to reduce the search space of candidate patterns. The concept of Conditional Training Image Sets (CTIS; Oriani et al., 2017) consists in selecting subsets of training images conditioned on auxiliary information, restricting simulations to contextually consistent patterns. This highlights that training image relevance is as critical as the simulation algorithm itself: poorly matched training images may yield realistic but incorrect patterns. Recent developments also improve scalability through tiling strategies that preserve global consistency while enabling local simulations (Gerber et al., 2026b).

Building on these ideas, we propose a framework in which analogue selection defines data-driven CTIS for each query date. Historical maps are selected based on similar meteorological, seasonal, and atmospheric conditions, and are then aggregated to produce the reconstruction. This separation allows us to disentangle two key aspects of performance: selecting relevant historical situations and preserving realistic spatial structures when combining them.

Within this framework, stochastic aggregation is explored through two complementary strategies. The first uses *chessQS* (Gerber et al., 2026b), a tile-based adaptation of QS, where selected analogue maps act as training images for MPS simulation and spatial patterns can be recombined across the domain. The second uses Anchor Sampling (AS), a new spatially constrained stochastic alternative introduced here in which candidate values are sampled from homologous locations across the selected analogue maps while still accounting for local neighbourhood patterns. AS therefore preserves location-specific controls on ET, such as topography, land cover (if consistent over time), river networks, or irrigation structures, while introducing variability through the analogue ensemble.

In this study, the framework is applied to the reconstruction of MODIS evapotranspiration (ET) maps over the Ebro watershed in Spain. Specifically, we investigate whether analogue-based reconstruction can be improved through (i) local analogue selection and (ii) stochastic pattern-based aggregation. We compare domain-wise analogue selection, where a single set of analogues represents the entire domain, with tile-wise selection, where local CTIS are defined independently. Each selection strategy is combined with three aggregation methods: inverse-distance weighted averaging as a deterministic baseline, *chessQS* and Anchor Sampling. Performance is evaluated under cross-validation, in which MODIS ET maps are withheld as reference query dates and reconstructed from predictor information and historical analogues.

The remainder of the paper is structured as follows. The next section presents the study area. Section 3 discusses the datasets used. Section 4 details the methodological framework developed in this study. Section 5 details the calibration and experimental design. Section 6 presents the results obtained. Section 7 offers a discussion of these results and, lastly, Section 8 concludes the study.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

2 Study area

The study area is the Ebro River basin, located in the northeastern Iberian Peninsula in Spain (Fig. 1). With an area of approximately 85,500 km^2, it is the largest hydrographic basin in Spain and covers about 17% of the Spanish peninsular territory (La Confederación Hidrográfica del Ebro, 2026). The basin is bounded by the Cantabrian Mountains and the Pyrenees to the north, the Iberian System to the south and southeast, and the Catalan Coastal Ranges to the east. The Ebro River drains the central depression from northwest to southeast before discharging into the Mediterranean Sea through the Ebro Delta.

The basin provides a suitable test case for ET map reconstruction because it presents strong spatial contrasts in topography, climate, and land-surface controls. Elevations exceed 3000 m in the Pyrenees, whereas extensive lowlands occupy the central Ebro depression. This topographic organization produces a pronounced hydroclimatic gradient: mean annual precipitation ranges from approximately 320 mm yr^{-1} in the semi-arid central valley to more than 2000 mm yr^{-1} in the Pyrenees and Cantabrian Mountains (European Commission, Joint Research Centre, 2021). The northern mountainous tributaries, particularly those draining the Pyrenees, contribute a substantial fraction of the basin discharge (López-Moreno et al., 2011).

In addition to climatic and topographic heterogeneity, the Ebro basin contains extensive irrigated areas, particularly in the central valley, which locally modify surface water availability and ET dynamics. The coexistence of humid mountainous regions, semi-arid lowlands, and irrigated agricultural areas results in spatially heterogeneous ET patterns and predictor-response relationships. These characteristics make the basin particularly relevant for evaluating whether locally weighted analogue selection and stochastic pattern-based aggregation improve the reconstruction of remotely sensed ET fields.

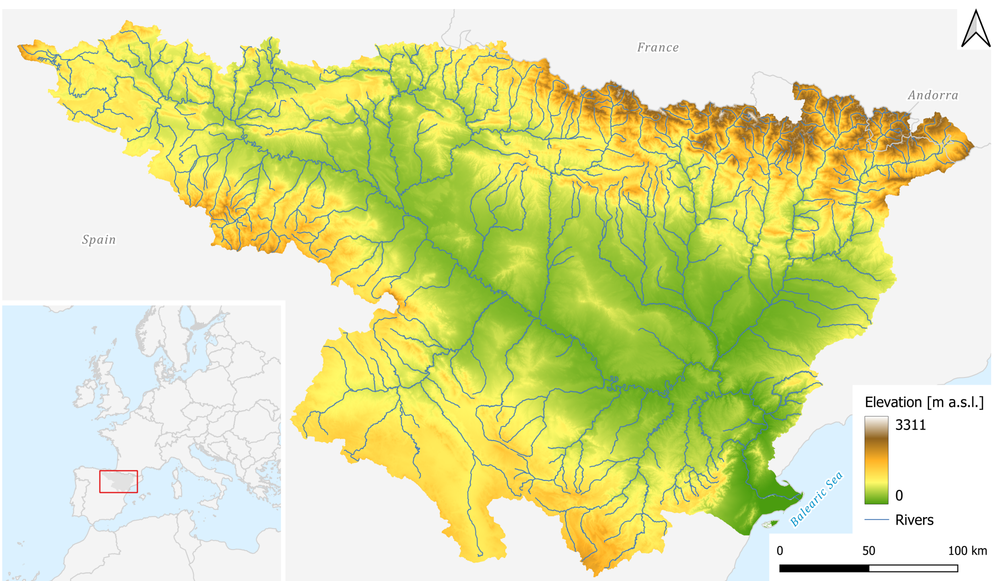

Figure 1. The Ebro watershed.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

3 Data

3.1 MODIS Net Evapotranspiration

The target variable used in this study is evapotranspiration (ET) from the MOD16A2GF product generated from observations of the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra satellite (Running et al., 2021). This dataset provides global estimates of terrestrial evapotranspiration and latent heat flux at a spatial resolution of 500 m with an 8-day temporal compositing period. The MOD16A2GF product is produced using an algorithm based on the Penman-Monteith framework, which integrates remotely sensed MODIS surface variables, such as vegetation dynamics, albedo, and land cover, with daily meteorological reanalysis data (Running et al., 2021).

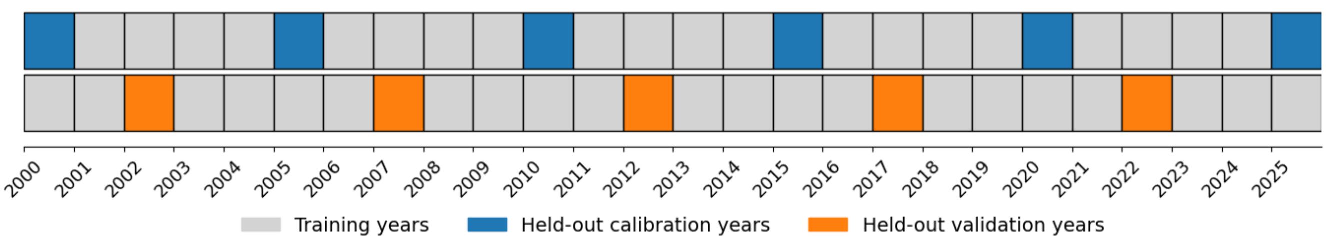

Figure 2. Separation of the training, calibration, and validation periods.

In this study, MOD16A2GF is used as the reference ET archive to be reconstructed. Selected ET maps are withheld as query dates and reconstructed from the remaining historical archive and predictor information. Here, two independent sets of withheld years are used (Fig. 2). The first set uses years 2000, 2005, 2010, 2015, 2020, and 2025 as a calibration dataset to calibrate the method parameters, where all remaining years are used as training dates. The second set uses years 2002, 2007, 2012, 2017, and 2022 as validation years for the different methods, while all other years are used as training data. The dataset is clipped to the Ebro watershed boundaries and, for computational efficiency, the MODIS ET maps are aggregated from their native 500 m resolution to a 1 km target grid using spatial averaging. All reconstruction methods are applied and evaluated on this common target grid.

The processed MODIS ET maps contain a small number of persistent missing pixels that occur at the same locations throughout the archive. These values are not temporally varying observation gaps, but correspond to pixels that are not assigned valid ET estimates in the MOD16A2GF product. These pixels are therefore not reconstructed in this study. They are kept as missing in the target maps, excluded from the analogue archive and validation metrics, and left empty in the reconstructed outputs.

3.2 ERA5-Land precipitation, temperature, wind, and total evaporation

Meteorological predictors describing regional surface conditions are derived from the ERA5-Land reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF; Muñoz-Sabater et al., 2021). The dataset provides global estimates of land-surface variables at a spatial resolution of 0.1° and an hourly temporal resolution.

In this study, daily total precipitation, daily minimum, average, and maximum near-surface air temperature are used as predictors. Daily mean 10 m u- and v-wind components are also used together with wind speed and direction derived from these components. These variables represent key meteorological drivers of evapotranspiration, consistent with Penman-Monteith-type formulations and recent ET-driver analyses (Allen et al., 1998; Fluhrer et al., 2025; Xu et al., 2024). Precipitation controls soil moisture availability, temperature strongly influences evaporative demand and surface energy fluxes, and near-surface wind affects turbulent exchange processes that regulate evapotranspiration. Daily values are obtained by aggregating the hourly ERA5-Land fields.

In addition, total evaporation from ERA5-Land is incorporated as an auxiliary variable. This variable is used as an additional covariate in the analogue selection, providing a physically consistent proxy of MODIS evapotranspiration that complements the meteorological predictors.

3.3 ERA5 geopotential heights

Large-scale atmospheric circulation predictors are taken from ERA5 (Hersbach et al., 2020). ERA5 combines numerical weather prediction models with a wide range of observational data through data assimilation to produce global atmospheric fields at an hourly temporal resolution and a spatial resolution of 0.25°.

Daily mean geopotential fields at 500 and 1000 hPa are used as large-scale atmospheric circulation predictors. Geopotential is converted to geopotential height before being interpolated to the predictor grid. Two spatial supports are prepared for these predictors: a larger regional domain used in the global-weight analogue configuration, and a watershed-shaped domain matching the MODIS ET target grid used in the pixel-wise-weight configuration. The rationale for these two configurations is described in Section 4.2.

3.4 Copernicus DEM

To guide the *chessQS* simulations and constrain the spatial organization of reconstructed patterns, we use the Copernicus GLO-90 Digital Elevation Model (DEM), a Digital Surface Model (European Space Agency and Airbus, 2022), as an auxiliary covariate. The DEM is reprojected and aggregated from its native 90 m resolution to the 1 km MODIS ET grid using area-weighted averaging. It is used exclusively as a covariate in the *chessQS* simulations.

4 Methods

4.1 Analogue-based reconstruction framework

The proposed framework uses analogue selection to condition the reconstruction of evapotranspiration (ET) maps on climatic and meteorological similarity. The archive is split into learning dates (LD), for which both predictors and observed ET maps are available, and query dates (QD), for which ET must be reconstructed. For each ET composite, the analogue descriptor is built from the predictor fields available at the composite reference date and over a short antecedent period. The parameter

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

T_w defines the total number of daily predictor fields included in the descriptor, ending on the composite reference date. The analogue search therefore accounts for the atmospheric and surface conditions on the query date together with short-term antecedent conditions.

The key idea is to separate two operations that are often combined in classical analogue methods. First, a set of historical
165 dates is selected because their predictor states are similar to the query state. This set of analogue maps can be interpreted as a data-driven Conditional Training Image Set (CTIS): it contains historical ET patterns that are plausible under the query conditions. Second, this CTIS must be transformed into a reconstructed ET field. For this second step, we compare three aggregation strategies: inverse-distance weighted averaging, *chessQS* simulation, and Anchor Sampling (AS; Fig. 3).

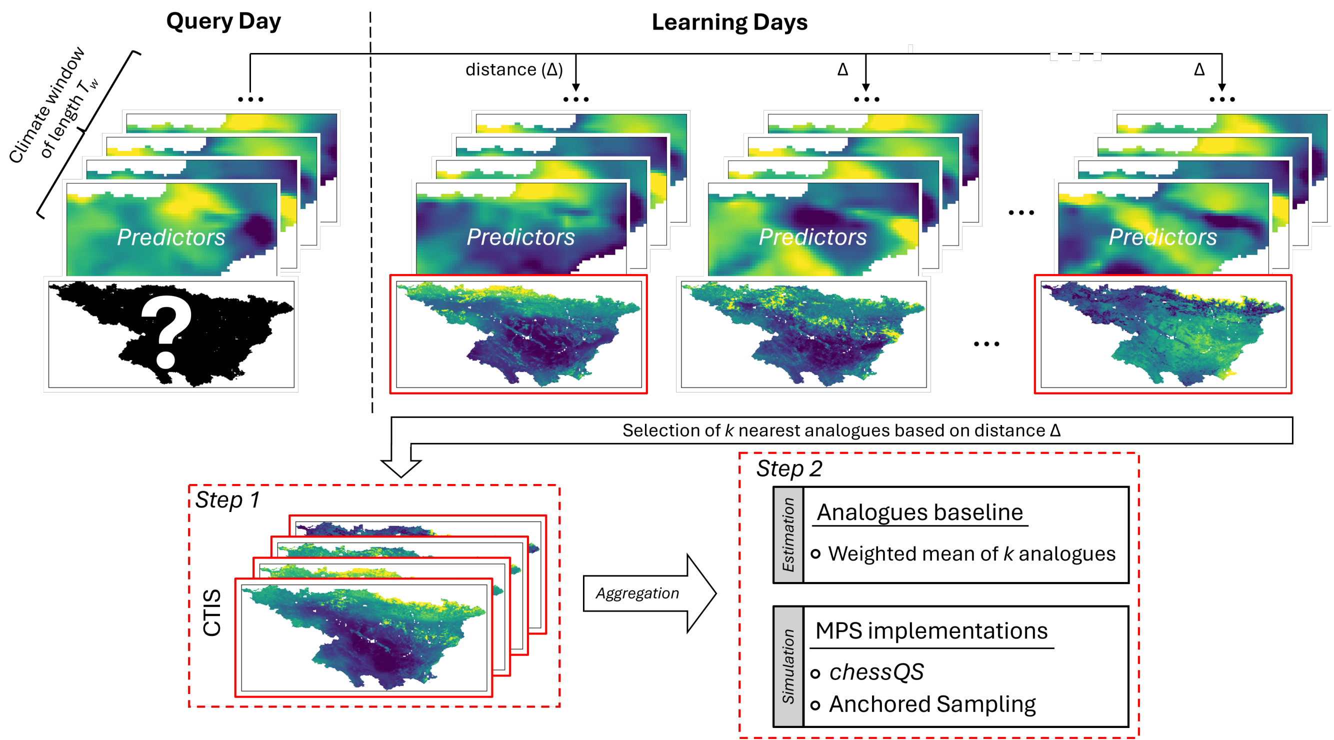

Figure 3. Representation of the analogue selection and aggregation framework. Predictor similarity is used to identify a CTIS of historical ET maps, which is then aggregated either deterministically or stochastically.

4.2 Predictor representation and Random Forest-based weights

170 The assumption underlying the analogue approach is that similar climatic and meteorological conditions lead to similar ET responses. For each query date, similarity to all available learning dates is quantified in predictor space. All predictors are normalised before comparison so that the distance is not dominated by variables with larger numerical ranges. The day of year

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

(DOY) is included as an additional predictor to constrain the analogue search to comparable seasonal conditions. The seasonal cycle is represented by sine and cosine transformations of the day of year (Eq. 1):

$$\mathrm{DOY}_{\sin} = \sin\left(\frac{2\pi \cdot \mathrm{DOY}}{365}\right), \qquad \mathrm{DOY}_{\cos} = \cos\left(\frac{2\pi \cdot \mathrm{DOY}}{365}\right), \tag{1}$$

which avoids artificial discontinuities between the end and beginning of the year.

Because the Ebro basin is spatially heterogeneous, the predictors are not expected to have the same relevance everywhere. Temperature, precipitation, wind, atmospheric circulation, and seasonal timing may contribute differently depending on elevation, land cover, or local water availability. We therefore test two RF-based predictor-weighting strategies. Both strategies use Random Forest regression and permutation importance (Breiman, 2001), implemented with the `RandomForestRegressor` class from `scikit-learn` (Pedregosa et al., 2011).

The first strategy uses global predictor weights. In this case, a single RF model is trained using predictor values over the larger spatial support used for the domain-wise predictor representation. Permutation importance is then used to derive one importance weight per predictor variable. These weights are spatially constant and therefore represent the average relevance of each predictor over the full support.

The second strategy uses pixel-level predictor weights. In this case, one RF model is trained independently at each target pixel using the local predictor time series and the corresponding MODIS ET values. Permutation importance is then used to derive pixel-specific weights for each predictor variable. This produces weights $w_f(\mathbf{s})$ that vary spatially and depend on location $\mathbf{s}$ and allow the analogue distance to adapt to local ET controls.

The resulting permutation-importance scores are used only to weight the predictor distance, not to directly predict the reconstructed ET fields. All RF models and predictor weights are estimated from the learning data available during calibration and are then kept fixed for the independent validation experiments, preventing information from the validation years from entering the analogue-selection procedure. This weighting strategy allows the analogue distance to adapt to spatially varying ET controls, for example between humid mountainous regions, semi-arid lowlands, and irrigated areas.

4.3 Analogue distance and analogue-set definition

For each query date, similarity to all learning dates is quantified using a weighted Euclidean distance computed across all predictors. Let $X_{q,f}(\mathbf{s})$ and $X_{l,f}(\mathbf{s})$ denote the standardised value of predictor feature f at location $\mathbf{s}$ for query date q and learning date l, respectively. The index f spans the full predictor-feature vector, including the different predictor variables and temporal lags. For a spatial support R, the weighted predictor distance is

$$\Delta_{q,l}^{(R)} = \left[\sum_{f=1}^{n_f} \sum_{\mathbf{s}\in\Omega_R^*} w_f(\mathbf{s}) \left(X_{q,f}(\mathbf{s}) - X_{l,f}(\mathbf{s})\right)^2\right]^{1/2}, \tag{2}$$

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

where Ω_R^* is the set of predictor-grid cells retained after excluding cells with non-finite values, and $w_f(\mathbf{s})$ is the RF-derived importance weight of feature f at location $\mathbf{s}$. For global RF weights, $w_f(\mathbf{s})$ is constant in space and reduces to a single value w_f for each predictor feature. For pixel-level RF weights, $w_f(\mathbf{s})$ varies spatially.

The spatial support R depends on the analogue-selection configuration. For domain-wise selection, R corresponds to the full predictor domain. For tile-wise selection, R corresponds to the predictor-grid cells within the considered tile. Let $\Delta^{(R)}_{q,(k_{\mathrm{NN}})}$ denote the k_{NN}-th smallest value among the distances $\{\Delta^{(R)}_{q,l} : l \in \mathcal{L}\}$, where $\mathcal{L}$ is the set of available learning dates. The analogue set is then defined as

$$\mathcal{A}_q^{(R)} = \left\{ l \in \mathcal{L} : \Delta^{(R)}_{q,l} \leq \Delta^{(R)}_{q,(k_{\mathrm{NN}})} \right\}. \tag{3}$$

In the DW case, a single analogue set $\mathcal{A}_q^{(\mathrm{DW})}$ is used for the whole domain. In the TW case, each tile m has its own analogue set $\mathcal{A}_q^{(m)}$.

4.4 Domain-wise and tile-wise analogue selection

Two spatial supports are investigated for analogue selection. In the domain-wise (DW) configuration, the distance in Eq. 2 is computed over the whole domain, and the same analogue dates are used everywhere. This preserves basin-wide temporal coherence: each analogue corresponds to a complete historical situation. However, it also forces the method to select dates that are compromises over the whole basin, even if no historical date matches all subregions equally well.

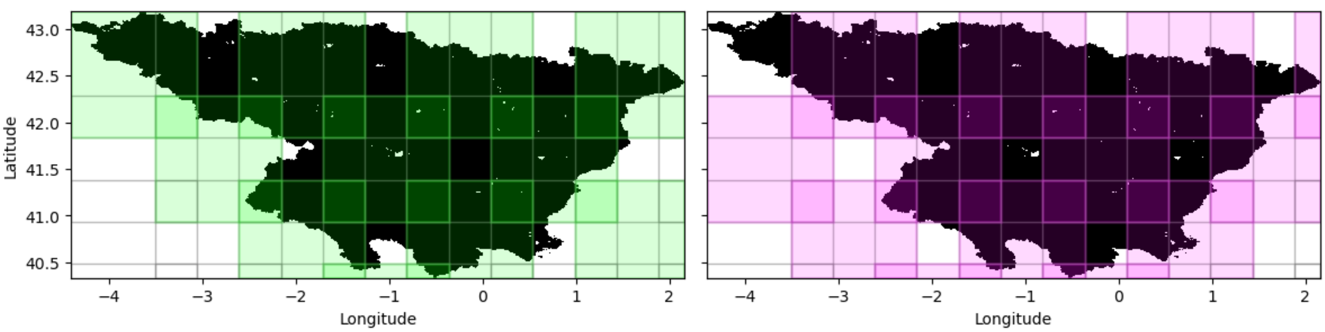

Figure 4. Chessboard-like overlapping tiling used for TW analogue selection, *chessQS* simulations, and TW aggregation. The coloured squares represent alternating tile groups for the simulations.

In the tile-wise (TW) configuration, the domain is divided into overlapping tiles of size S with overlap O (Fig. 4), and analogue selection is performed independently for each tile. The motivation is that no single historical date may be the best analogue for all parts of a heterogeneous basin. Local analogue selection can therefore adapt to regional differences in the predictor-ET relationship and select the most relevant historical information for each subregion. By allowing different tiles to draw on different analogue dates, TW selection can generate basin-wide configurations that were never observed as complete maps, while still using ET values and spatial patterns taken from the historical archive. This may become particularly relevant

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

when predictor combinations depart from the historical record, including under climate-change-driven shifts in atmospheric or surface conditions.

The tile size S controls how much spatial recombination can be introduced. Smaller tiles allow stronger local adaptation but may reduce spatial coherence, whereas larger tiles preserve broader structures but approach the DW configuration. The drawback of TW selection is that neighbouring tiles may be conditioned on different historical dates, so the aggregation step must maintain spatial continuity when local CTIS are combined. Comparing DW and TW analogues allows us to test whether local matching and recombination leads to improved results compared to basin-wide reconstructions.

4.5 Aggregation strategies

Three aggregation strategies are applied to the analogue-selected CTIS: inverse-distance weighted averaging (WA), *chessQS*, and Anchor Sampling (AS). Each aggregation strategy is combined with the two analogue-selection supports introduced above, namely domain-wise (DW) and tile-wise (TW) selection. This results in six reconstruction scenarios: DW-WA, DW-*chessQS*, DW-AS, TW-WA, TW-*chessQS*, and TW-AS. Table 1 summarizes the six reconstruction scenarios, and the three aggregation strategies are described in the following subsections.

Table 1. Reconstruction configurations considered in the comparison.

Analogue selection	Aggregation	Scenario code
Domain-wise	Weighted average	DW-WA
	chessQS	DW-*chessQS*
	Anchor Sampling	DW-AS
Tile-wise	Weighted average	TW-WA
	chessQS	TW-*chessQS*
	Anchor Sampling	TW-AS

4.5.1 Inverse-distance weighted averaging

The classical analogue estimate is obtained by averaging the ET maps associated with the selected analogues of the CTIS. Analogues closer to the query date in predictor space receive larger weights:

$$\hat{Y}_q^{(R)} = \sum_{l \in \mathcal{A}_q^{(R)}} \alpha_{q,l}^{(R)} Y_l, \qquad \alpha_{q,l}^{(R)} = \frac{\left(\Delta_{q,l}^{(R)}\right)^{-1}}{\sum_{j \in \mathcal{A}_q^{(R)}} \left(\Delta_{q,j}^{(R)}\right)^{-1}}, \tag{4}$$

where $\hat{Y}_q^{(R)}$ is the reconstructed ET map for query date q over the spatial support R, Y_l is the observed ET map for learning date l, and $\Delta_{q,l}^{(R)}$ is the predictor-space distance defined in Eq. 2. This estimator is robust and provides a natural baseline, but

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

it also represents the conditional mean of several maps. It is therefore expected to smooth local contrasts and underestimate spatial variability, especially when the selected analogues contain similar large-scale conditions but different fine-scale patterns. Extremes are also dampened as soon as more than one analogue is selected.

For TW analogues, the same approach is applied independently within each tile, using the tile-specific analogue set $\mathcal{A}_q^{(m)}$ and distance $\Delta_{q,l}^{(m)}$. In overlapping areas, the final value is obtained by averaging the estimates from all tiles covering the pixel, in order to smooth transitions and reduce visible discontinuities between neighbouring tiles.

4.5.2 *chessQS* aggregation

To avoid collapsing the CTIS into a single average, the selected analogue maps can instead be used as training images for Multiple-Point Statistics (MPS) simulation. QuickSampling (QS; Gravey and Mariethoz, 2020) generates stochastic realizations by sampling spatial patterns from training images, allowing the reconstruction to preserve structures observed in the analogue maps rather than blending or smoothing them. In this context, the analogue step ensures hydroclimatic relevance, while QS provides stochastic pattern-based spatial aggregation.

QS can become computationally demanding or inappropriate for large non-stationary images. We therefore use *chessQS* (Gerber et al., 2026b), which subdivides the simulation grid and training images into overlapping tiles arranged in two alternating groups (Fig. 4). Tiles in the first group are simulated first, and the second group is then simulated conditionally on the first. The overlap helps reduce boundary artefacts and allows continuity with already simulated neighbouring tiles, while local simulations reduce computational cost and help address non-stationarity across the domain. In addition, MPS can help preserve spatial variability and, when multiple realizations are generated, provides a natural framework for uncertainty quantification.

The same tiling geometry, defined by tile size S and overlap O, is used for both TW analogue selection and *chessQS* simulation. These two uses of tiles have different roles: in TW analogue selection, tiles define the spatial support over which local analogue distances are computed and provide tile-specific analogue sets $\mathcal{A}_q^{(m)}$; in *chessQS*, tiles define the local simulation domains used to reduce computational cost and handle non-stationarity. Reusing the same tile grid keeps the local CTIS consistent with the simulation domains and simplifies the comparison between DW and TW scenarios. In DW-*chessQS*, the tiling is used only for simulation, and all simulation tiles draw from the same domain-wise CTIS. In TW-*chessQS*, each simulation tile draws from the CTIS selected for the corresponding analogue-selection tile.

To guide the *chessQS* simulations, which are otherwise unconditional, additional covariates are incorporated. Latitude, longitude, and elevation are used to limit unwanted spatial displacement of ET patterns. In addition, a similarity-based covariate is introduced to reflect the analogue selection step. The simulation grid includes a constant 0-valued layer, while each training image is augmented with a layer equal to its analogue distance $\Delta_{q,l}^{(R)}$. This favours training images associated with smaller analogue distances during pattern matching.

The pattern search is further constrained using a kernel ω that limits the maximum spatial range of neighbouring pixels considered during pattern matching (Gravey and Mariethoz, 2020). The kernel defines the local neighbourhood used to compare patterns between the simulation grid and the training images, and is also used to assign relative weights to the target variable and auxiliary covariates during pattern comparison.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

4.5.3 Anchor Sampling aggregation

Anchor Sampling (AS) is introduced here as the most spatially constrained adaptation of the QS algorithm. In standard QS, candidate patterns can be searched across the training images and across different spatial locations, allowing spatial structures to be displaced and recombined. Increasing the spatial constraint, for example by assigning larger weights to latitude and longitude covariates, progressively restricts the set of feasible matches. AS corresponds to the limiting case of this idea: for each pixel, candidate values are sampled only from the same spatial location across the selected analogue maps.

This anchoring is motivated by the fact that several controls on ET, such as topography, land cover, river networks, or irrigation structures, are fixed in space. AS therefore preserves location-specific features while allowing temporal variability through the selected analogue maps. In this sense, variability is introduced through the analogue-selected CTIS rather than through spatial relocation of patterns.

Despite this strict spatial anchoring, AS remains pattern-based. Candidate values are evaluated using the spatial neighbourhood of the pixel to be simulated, similarly to QS. The method is therefore stochastic and accounts for spatial context through neighbouring pixels, but restricts pattern matching to homologous locations. As a result, AS does not require additional spatial covariates, such as latitude, longitude, or elevation, to guide the simulations.

In addition, the distance values computed during analogue selection are reused to define training-image weights within the AS simulation. These weights bias the stochastic selection of candidate values toward analogue maps that are more similar to the query date. For each location, the analogue map from which the candidate value is drawn is sampled with a probability proportional to the inverse selection distance, using the weights defined in Eq. 4.

4.6 Validation metrics

The six reconstruction methods are evaluated using complementary metrics that target different aspects of ET map reconstruction. Pixel-wise accuracy is quantified with RMSE, while spatial agreement is assessed with the Spatial Pattern Efficiency Metric (SPEM; Dembélé et al., 2020) and the Structural Similarity Index Measure (SSIM; Wang et al., 2004). These image-based metrics are complemented by variogram-based, water-balance, and upper-tail metrics, which respectively evaluate spatial continuity, basin-scale volume consistency, and the representation of the upper tail of the ET distribution.

Spatial continuity and variability are further assessed using empirical variograms of each reconstructed ET map with that of the corresponding observed map. For each validation date, semivariance is computed as a function of spatial lag from all valid pixels, producing one variogram curve for the observed map and one for each reconstruction. The discrepancy between these curves is then summarized using a relative variogram RMSE (Eq. 5):

$$\mathrm{VarioRMSE}_t = \frac{\left[\frac{1}{N_h}\sum_{j=1}^{N_h}\left(\gamma_t^{\mathrm{pred}}(h_j)-\gamma_t^{\mathrm{obs}}(h_j)\right)^2\right]^{1/2}}{\frac{1}{N_h}\sum_{j=1}^{N_h}\gamma_t^{\mathrm{obs}}(h_j)}, \tag{5}$$

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

where $\gamma_t^{\mathrm{pred}}(h_j)$ and $\gamma_t^{\mathrm{obs}}(h_j)$ are the reconstructed and observed empirical semivariances at lag bin h_j, and N_h is the number of valid lag bins for which both variograms are finite. The numerator is the RMSE between the reconstructed and observed variogram curves. This error is normalized by the mean observed semivariance over the same lag bins, so that variogram discrepancies are evaluated relative to the overall spatial variability of each observed map. This prevents dates with larger absolute semivariance values from dominating the metric. The reported value is the average of $\mathrm{VarioRMSE}_t$ over all validation dates.

Hydrological consistency is evaluated using basin-scale water-balance metrics. These metrics assess whether the reconstructed ET fields preserve the total amount of evapotranspiration when integrated over the basin, which is important for hydrological applications even when pixel-wise errors or spatial-pattern metrics are satisfactory. These metrics are interpreted here as relative errors in basin-scale ET volume. For any set of validation timesteps $\mathcal{T}$, the water-balance bias (Eq. 6) is defined as

$$B_{\mathrm{WB}}(\mathcal{T}) = \frac{\sum_{t\in\mathcal{T}}\left(V_t^{\mathrm{pred}} - V_t^{\mathrm{obs}}\right)}{\sum_{t\in\mathcal{T}} V_t^{\mathrm{obs}}}, \qquad \mathrm{WBE}(\mathcal{T}) = \left|B_{\mathrm{WB}}(\mathcal{T})\right|, \tag{6}$$

where V_t^{obs} and V_t^{pred} are the basin-integrated observed and reconstructed ET volumes at timestep t. When $\mathcal{T}$ contains all validation dates, the metric evaluates whether the reconstruction conserves total ET volume over the complete validation period.

The same definition is also applied independently to each timestep by setting $\mathcal{T} = t$. The reported timestep-scale metrics are the mean timestep WB bias ($\overline{B_{\mathrm{WB},t}}$) and mean timestep WBE ($\overline{\mathrm{WBE}_t}$), obtained by averaging $B_{\mathrm{WB}}(t)$ and $\mathrm{WBE}(t)$ over all validation dates. Unlike the global metric, these summaries give equal weight to each validation timestep. They therefore describe the typical basin-scale volume bias and absolute volume error of individual ET maps.

The upper tail of the ET distribution is evaluated using the signed relative bias of the 0.90-quantile, denoted $Q_{0.90}$ (Eq. 7). This metric is included to specifically evaluate high ET values, which are expected to be sensitive to the smoothing effect of weighted-average analogue aggregation. Although a full quantile-bias curve could be computed, the objective here is not to compare the entire distribution, but to assess whether the reconstruction preserves high ET values. The corresponding bias is defined as

$$B_{Q_{0.90}} = \frac{Q_{0.90}\left(\hat{Y}\right) - Q_{0.90}\left(Y\right)}{Q_{0.90}\left(Y\right)}, \tag{7}$$

where Y and $\hat{Y}$ denote the sets of observed and reconstructed ET values restricted to pixels and timesteps for which both values are finite.

A timestep-scale version of the same metric is also computed by evaluating the observed and reconstructed $Q_{0.90}$ independently for each validation date before computing the signed relative bias. The reported timestep-scale bias ($\overline{B_{Q_{0.90},t}}$) is the average of these date-wise biases over all validation timesteps.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

5 Calibration and experimental design

5.1 Calibration and validation periods

As previously explained, two independent sets of withheld years are used in this study (Fig. 2). The first set is used to tune the method parameters, and the second set is used only to compare the final reconstruction scenarios. During calibration, the years 2000, 2005, 2010, 2015, 2020, and 2025 are withheld simultaneously as query years, while all remaining years are used as learning dates. These calibration years are used to select the RF weighting strategy, analogue-selection parameters, tile geometry, and stochastic simulation parameters. The method comparison is then performed on an independent validation period, for which the years 2002, 2007, 2012, 2017, and 2022 are withheld as query years, while all other years are used as learning dates.

The calibration is performed sequentially rather than by jointly optimising all parameters to keeps the search computationally feasible. The sequence of calibration steps is summarized in Table 2.

Table 2. Sequential calibration steps used before the final method comparison.

Step	Calibrated component	Configuration used for calibration	Selected parameter(s)
1	Predictor weighting strategy	DW weighted-average analogue baseline	Global vs. pixel-level RF weights
2	DW analogue selection	DW weighted-average analogue baseline with selected RF weights	Number of analogues k_{NN} and temporal window T_w
3	Tile geometry	TW weighted-average analogue baseline	Tile size S and overlap O
4	TW analogue selection	TW weighted-average analogue baseline with selected tile geometry	Number of analogues k_{NN} and temporal window T_w
5	Stochastic simulation	DW-*chessQS* and DW-AS on reduced calibration dates, calibrated separately	Method-specific pattern-search parameters: $(k_{\mathrm{QS}}, n_{\mathrm{QS}})$ for *chessQS* and $(k_{\mathrm{AS}}, n_{\mathrm{AS}})$ for AS

At each step, performance is assessed using mean RMSE over the calibration years. Once selected, the calibrated parameters are kept fixed for the validation.

5.2 RF weighting strategy

Two predictor-weighting strategies are compared: globally derived RF weights and pixel-level RF weights. This comparison is performed using the DW weighted-average analogue baseline, so that the effect of the weighting strategy can be assessed before introducing tile-wise selection or stochastic aggregation.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

Table 3. Performance comparison between global and pixel-level RF predictor-weighting strategies.

	RMSE ↓	SPEM ↑	VarioRMSE ↓
Global weights	1.861	0.343	0.388
Pixel-level weights	**1.786**	**0.361**	**0.357**

The improvement obtained with pixel-level RF weights (Table 3) indicates that the relevance of the predictors varies spatially across the basin. Compared with global weights, pixel-level weights reduce RMSE and VarioRMSE and increase SPEM, suggesting that locally adapted analogue distances select more appropriate historical situations than a single basin-wide weighting scheme. This result is consistent with the hydroclimatic heterogeneity of the Ebro basin, where ET is controlled by different combinations of atmospheric forcing, topography, and water availability depending on location. Pixel-level RF weights are therefore retained for all subsequent calibration steps and validation scenarios.

5.3 Domain-wise analogue-selection calibration

Using the pixel-level RF weights, the DW analogue-selection parameters are calibrated by grid search over the number of analogues k_{NN} and the temporal window length T_w in days. The DW weighted-average analogue method is used as the baseline configuration. The minimum RMSE is obtained for $k_{\mathrm{NN}} = 16$ and $T_w = 2$ (Fig. 5a), which are retained for all DW scenarios in the final comparison.

5.4 Tile size and overlap calibration

The tile size S and overlap O are calibrated for TW analogue selection and then also used for *chessQS* simulations, ensuring consistency between local analogue conditioning and local stochastic simulation.

The best RMSE is obtained for a tile size of $S = 400$ pixels and an overlap of $O = 200$ pixels, corresponding to 400 km and 200 km, respectively, on the 1 km target grid (Fig. 5b). However, performance differences across configurations are limited, with mean RMSE values ranging from approximately 1.790 to 1.815 mm/8-day. Because the objective of TW selection is to allow local analogue conditioning, the largest tiles are not necessarily the most appropriate choice despite their slightly lower RMSE. Large tiles reduce the local character of the method and increase the computational cost of *chessQS*. We therefore adopt a compromise configuration of $S = 150$ and $O = 50$, which keeps the RMSE close to the optimum while preserving more local conditioning.

This choice preserves tiles large enough to capture relevant spatial correlation structures while keeping the simulations sufficiently local to address non-stationarity and reduce computational cost. All subsequent TW and *chessQS* experiments use this tile size and overlap.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

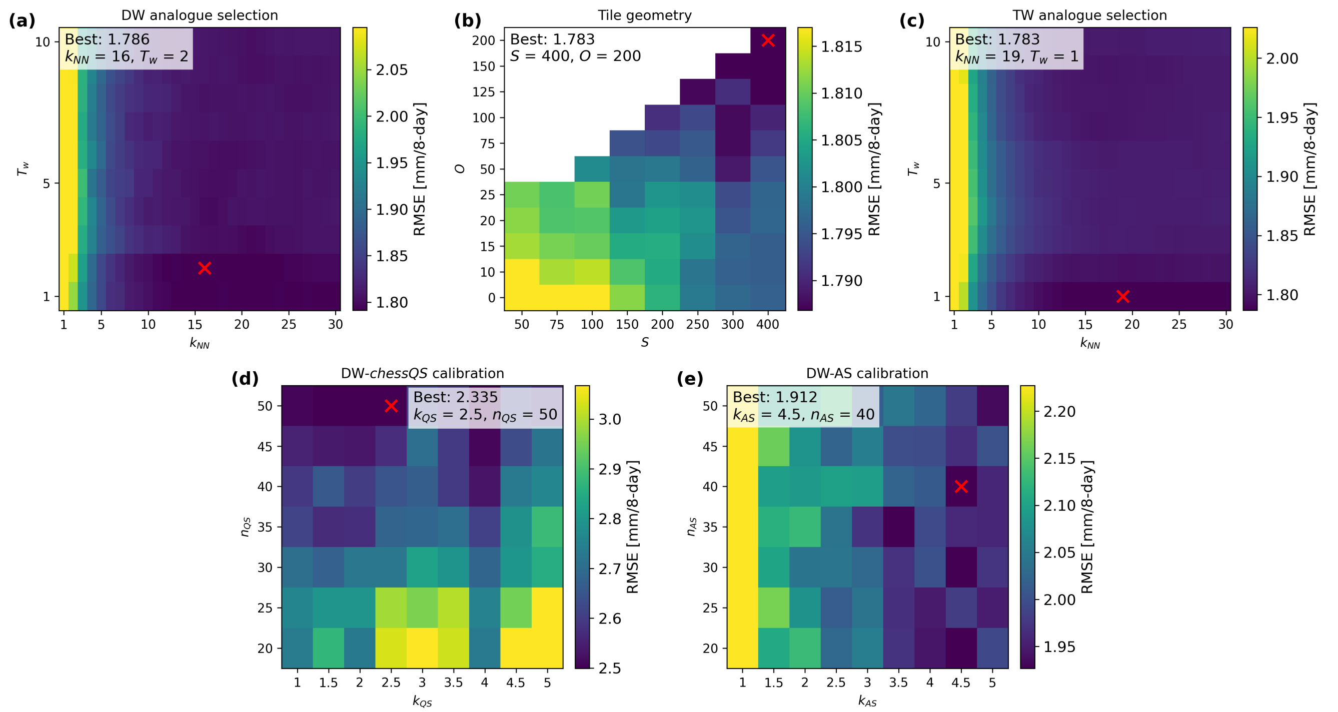

Figure 5. Calibration grid searches used before the final method comparison. (a) DW analogue-selection calibration over the number of analogues k_{NN} and temporal window length T_w. (b) Tile-size and overlap calibration for TW analogue selection. (c) TW analogue-selection calibration over k_{NN} and T_w, using $S = 150$ and $O = 50$. (d) *chessQS* stochastic simulation calibration over neighbourhood size k_{QS} and number of pixels n_{QS} used in the pattern comparison. (e) Anchor Sampling calibration over neighbourhood size k_{AS} and number of pixels n_{AS} used in the pattern comparison. Red crosses indicate the selected parameter combinations.

5.5 Tile-wise analogue-selection calibration

After fixing $S = 150$ and $O = 50$, the analogue-selection parameters are recalibrated for the TW configuration. This second grid search is necessary because analogue distances are computed over local tiles rather than over the full domain, and the optimal number of analogues and temporal window may therefore differ from the DW case.

The minimum RMSE is obtained for $k_{\mathrm{NN}} = 19$ and $T_w = 1$ (Fig. 5c). The TW optimum differs from the DW optimum and yields a small improvement over the DW weighted-average baseline. This indicates that local analogue conditioning can provide additional skill, even at the scale of the Ebro watershed, although the improvement remains limited.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

5.6 Stochastic simulation calibration

The *chessQS* and AS parameters controlling stochastic pattern selection are calibrated independently from the analogue-selection step. This calibration is performed on a reduced set of validation dates to limit computational cost, while keeping the analogue-selection parameters fixed. This reduction is necessary because a single *chessQS* reconstruction requires approximately three minutes on the computing environment used in this study, whereas weighted averaging and AS are substantially faster. The selected dates correspond to 01.01, 06.03, 10.06, and 14.09 for the years 2000, 2005, 2010, 2015, 2020, and 2025.

For computational tractability, the stochastic simulation parameters are calibrated in the DW configuration only. This choice avoids multiplying the number of expensive stochastic simulations while still providing representative pattern-search settings. The same calibrated stochastic parameters are then used in the corresponding DW and TW stochastic scenarios.

Separate grid searches are conducted for *chessQS* and AS because the two methods differ in their sampling constraints and computational cost. The searches evaluate combinations of neighbourhood size $k_{\mathrm{QS/AS}}$ and number of pixels $n_{\mathrm{QS/AS}}$ used in the pattern comparison.

For *chessQS*, the kernel size is set to 150×150 pixels, matching the selected tile size. This choice allows the pattern search to use neighbourhood information at the same spatial scale as the local simulation domains. All variables have a spatially constant weight of 1, except for latitude and longitude which have a weight of 3. The larger coordinate weights are used to favour candidates from similar geographic positions and to limit unrealistic displacement of ET patterns across the domain. These values are kept fixed for all *chessQS* experiments and should be interpreted as pragmatic conditioning parameters rather than optimized weights. The lowest RMSE is obtained for $k_{\mathrm{QS}} = 2.5$ and $n_{\mathrm{QS}} = 50$ (Fig. 5d), which are retained for both DW-*chessQS* and TW-*chessQS*.

A similar grid search is performed for Anchor Sampling using the same validation dates and evaluation framework. As AS is considerably faster than QS and *chessQS*, the grid search is expanded to cover more parameter combinations. The kernel used for AS is also set to 150×150 pixels and has a single weight of 1 for the target variable, as no covariates are used in the AS simulations. The lowest RMSE is obtained for $k_{\mathrm{AS}} = 4.5$ and $n_{\mathrm{AS}} = 40$ (Fig. 5e), which are retained for both DW-AS and TW-AS.

5.7 Comparison of aggregation scenarios

The final experiments compare the six reconstruction scenarios obtained by combining the two analogue-selection supports, DW and TW, with the three aggregation strategies: WA, *chessQS*, and AS. All scenarios use the same predictor set, pixel-level RF weights, and independent validation years. The analogue-selection parameters are fixed separately for the DW and TW configurations according to their calibration results. Latitude, longitude, and elevation are used only as auxiliary covariates for the *chessQS* simulations.

Although *chessQS* and AS can generate ensembles, the present comparison uses one realization per validation date because of the number of reconstructed dates and the computational cost of the stochastic simulations. The evaluation therefore focuses on reconstruction behaviour across the validation years.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

6 Results

6.1 Overall reconstruction performance

Table 4. Validation results for the six reconstruction scenarios. Hyperparameters are calibrated on independent calibration years, and final performance is evaluated on the validation years 2002, 2007, 2012, 2017, and 2022. For signed bias metrics, the best value is the value closest to zero.

	DW			TW		
	WA	*chessQS*	AS	WA	*chessQS*	AS
RMSE [mm/8-day] ↓	**1.718**	2.564	2.064	**1.718**	2.844	2.056
VarioRMSE [-] ↓	0.361	0.440	**0.317**	0.355	0.422	0.341
SPEM [-] ↑	0.380	0.149	0.230	**0.383**	0.022	0.230
SSIM [-] ↑	**0.874**	0.738	0.789	**0.874**	0.687	0.793
B_{WB} [%] $\rightarrow 0$	3.226	**1.747**	1.760	3.856	3.532	2.746
$\overline{\mathrm{WBE}_t}$ [%] ↓	8.900	10.641	8.452	8.925	9.054	**8.187**
$\overline{B_{\mathrm{WB},t}}$ [%] $\rightarrow 0$	4.065	**2.573**	2.615	4.678	4.117	3.639
$B_{Q_{0.90}}$ [%] $\rightarrow 0$	1.595	1.201	**0.901**	2.136	2.102	1.351
$\overline{B_{Q_{0.90},t}}$ [%] $\rightarrow 0$	-0.562	1.454	0.249	**0.045**	1.962	1.322

Table 4 summarizes the performance of the six reconstruction scenarios. The weighted-average analogue approaches provide the best pixel-wise and image-structure performance. The lowest RMSE is obtained by the two WA configurations, with DW-WA and TW-WA both equal to 1.718 mm/8-day after rounding. TW-WA also gives the highest SPEM (0.383) and SSIM (0.874), although DW-WA remains very close.

The stochastic aggregation methods increase RMSE and reduce SPEM and SSIM relative to WA. AS remains closer to the deterministic baseline than *chessQS*, with RMSE values of 2.064 and 2.056 mm/8-day for DW-AS and TW-AS, respectively. In comparison, DW-*chessQS* and TW-*chessQS* obtain higher RMSE values of 2.564 and 2.844 mm/8-day. The same ranking is observed for SPEM and SSIM, indicating that the greater flexibility of *chessQS* introduces larger local and structural mismatches than AS.

In contrast, AS improves the variogram-based metric. DW-AS gives the lowest VarioRMSE (0.317), followed by TW-AS (0.341). Both WA configurations remain competitive, whereas the two *chessQS* scenarios show larger VarioRMSE values. This indicates that, in the present validation, the spatially anchored stochastic sampling of AS better preserves the observed spatial variability than the more flexible pattern recombination of *chessQS*.

Hydrological consistency shows a more nuanced behaviour. The lowest whole-period B_{WB} is obtained by DW-*chessQS* (1.747%), closely followed by DW-AS (1.760%). However, at the timestep scale, TW-AS gives the lowest $\overline{\mathrm{WBE}_t}$ (8.187%), while DW-*chessQS* gives the mean timestep bias $\overline{B_{\mathrm{WB},t}}$ closest to zero. For the upper-tail metrics, DW-AS gives the lowest global $B_{Q_{0.90}}$ (0.901%), whereas TW-WA gives the mean timestep bias $\overline{B_{Q_{0.90},t}}$ closest to zero.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

6.2 Spatial and temporal behaviour of the reconstructions

Figure 6 illustrates the spatial behaviour of the six scenarios for the 09.06.2012, as an example. The weighted-average reconstructions reproduce the main large-scale ET gradients visible in the MODIS reference map, including the higher ET values in the northern part of the basin and lower values over the central and southern regions. The DW-WA and TW-WA maps are very similar, consistent with the small differences observed in Table 4. This example illustrates that, for this date at least, tile-wise analogue selection does not strongly alter or improve the large-scale reconstruction relative to the domain-wise analogue set.

The stochastic methods show more contrasted behaviours. AS preserves much of the spatial structure of the weighted-average reconstructions while introducing stronger local variability. This behaviour is consistent with its lower VarioRMSE and improved upper-tail bias, but also with its larger RMSE and lower SPEM and SSIM. AS therefore improves some variability- and distribution-related properties while increasing local pixel-wise mismatch. The *chessQS* reconstructions, however, show stronger departures from the reference map in this reconstruction setting. Both DW-*chessQS* and TW-*chessQS* produce lower SPEM and SSIM values than AS, indicating less agreement with the reference spatial structure for the present validation experiment. TW-*chessQS* appears smoother and less consistent with some of the observed regional contrasts. This suggests that combining tile-wise analogue selection with flexible spatial pattern recombination can introduce spatial configurations that are not sufficiently constrained by the available local CTIS or covariates.

Figure 7 shows the temporal evolution of ET at a representative central pixel for 2012, as an example. Both WA scenarios follow the seasonal cycle well, but their trajectories are smoother than the observations. In particular, they underestimate the amplitude of several high-ET events during summer, which illustrates the tendency of the aggregation method to dampen extremes.

AS produces more variable time series and better captures some short-term fluctuations, although individual peaks may still be over- or underestimated. In particular, DW-AS produces pronounced summer peaks that are not always aligned with the observations. In contrast, the *chessQS* time series shown here underestimate the summer maximum, especially for TW-*chessQS*, and show unrealistic variability between timesteps.

6.3 Computational cost

Table 5 reports indicative wall-clock runtimes for the six reconstruction scenarios. The values are intended as implementation-dependent indicators of relative computational cost rather than hardware-independent benchmarks. The timings were measured in the calling script around the full reconstruction of one validation date for each scenario. They therefore reflect the cost of running the corresponding reconstruction workflow, rather than only the internal runtime of the aggregation routine. Weighted averaging is the fastest approach for both DW and TW analogue selection, with TW-WA requiring only a small additional cost relative to DW-WA.

AS adds only a moderate computational overhead in the DW configuration, with DW-AS requiring about 1.4 times the runtime of DW-WA. TW-AS is more expensive, requiring about 7.3 times the runtime of DW-WA, because the AS procedure

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

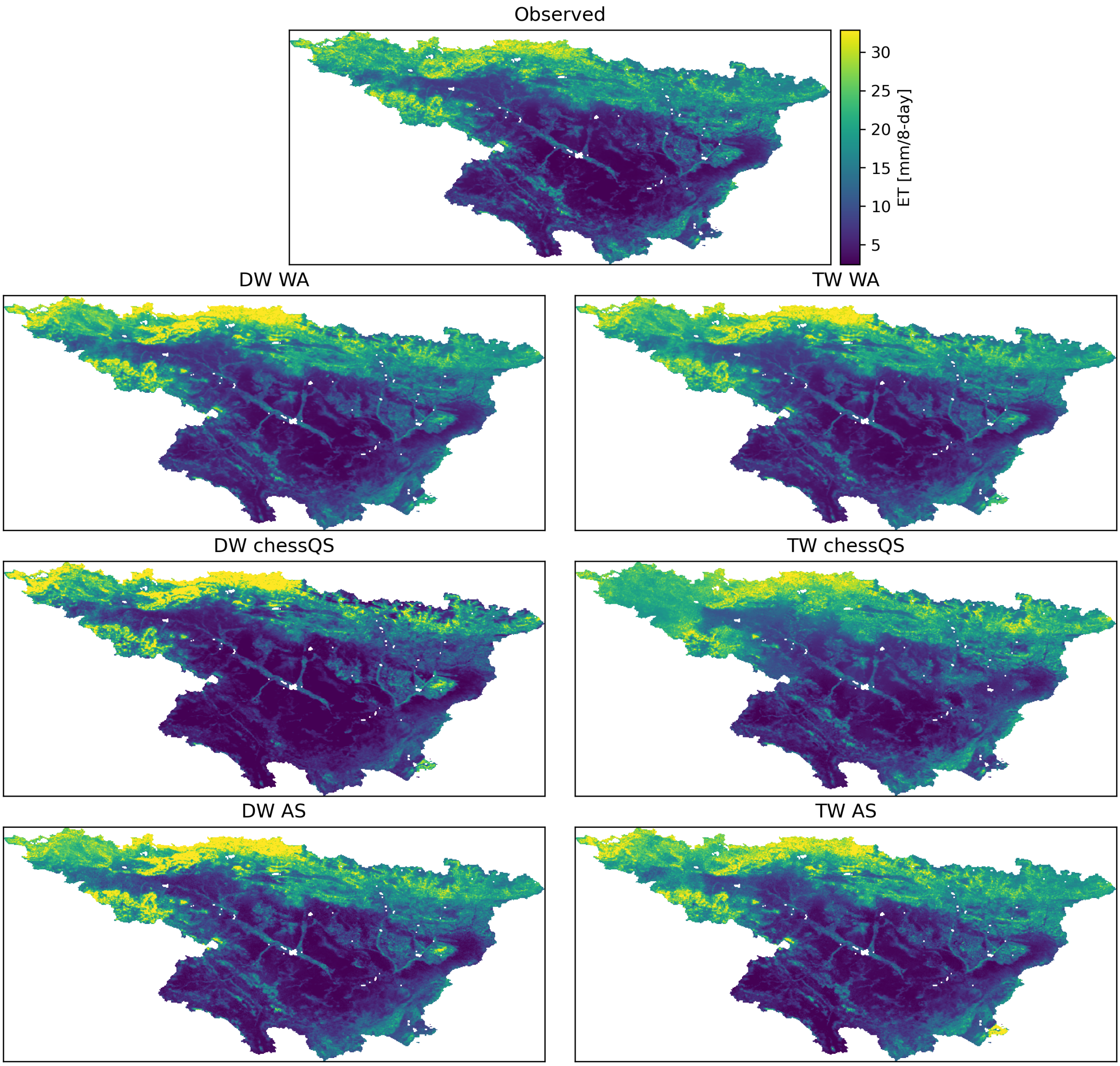

Figure 6. Example reconstruction maps for one validation date (09.06.2012). The observed MODIS ET map is shown at the top, followed by the six reconstruction scenarios. White areas indicate persistent missing or invalid MODIS pixels, which are kept empty and not reconstructed.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

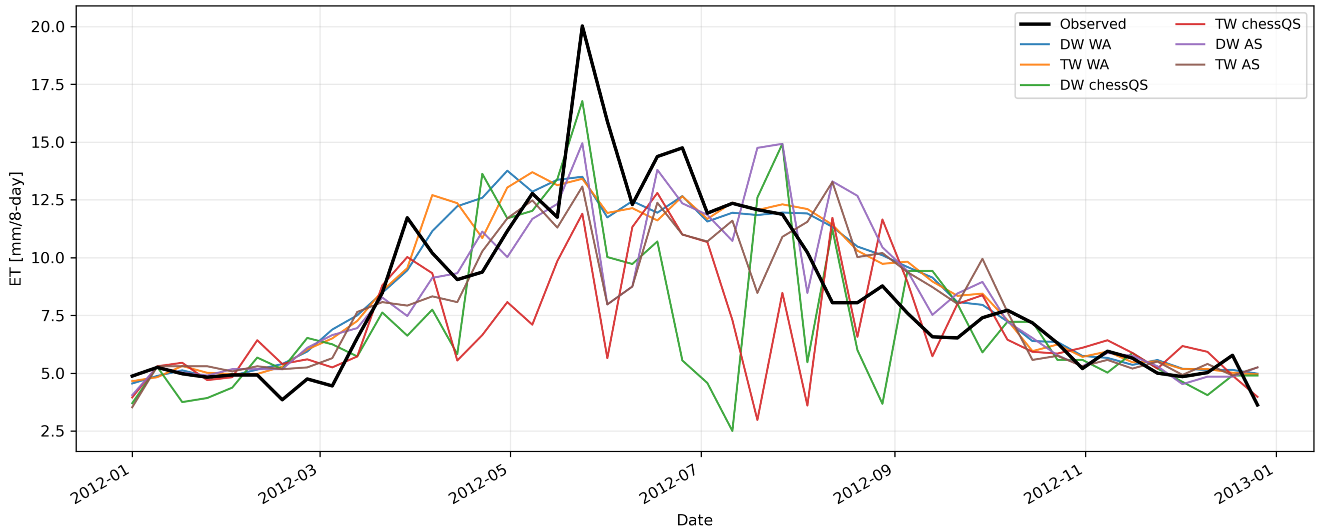

Figure 7. Observed and reconstructed ET time series for a representative central pixel during the 2012 validation year.

Table 5. Indicative computational cost of the reconstruction scenarios. Reported times correspond to the average wall-clock time required to reconstruct one validation date on the computing environment used in this study. Timings were measured externally in the calling script around the complete reconstruction call for each scenario. Relative costs are computed with respect to DW-WA.

Scenario	Mean runtime per date	Relative cost
DW-WA	2.8 s	1.0
TW-WA	3.5 s	1.2
DW-AS	4.1 s	1.4
TW-AS	20.3 s	7.3
DW-*chessQS*	2.52 min	53.8
TW-*chessQS*	2.95 min	63.2

is called repeatedly for the different tiles of each query date. Nevertheless, AS remains substantially less demanding than *chessQS*.

The *chessQS* scenarios are the most computationally demanding, requiring about 54-63 times the runtime of DW-WA. This is expected, as each reconstruction requires stochastic pattern searches within the selected CTIS and local simulations over the tiled domain. DW-*chessQS* and TW-*chessQS* have comparable runtimes because the same *chessQS* tiling structure, kernel size, and simulation parameters are used in both cases; only the CTIS assigned to each tile differs.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

7 Discussion

The results show that weighted-average analogue reconstruction remains difficult to outperform in terms of pixel-wise accuracy and image-structure agreement. Both DW-WA and TW-WA achieve the lowest RMSE and the highest SPEM and SSIM values, indicating that inverse-distance averaging provides an accurate reconstruction of MODIS ET over the Ebro watershed. This suggests that the analogue-selection step captures much of the relationship between the climatic and meteorological predictors and ET. However, the time-series example (Fig. 7) also illustrates its main limitation: averaging reproduces the seasonal evolution but dampens short-term variability and high ET values. WA is therefore not systematically optimal for all criteria. The stochastic methods, especially AS, improve some metrics related to spatial variability, water-balance consistency, and the upper tail of the ET distribution. This confirms that the preferred aggregation strategy depends on whether the objective is accurate deterministic reconstruction or better preservation of variability, both spatial and temporal, and distributional properties.

The calibration results also support the use of spatially varying predictor weights. Pixel-level RF weights improve the weighted-average baseline relative to global weights, indicating that a single predictor weighting is less effective at representing the heterogeneous controls on ET across the basin. This supports the broader motivation for local analogue selection, although the final TW results show that locally weighting the predictor distance is more beneficial than recombining analogues through square tiles in the present domain.

The added value of tile-wise analogue selection is limited in the present configuration. TW-WA only marginally improves over DW-WA, suggesting that a single basin-wide analogue set already provides a strong compromise for the Ebro watershed and the available predictors. Nevertheless, the similar performance of both configurations shows that local analogue recombination is feasible without degrading reconstruction quality. This is relevant because TW selection relaxes the requirement that a complete historical map must represent the whole basin simultaneously. Its potential may therefore become more apparent over larger or more heterogeneous domains, or under new combinations of predictors for which no single complete historical analogue is sufficiently representative.

Stochastic aggregation shows a clear trade-off. AS and *chessQS* do not improve RMSE, SPEM, or SSIM relative to WA, which indicates that stochastic simulation introduces local departures from the observed maps. However, AS improves VarioRMSE and the global upper-tail bias $B_{Q_{0.90}}$, and TW-AS gives the lowest mean timestep water-balance error $\overline{\mathrm{WBE}_t}$. These results show that stochastic aggregation should not be assessed only through pixel-wise accuracy. AS appears to provide the best stochastic alternative, as it remains closer to WA than *chessQS* in terms of RMSE and spatial-structure metrics while improving several variability- and hydrology-related criteria.

The upper-tail metrics provide additional insight into the effect of aggregation on the upper tail of the ET distribution. Weighted averaging remains accurate in terms of RMSE, but it does not result in the lowest global $Q_{0.90}$ bias. DW-AS gives the smallest global upper-tail bias, suggesting that spatially anchored stochastic sampling can better preserve the upper tail of ET values when all valid pixels and validation dates are pooled. However, TW-WA gives the timestep-averaged bias $\overline{B_{Q_{0.90},t}}$ closest to zero, indicating that the benefit of stochastic sampling depends on whether upper-tail ET values are evaluated globally or date by date.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

In the present complete-map reconstruction application, the greater flexibility of *chessQS* does not lead to improved reconstruction quality. Both DW-*chessQS* and TW-*chessQS* perform worse than AS and WA in terms of RMSE, SPEM, SSIM, and VarioRMSE. Although DW-*chessQS* gives the lowest whole-period water-balance error, this improvement is small relative to DW-AS and does not translate into better spatial or pixel-wise agreement. These results suggest that, for the present application, the spatial structures generated by *chessQS* are not sufficiently constrained by the available covariates and CTIS to match the observed ET fields. This does not preclude potential advantages of *chessQS* in more degraded reconstruction settings, for example when observations contain large gaps, strong noise, or spatial structures that cannot be represented by location-constrained sampling. The issue is more pronounced for TW-*chessQS*, where each tile is simulated from a locally selected CTIS. If neighbouring tiles are conditioned on substantially different analogue sets, the resulting simulations can display discontinuities across tile boundaries or locally inconsistent structures despite the use of overlaps.

Computational cost reinforces the distinction between the methods. WA is by far the fastest approach and also provides the best pixel-wise and image-structure performance, making it the most practical option when the objective is deterministic reconstruction over many dates. AS represents an intermediate compromise: DW-AS remains relatively efficient because it requires a single AS simulation for each query date, whereas TW-AS becomes slower because the AS procedure is called repeatedly for the different tiles. *chessQS* is substantially more demanding computationally, as each reconstruction requires local pattern searches over the tiled simulation grid. In the present configuration, this additional cost is not compensated by improved accuracy or spatial-structure performance.

Several limitations should be considered when interpreting these results. First, the stochastic simulation parameters were calibrated in the DW configuration only and subsequently transferred to TW simulations for computational cost reasons. Consequently, the poorer performance of TW stochastic scenarios may partly reflect parameter transfer across conditioning scales rather than the limitation of tile-wise stochastic aggregation alone. Second, the stochastic methods are evaluated using one realization per validation date because of the large number of dates and the computational cost of the simulations. This is sufficient to assess their average reconstruction behaviour in the present study, but not to fully evaluate their potential for uncertainty quantification. Third, the limited improvement of TW selection, as well as the poor behaviour of TW-*chessQS*, may partly result from using square overlapping tiles. Although convenient, such tiles are not necessarily hydrologically or climatologically coherent and may either combine areas governed by different ET controls or separate areas with similar behaviour. Alternative partitions based on elevation, land cover, climatic zones, or predictor similarity could provide more meaningful local CTIS and reduce inconsistencies between neighbouring simulations. Finally, testing the framework over larger and more heterogeneous domains would help determine whether local analogue selection and stochastic spatial recombination become more advantageous when basin-wide historical analogues are less representative. For such larger-domain applications, such as a European-scale reconstruction, graph-based partitioning or graph-cut approaches could be used to define contiguous zones that are internally coherent with respect to these controls (Thao et al., 2022; Schmutz et al., 2025; Vrac et al., 2024). More generally, the results should be interpreted in terms of suitability for different reconstruction tasks rather than as a single ranking of methods. Weighted averaging is well suited to deterministic reconstruction when the objective is pixel-wise accuracy and temporal robustness over many dates. AS provides a more spatially constrained stochastic alternative, allowing variability

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

to be introduced while preserving location-specific controls. *chessQS*, by design, offers greater flexibility in spatial pattern recombination. This flexibility was not advantageous in the present complete-map reconstruction application, where the target fields are already fully observed and the main challenge is to reproduce them as accurately as possible. However, it may become useful in more degraded or structurally complex settings, for example when observations contain large gaps, strong noise, or changes in surface structure. A related perspective concerns applications where land-use and land-cover change is explicitly considered. Although land-use land-cover change is not addressed in the present study, it represents a setting in which spatial structures themselves may evolve through time. In such cases, *chessQS* could help recombine spatial structures from historical analogues in ways that are not possible with weighted averaging or strictly spatially anchored sampling, provided that dynamic surface covariates are available to constrain the simulations.

8 Conclusion

This study evaluated whether analogue-based reconstruction of MODIS evapotranspiration (ET) maps over the Ebro watershed can be improved by introducing local analogue selection and stochastic pattern-based aggregation. The proposed framework separates the analogue-selection step, which identifies climatically similar historical situations, from the aggregation step, which determines how the selected analogue maps are transformed into a reconstruction. Six scenarios were compared by combining domain-wise and tile-wise analogue selection with inverse-distance weighted averaging, *chessQS*, and Anchor Sampling.

The results show that the weighted-average analogue method remains a strong baseline. Both domain-wise and tile-wise weighted averaging produced accurate and spatially consistent reconstructions, with acceptable water-balance errors and very similar performance between DW and TW analogue-selection. This indicates that, for the Ebro watershed and the predictors used here, the selected basin-wide analogue sets already provide a good representation of the ET response. Tile-wise analogue selection was viable and did not degrade the reconstruction, but its added value remained limited in this application.

Stochastic aggregation introduced additional spatial variability but did not improve pixel-wise accuracy or image-structure agreement relative to weighted averaging. Anchor Sampling provided the most balanced stochastic alternative: it improved variogram agreement, mean timestep water-balance error, and the global upper-tail ET bias, while remaining closer to weighted averaging than *chessQS* in terms of RMSE, SPEM, and SSIM. In contrast, *chessQS* provided greater spatial flexibility but did not improve spatial-continuity metrics or image-structure agreement in the present configuration. These results suggest that MPS-based stochastic aggregation can help preserve both spatial and temporal variability, but requires sufficiently strong conditioning to avoid misplaced or inconsistent patterns.

From a practical perspective, the results highlight a trade-off between accuracy, stochasticity, and computational cost. Weighted averaging is fast, simple, interpretable, and remains the most suitable option when the objective is deterministic reconstruction over many dates. Anchor Sampling offers a promising compromise when stochasticity, spatial variability, accurate timestep water-balance behaviour, or good upper-tail ET distribution representation are desired at moderate computational cost. *chessQS* provides greater spatial flexibility, but its high computational cost and reduced reconstruction accuracy limit its practical advantage in the present configuration.

Future work should investigate whether the benefits of local analogue selection and MPS-based aggregation become stronger over larger or more heterogeneous domains, where a single basin-wide analogue set may be insufficient. The potential of *chessQS* may also be greater in applications where land-use and land-cover change is explicitly considered. Although land use land cover change is not addressed in the present study, it represents a setting in which spatial structures themselves may evolve and flexible pattern recombination could become useful, provided that dynamic surface covariates are available to constrain the simulations. The definition of local analogue domains could also be improved by replacing square tiles with hydrologically or climatologically meaningful regions. Finally, generating multiple stochastic realizations per query date would allow a fuller assessment of uncertainty representation and probabilistic performance. Overall, this study shows that coupling analogues with stochastic simulation is promising but not trivial.

Code and data availability. The exact version of the comparison workflow used to produce the results in this paper, including the *chessQS* implementation used in the experiments, together with the input files, selected reconstruction outputs, and the frozen version of the G2S package containing Anchor Sampling, is archived on Zenodo at https://doi.org/10.5281/zenodo.21105051 (Gerber, 2026). The comparison code is distributed under the GPL-3.0 licence. The development version of the comparison workflow is available from https://github.com/LoicGerber/stochastic-analogue-reconstruction. A stand-alone development version of *chessQS* is available from https://github.com/LoicGerber/ChessQS, while the current development version of G2S is available from https://github.com/GAIA-UNIL/G2S.

Author contributions. **LG:** Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Validation, Visualization, Writing - original draft, Writing - review & editing. **MG:** Methodology, Software, Writing - review & editing. **GM:** Conceptualization, Methodology, Supervision, Project administration, Writing - review & editing.

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

Acknowledgements. During the preparation of this work the authors used AI to assist with coding implementation of the method and with proof-reading. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., and Hain, C. R.: Remote sensing of drought: Progress, challenges and opportunities, Reviews of Geophysics, 53, 452–480, https://doi.org/10.1002/2014RG000456, 2015.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration: Guidelines for computing crop water requirements, no. 56 in FAO Irrigation and Drainage Paper, Food and Agriculture Organization of the United Nations, Rome, 1998.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Chilès, J.-P. and Delfiner, P.: Geostatistics: Modeling Spatial Uncertainty, Wiley, 2 edn., https://doi.org/10.1002/9781118136188, 2012.

Chinasho, A., Bedadi, B., Lemma, T., Tana, T., Hordofa, T., and Elias, B.: Evaluation of Seven Gap-Filling Techniques for Daily Station-Based Rainfall Datasets in South Ethiopia, Advances in Meteorology, 2021, 9657 460, https://doi.org/10.1155/2021/9657460, 2021.

De Lannoy, G. J. M., Bechtold, M., Albergel, C., Brocca, L., Calvet, J.-C., Carrassi, A., Crow, W. T., De Rosnay, P., Durand, M., Forman, B., Geppert, G., Girotto, M., Hendricks Franssen, H.-J., Jonas, T., Kumar, S., Lievens, H., Lu, Y., Massari, C., Pauwels, V. R. N., Reichle, R. H., and Steele-Dunne, S.: Perspective on satellite-based land data assimilation to estimate water cycle components in an era of advanced data availability and model sophistication, Frontiers in Water, 4, https://doi.org/10.3389/frwa.2022.981745, 2022.

Dembélé, M., Ceperley, N., Zwart, S. J., Salvadore, E., Mariethoz, G., and Schaefli, B.: Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies, Advances in Water Resources, 143, 103 667, https://doi.org/10.1016/j.advwatres.2020.103667, 2020.

European Commission, Joint Research Centre: Fact Sheet: Ebro River Basin, Tech. rep., European Commission, Joint Research Centre, https://water.jrc.ec.europa.eu/pdf/ebro-fs.pdf, 2021.

European Space Agency and Airbus: Copernicus DEM, https://doi.org/10.5270/ESA-c5d3d65, institution: European Space Agency, 2022.

Fluhrer, A., Baur, M. J., Piles, M., Bayat, B., Rahmati, M., Chaparro, D., Dubois, C., Hellwig, F. M., Montzka, C., Kübert, A., Mueller, M. M., Augscheller, I., Jonard, F., Schellenberg, K., and Jagdhuber, T.: Assessing evapotranspiration dynamics across central Europe in the context of land-atmosphere drivers, Biogeosciences, 22, 3721–3746, https://doi.org/10.5194/bg-22-3721-2025, 2025.

Gerber, L.: Novel analogue-based and geostatistical approaches for space-time prediction of environmental variables, https://doi.org/10.5281/zenodo.21105051, 2026.

Gerber, L., Dembélé, M., and Mariéthoz, G.: Analogue-based reconstruction of hydrological forcing for modelling in data-scarce regions: Application to the Volta River Basin, Journal of Hydrology: Regional Studies, 66, 103 571, https://doi.org/10.1016/j.ejrh.2026.103571, 2026a.

Gerber, L., Lambiel, A., Külling, N., and Mariéthoz, G.: chessQS: An efficient tile-based gap-filling method for large and complex remote sensing images, Ecological Informatics, 94, 103 666, https://doi.org/10.1016/j.ecoinf.2026.103666, 2026b.

Goovaerts, P.: Geostatistics for Natural Resources Evaluation, Applied Geostatistics Series, Oxford University Press, New York, ISBN 9780195115383, 1997.

Gravey, M. and Mariethoz, G.: QuickSampling v1.0: a robust and simplified pixel-based multiple-point simulation approach, Geoscientific Model Development, 13, 2611–2630, https://doi.org/10.5194/gmd-13-2611-2020, 2020.

Gravey, M., Rasera, L. G., and Mariethoz, G.: Analogue-based colorization of remote sensing images using textural information, ISPRS Journal of Photogrammetry and Remote Sensing, 147, 242–254, https://doi.org/10.1016/j.isprsjprs.2018.11.003, 2019.

Hadjipetrou, S., Mariethoz, G., and Kyriakidis, P.: Gap-Filling Sentinel-1 Offshore Wind Speed Image Time Series Using Multiple-Point Geostatistical Simulation and Reanalysis Data, Remote Sensing, 15, 409, https://doi.org/10.3390/rs15020409, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., De Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hrachowitz, M. and Clark, M. P.: The complementary merits of competing modelling philosophies in hydrology, Hydrology and Earth System Sciences, 21, https://doi.org/10.5194/hess-21-3953-2017, 2017.

Jha, S. K., Mariethoz, G., Evans, J., McCabe, M. F., and Sharma, A.: A space and time scale-dependent nonlinear geostatistical approach for downscaling daily precipitation and temperature, Water Resources Research, 51, 6244–6261, https://doi.org/10.1002/2014WR016729, 2015.

Journel, A. and Zhang, T.: The necessity of a multiple-point prior model, Mathematical Geology, 38, 591–610, https://doi.org/10.1007/s11004-006-9031-2, 2006.

Kessler, T. C., Comunian, A., Oriani, F., Renard, P., Nilsson, B., Klint, K. E., and Bjerg, P. L.: Modeling Fine-Scale Geological Heterogeneity—Examples of Sand Lenses in Tills, Groundwater, 51, 692–705, https://doi.org/10.1111/j.1745-6584.2012.01015.x, 2013.

La Confederación Hidrográfica del Ebro: La cuenca del Ebro, https://www.chebro.es/en-GB/web/guest/la-cuenca-del-ebro, 2026.

López-Moreno, J. I., Vicente-Serrano, S. M., Morán-Tejeda, E., Zabalza, J., Lorenzo-Lacruz, J., and García-Ruiz, J. M.: Impact of climate evolution and land use changes on water yield in the Ebro basin, Hydrology and Earth System Sciences, 15, 311–322, https://doi.org/10.5194/hess-15-311-2011, 2011.

Lorenz, E. N.: Atmospheric Predictability as Revealed by Naturally Occurring Analogues, Journal of the Atmospheric Sciences, 26, https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2, 1969.

Mariethoz, G. and Caers, J.: Multiple-point Geostatistics: Stochastic Modeling with Training Images, vol. 9781118662755 of *Multiple-point Geostatistics: Stochastic Modeling with Training Images*, Wiley-Blackwell, 2014.

Mariethoz, G., Renard, P., and Straubhaar, J.: The Direct Sampling Method to Perform Multiple-Point Geostatistical Simulations, Water Resources Research, 46, W11 536, https://doi.org/10.1029/2008WR007621, 2010.

Miller, L., Pelletier, C., and Webb, G. I.: Deep Learning for Satellite Image Time-Series Analysis: A Review, IEEE Geoscience and Remote Sensing Magazine, 12, 81–124, https://doi.org/10.1109/MGRS.2024.3393010, 2024.

Mo, Y., Xu, Y., Liu, Y., Xin, Y., and Zhu, S.: Comparison of gap-filling methods for producing all-weather daily remotely sensed near-surface air temperature, Remote Sensing of Environment, 296, 113 732, https://doi.org/10.1016/j.rse.2023.113732, 2023.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth System Science Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Noyelle, R., Zhang, Y., Yiou, P., and Faranda, D.: Maximal reachable temperatures for Western Europe in current climate, Environmental Research Letters, 18, https://doi.org/10.1088/1748-9326/acf679, 2023.

Oriani, F., Ohana-Levi, N., Marra, F., Straubhaar, J., Mariethoz, G., Renard, P., Karnieli, A., and Morin, E.: Simulating Small-Scale Rainfall Fields Conditioned by Weather State and Elevation: A Data-Driven Approach Based on Rainfall Radar Images, Water Resources Research, 53, 8512–8532, https://doi.org/10.1002/2017WR020876, 2017.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Pirot, G., Straubhaar, J., and Renard, P.: Simulation of braided river elevation model time series with multiple-point statistics, Geomorphology, 214, 148–156, https://doi.org/10.1016/j.geomorph.2014.01.022, 2014.

Running, S., Mu, Q., Zhao, M., and Moreno, A.: MODIS/Terra Net Evapotranspiration Gap-Filled 8-Day L4 Global 500m SIN Grid V061, https://doi.org/10.5067/MODIS/MOD16A2GF.061, 2021.

Schmutz, L., Thao, S., Vrac, M., Allard, D., and Mariethoz, G.: Improved climate projections by combining CMIP6 models according to their local multivariate performance, https://doi.org/10.5194/egusphere-egu25-18233, 2025.

Siabi, N., Sanaeinejad, S. H., and Ghahraman, B.: Effective method for filling gaps in time series of environmental remote sensing data: An example on evapotranspiration and land surface temperature images, Computers and Electronics in Agriculture, 193, 106 619, https://doi.org/10.1016/j.compag.2021.106619, 2022.

Strebelle, S.: Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics, Mathematical Geology, 34, 1–21, https://doi.org/10.1023/A:1014009426274, 2002.

Thao, S., Garvik, M., Mariethoz, G., and Vrac, M.: Combining global climate models using graph cuts, Climate Dynamics, 59, 2345–2361, https://doi.org/10.1007/s00382-022-06213-4, 2022.

Vautard, R. and Yiou, P.: Control of recent European surface climate change by atmospheric flow, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL040480, 2009.

Vrac, M., Allard, D., Mariéthoz, G., Thao, S., and Schmutz, L.: Distribution-based pooling for combination and multi-model bias correction of climate simulations, Earth System Dynamics, 15, 735–762, https://doi.org/10.5194/esd-15-735-2024, 2024.

Wang, L. and Zhang, Y.: Filling GRACE data gap using an innovative transformer-based deep learning approach, Remote Sensing of Environment, 315, 114 465, https://doi.org/10.1016/j.rse.2024.114465, 2024.

Wang, Q., Wang, L., Zhu, X., Ge, Y., Tong, X., and Atkinson, P. M.: Remote sensing image gap filling based on spatial-spectral random forests, Science of Remote Sensing, 5, 100 048, https://doi.org/10.1016/j.srs.2022.100048, 2022.

Wang, Q., Tang, Y., Ge, Y., Xie, H., Tong, X., and Atkinson, P. M.: A comprehensive review of spatial-temporal-spectral information reconstruction techniques, Science of Remote Sensing, 8, 100 102, https://doi.org/10.1016/j.srs.2023.100102, 2023.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E.: Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing, 13, 600–612, https://doi.org/10.1109/TIP.2003.819861, 2004.

Wu, J., Li, T., Lin, L., and Zeng, C.: Progressive gap-filling in optical remote sensing imagery through a cascade of temporal and spatial reconstruction models, Remote Sensing of Environment, 311, 114 245, https://doi.org/10.1016/j.rse.2024.114245, 2024.

Xu, S., Qin, T., Lu, J., Liu, S., Hou, J., Feng, J., Li, W., Liu, H., and Abebe, S. A.: Identification of driving mechanisms of actual evapotranspiration in the Yiluo River Basin based on structural equation modeling, Ecological Processes, 13, 69, https://doi.org/10.1186/s13717-024-00551-3, 2024.

Yin, G., Mariethoz, G., and McCabe, M.: Gap-Filling of Landsat 7 Imagery Using the Direct Sampling Method, Remote Sensing, 9, 12, https://doi.org/10.3390/rs9010012, 2016.

Yiou, P.: AnaWEGE: a weather generator based on analogues of atmospheric circulation, Geoscientific Model Development, 7, https://doi.org/10.5194/gmd-7-531-2014, 2014.

https://doi.org/10.5194/egusphere-2026-4283
Preprint. Discussion started: 7 September 2026

Yiou, P. and Jézéquel, A.: Simulation of extreme heat waves with empirical importance sampling, Geoscientific Model Development, 13, https://doi.org/10.5194/gmd-13-763-2020, 2020.

Yiou, P., Salameh, T., Drobinski, P., Menut, L., Vautard, R., and Vrac, M.: Ensemble reconstruction of the atmospheric column from surface pressure using analogues, Climate Dynamics, 41, https://doi.org/10.1007/s00382-012-1626-3, 2013.

Yule, E. L., Hegerl, G. C., Schurer, A., Ballinger, A., and Hawkins, E.: Using analogues to predict changes in future UK heatwaves*, Environmental Research: Climate, 3, https://doi.org/10.1088/2752-5295/ad57e3, 2024.

Zakeri, F. and Mariethoz, G.: Synthesizing long-term satellite imagery consistent with climate data: Application to daily snow cover, Remote Sensing of Environment, 300, https://doi.org/10.1016/j.rse.2023.113877, 2024.

Zhang, C., Zeng, J., Shi, P., Ma, H., Letu, H., Zhang, X., Wang, P., Bi, H., and Rong, J.: Global-scale gap filling of satellite soil moisture products: Methods and validation, Journal of Hydrology, 653, 132 762, https://doi.org/10.1016/j.jhydrol.2025.132762, 2025.

Zhang, T., Zhou, Y., Zhao, K., Zhu, Z., Asrar, G. R., and Zhao, X.: Gap-filling MODIS daily aerosol optical depth products by developing a spatiotemporal fitting algorithm, GIScience & Remote Sensing, 59, 762–781, https://doi.org/10.1080/15481603.2022.2060596, 2022.