Titre: Title:	Spatial pattern regression for meteorological fields interpolation
Auteurs: Authors:	Vihotogbé Houssou, & Julie Carreau
Date:	2026
Type:	Article de revue / Article
Référence: Citation:	Houssou, V., & Carreau, J. (2026). Spatial pattern regression for meteorological fields interpolation. Hydrology and Earth System Sciences, 30(18), 5791-5807. https://doi.org/10.5194/hess-30-5791-2026

Hydrol. Earth Syst. Sci., 30, 5791–5807, 2026
https://doi.org/10.5194/hess-30-5791-2026

Hydrology and Earth System Sciences
Open Access

Spatial pattern regression for meteorological fields interpolation

Vihotogbé Houssou[1,2,4] **and Julie Carreau**[1,2,3,4]

[1]Department of Mathematics and Industrial Engineering, Polytechnique Montréal, 2500 chemin de Polytechnique, Montréal, H3T 1J4, Québec, Canada
[2]GERAD – Groupe d'Études et de Recherche en Analyse des Décisions, 2920 Chemin de la Tour, Montréal, H3T 1N8, Québec, Canada
[3]Mila – Quebec Artificial Intelligence Institute, 6666 Saint-Urbain, Montréal, H2S 3H1, Québec, Canada
[4]IVADO – Institute for Data Valorization, 950 Av. Beaumont, Montréal, H3N 1V5, Québec, Canada

Correspondence: Vihotogbé Houssou (vihotogbe-2.houssou@polymtl.ca)

Received: 26 March 2026 – Discussion started: 17 April 2026
Revised: 27 August 2026 – Accepted: 1 September 2026 – Published: 15 September 2026

Abstract. High-resolution gridded meteorological data are essential for hydrological impact studies, yet their reconstruction from sparse station networks remains challenging. We introduce Spatial Pattern Regression (SPR), a data-driven method that reconstructs gridded meteorological fields by combining spatial information extracted from high-resolution regional climate model (RCM) simulations with station observations. SPR operates in two steps: spatial patterns are first extracted from RCM data using principal component analysis, then daily fields are reconstructed through linear regression using available observations. The method is first evaluated using controlled synthetic experiments, where virtual stations selected as a subset of the RCM grid emulate observational networks with varying density, size, and location. SPR is then validated using real station observations. Daily precipitation, minimum temperature, and maximum temperature are considered. Results show that SPR performs better than inverse distance weighting, ordinary kriging, and kriging with external drift, particularly under sparse network conditions. Sensitivity analyses highlight the dominant role of station density and location on interpolation accuracy, supporting the robustness and applicability of SPR for hydrological studies.

1 Introduction

High-resolution gridded meteorological data play a central role in hydrological impact studies, particularly in the context of climate change, where the frequency and intensity of extreme events such as heatwaves, heavy precipitation, and flooding are increasing (Warren et al., 2022). Hydrological models used for flood forecasting, water resource management, or climate impact assessments require meteorological inputs that accurately represent the fine spatio-temporal variability of meteorological fields, especially during extreme events. For example, Lucas-Picher et al. (2020) successfully reproduced the extreme flooding of the Richelieu River (southern Quebec, Canada) in spring 2011 using a high-resolution gridded hydrometeorological dataset at approximately 7 km resolution, which accounted for orographic precipitation effects (Livneh et al., 2015). Similarly, high-resolution meteorological fields have been shown to be critical for assessing heat-related impacts in urban environments (Lauer et al., 2023).

In practice, gridded meteorological data used in hydrological studies are typically obtained through four main approaches: spatial interpolation of station observations (Cornes et al., 2018), physics-based numerical models (including numerical weather prediction and regional climate models) (Muñoz-Sabater et al., 2021), reanalysis products (Leduc et al., 2019; Gasset et al., 2021), and radar-based products (Vernay et al., 2025). Physics-based models, such as numerical weather prediction and regional climate models (RCMs), provide spatially complete and physically con-

Published by Copernicus Publications on behalf of the European Geosciences Union.

sistent meteorological fields at increasingly fine resolutions. These datasets are widely used in hydrological impact studies, either directly or after bias correction (e.g., Faghih and Brissette, 2023). However, raw model outputs may exhibit systematic biases and do not necessarily reproduce observed weather conditions at specific locations or times.

Spatial interpolation methods rely on statistical techniques to estimate meteorological variables at ungauged locations based on nearby observations. While widely used, their performance is strongly sensitive to station density and spatial variability, which are often limited, particularly in mountainous or sparsely instrumented regions (Ly et al., 2013). Under such conditions, spatial interpolation methods often struggle to capture realistic spatial variability, which can propagate errors into hydrological simulations (Hwang et al., 2012). This limitation becomes particularly critical during extreme rainfall events, where forecast accuracy strongly depends on network density (Segond et al., 2007; Looper and Vieux, 2012). To mitigate these limitations, auxiliary gridded information such as topography or climatology is commonly incorporated (Livneh et al., 2015; Werner et al., 2019; Daly et al., 1997). However, the use of climatology, often derived from RCM, is a limited way of exploiting the rich spatial structure present in RCM simulations. More advanced approaches seek to exploit higher-order spatial structures beyond mean climatology, capturing coherent patterns of variability across space (Taylor et al., 2013; Carreau and Guinot, 2021).

The main contribution of this paper is the development of an interpolation method that exploits the spatial structure of RCM simulations through spatial patterns, which succinctly summarize the spatial structures present in these simulations. Specifically, we introduce Spatial Pattern Regression (SPR), a physically-constrained spatial reconstruction framework designed to bridge the gap between traditional statistical interpolation and model-based reanalyses. Unlike reanalysis systems, which assimilate observations into dynamically evolving model states, SPR relies on a fixed set of spatial patterns extracted from the RCM, whose daily amplitudes are estimated from sparse station observations. Unlike purely local interpolation methods, SPR leverages physically consistent spatial structures derived from RCM simulations to constrain the reconstruction. SPR assumes that high-resolution regional climate model simulations adequately represent the dominant spatial structures of meteorological variables. These spatial structures are first extracted from RCM data using principal component analysis (PCA), yielding spatial patterns that represent the principal modes of spatial organization of the variable. When handling the high dimensionality of gridded climate data, PCA (equivalent to Empirical Orthogonal Function – EOF) provides a well-established framework for identifying a limited number of dominant spatial patterns that explain most of the variability and that can be exploited in regression-type reconstructions (Monahan et al., 2009). Daily meteorological fields are then reconstructed through linear regression using observations from available stations, allowing spatial coherence to be imposed even under sparse network conditions.

The use of dominant modes of variability extracted from covariance structures has been explored as a means to reconstruct spatially continuous fields when observations are sparse (Taylor et al., 2013). Reduced-space approaches based on EOF have also been used in meteorological reconstruction and data assimilation, notably in the Reduced Space Optimal Interpolation (RSOI) framework proposed by Schiemann et al. (2010). In RSOI, spatial variability is represented in a reduced EOF space, and observations are combined with a background field through an optimal interpolation formulation relying on prescribed error covariance structures. While effective, such approaches typically require historical observational archives to estimate these covariances and often rely on gridded datasets derived from prior interpolation of station data to define the reduced space. Furthermore, since these observation-based gridded fields only exist for the historical past, the spatial patterns learned by RSOI are related to past climate conditions and cannot be updated to reflect future changes in spatial structure. In contrast, the approach proposed in this work builds upon the idea of reduced-space reconstruction but adopts a regression-based formulation that does not require explicit error covariance modelling or historical observations. Instead, spatial patterns are extracted directly from regional climate model simulations, which span both historical and future periods, allowing the auxiliary period to be chosen freely. This makes SPR adaptable to changing climate conditions and applicable in data-sparse or ungauged regions.

SPR is evaluated against commonly used baseline interpolation methods. Performance is assessed using controlled synthetic experiments which enable systematic evaluations that would not be feasible using observations alone. In addition to these controlled synthetic experiments, an evaluation using real station observations is performed, thereby assessing both methodological robustness and practical applicability. Daily precipitation, minimum temperature, and maximum temperature are considered, as these variables are central to hydrological modeling and climate impact studies.

The remainder of this paper is organized as follows. Section 2 presents the study area and the datasets used, including regional climate model simulations and station observations. Section 3 describes the proposed Spatial Pattern Regression (SPR) methodology, the baseline interpolation methods, and the experimental design. Section 4 presents the results obtained from both synthetic experiments and the independent validation using real observations. Finally, Sect. 5 discusses the implications of the results for hydrological applications, as well as the limitations of the proposed approach. It also summarizes the main findings and outlines perspectives for future work.

2 Study area and Data

To evaluate and compare the interpolation methods, we considered daily precipitation (mm d^{-1}) and minimum and maximum temperature (°C). These meteorological variables are commonly used as inputs in hydrological modeling and climate impact studies.

2.1 Regional climate model data

High-resolution gridded meteorological data were obtained from the ClimEx project, which investigates the impacts of climate change on extreme meteorological and hydrological events using the Canadian RCM (Leduc et al., 2019). The ClimEx simulations cover a North American domain discretized on a regular grid of 280×280 cells, with a horizontal resolution of approximately 11 km (see Fig. 1).

From the full ClimEx simulation period (1950–2099), two distinct periods were defined. An auxiliary period (1980–2009) was used to extract spatial information, including climatologies and spatial patterns. An interpolation period (2000–2009) was used to perform spatial reconstructions and to evaluate method performance (see Sect. 3.4.1). Although the auxiliary and interpolation periods overlap in this study, the proposed methodology does not require temporal overlap or synchronicity between these periods, allowing flexibility in practical applications. It should be noted, however, that for temperature variables, long-term warming trends may introduce systematic biases when the auxiliary and interpolation periods are substantially separated in time, as the spatial mean field extracted from the auxiliary period may not accurately represent the mean level of the interpolation period. In such cases, detrending of the RCM fields prior to pattern extraction, or the selection of an auxiliary period temporally close to the interpolation period, would be advisable.

Within the North American ClimEx domain, two study regions were defined: a southern region and a northern region, reflecting contrasting climatic and spatial variability conditions. For each region, three nested spatial extents (large, medium, and small) were considered to assess the sensitivity of interpolation performance to region size (see Fig. 1).

2.2 Observational data

Instrumented meteorological observations were obtained from the official monitoring network operated by Environment and Climate Change Canada (ECCC), which provides long-term, quality-controlled climate observations (https://climate.weather.gc.ca/climate_data/bulk_data_e.html, last access: 30 January 2026). The analysis is restricted to stations located in southern Quebec, within a spatial domain of approximately 70 000 km^2, defined by latitudes 45.0–47.0° N and longitudes 75.0–70.0° W.

Stations were selected to ensure strictly continuous observation periods, with no missing daily values and no gap filling. For each variable, a two-year period with 100 % data completeness was retained in order to balance temporal representativeness and station availability. Different periods were selected across variables to maximize the number of available complete stations (Precipitation: 8 October 2000–7 October 2002; Minimum temperature: 7 December 2000–6 December 2002; Maximum temperature: 12 December 2016–11 December 2018).

The final dataset consists of 15 stations for precipitation and 44 stations for minimum and maximum temperature, each providing complete daily records over the corresponding two-year period. These observations are used exclusively for independent validation and do not contribute to the extraction of spatial patterns, ensuring a strict separation between auxiliary model information and real-world observations.

3 Methods

3.1 From Climatology to Spatial Patterns

Climatologies, often used as auxiliary information in interpolation methods, are broadly defined as inter-annual averages of a given meteorological variable. When applied to RCM gridded data, this averaging process reveals little about the underlying spatial structure inherent to RCM simulations. In this study, we extract richer information from RCM simulations by characterizing their spatial patterns – a concept related to weather types or regimes frequently used in climate science, identified through clustering in empirical orthogonal function (EOF) space (Cattiaux et al., 2010).

More specifically, we define a spatial pattern as a spatially coherent structure that captures a recurrent mode of variability of a meteorological field over a given region. Such patterns describe how values tend to co-vary in space, independently of their instantaneous amplitude. Unlike climatologies, which represent long-term mean states, spatial patterns retain detailed information on spatial variability. Furthermore, unlike local interpolation weights or distance-based kernels, they provide a global representation of spatial organization that remains meaningful even in areas with no direct observations.

Spatial patterns play a central role in the proposed SPR method by serving as a low-dimensional basis onto which daily meteorological data can be projected. Indeed, in practice, we compute the spatial patterns as the EOFs obtained through principal component analysis (PCA) of gridded RCM simulations. Each pattern corresponds to an eigenvector of the spatial covariance structure and represents an orthogonal mode of variability ordered by explained variance. Importantly, these patterns are extracted independently of the station observations used for reconstruction, ensuring that the spatial structures are not conditioned by the observational network geometry.

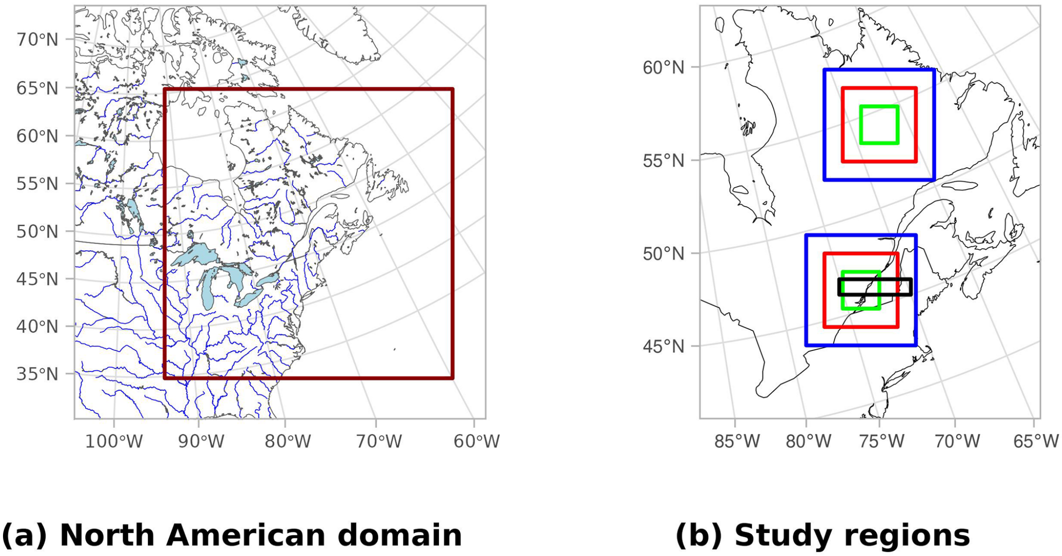

Figure 1. Spatial domains used in this study. Panel **(a)** shows the North American domain of the ClimEx project, with the study domain outlined by the dark red rectangle. Panel **(b)** shows the two study regions, referred to as the south and north regions, at three different spatial scales. The black rectangle indicates the spatial domain used for the real-data experiments.

Unlike external covariates or drift terms used in some interpolation methods, spatial patterns encode intrinsic spatial variability derived directly from the meteorological field itself (Monahan et al., 2009; Carreau and Guinot, 2021). Their relevance depends on the variable and regional context, with complex fields such as precipitation typically requiring more patterns than temperature. As such, spatial patterns do not act as predictors of local magnitude but as structural building blocks that constrain the shape of the reconstructed field.

The extracted spatial patterns (see Fig. 2) are illustrated in the following to highlight their physical interpretability and role in the reconstruction framework. The PCA is computed using all daily fields over the auxiliary period, allowing the extracted spatial patterns to represent the dominant modes of spatial variability across a broad range of meteorological situations.

3.2 Spatial Pattern Regression (SPR)

SPR is the main methodological contribution of this paper. It builds upon reduced-space representations of spatial variability and is based on the projection – performed via linear regression – of sparse station observations onto spatial patterns derived from RCM simulations. As SPR reconstructs spatially continuous meteorological fields from such observations, it can be regarded as an interpolation method, although its underlying mechanisms differ from those of classical spatial interpolation approaches. SPR is conceptually related to reduced-space methods developed in data assimilation, notably the Reduced Space Optimal Interpolation (RSOI) framework of Schiemann et al. (2010), which also relies on EOF-based representations of spatial variability. However, SPR adopts a regression-based formulation in which spatial pattern amplitudes are estimated independently at each time step, without requiring explicit modelling of background and observation error covariances. In addition, the spatial patterns used in SPR are derived exclusively from RCM simulations, rather than from gridded observational analyses, allowing the method to rely on physically consistent spatial structures. It makes no isotropy or stationarity assumptions and relies on the representativeness of the dominant spatial patterns. It consists of two main steps, each described in turn below.

3.2.1 Identification of a representative basis of spatial patterns

The first step of SPR consists in extracting a representative basis of spatial patterns from high-resolution RCM simulations. These patterns are identified using Principal Component Analysis (PCA), computed via Singular Value Decomposition (SVD). PCA provides an efficient way to extract a reduced set of orthogonal spatial modes that capture the dominant spatial variability of gridded meteorological fields (Link et al., 2019).

Let $\mathbf{Z}^{\mathrm{grid}}_{n,p}$ denote the RCM dataset over the auxiliary period, where n is the number of daily time steps and p is the number of grid cells in the study region. Applying SVD to the centered data matrix yields

$$\mathbf{Z}^{\mathrm{grid}}_{n,p} = \mathbf{U}^{\mathrm{grid}}_{n,k}\mathbf{S}^{\mathrm{grid}}_{k,k}(\mathbf{V}^{\mathrm{grid}}_{p,k})^T + (\mathbf{Z}^{\mathrm{grid}}_{p}\,1^T_n)^T, \quad (1)$$

where $\mathbf{V}^{\mathrm{grid}}_{p,k}$ contains the first k spatial patterns, $\mathbf{S}^{\mathrm{grid}}_{k,k}$ is the diagonal matrix of singular values, and $\mathbf{U}^{\mathrm{grid}}_{n,k}$ contains the asso-

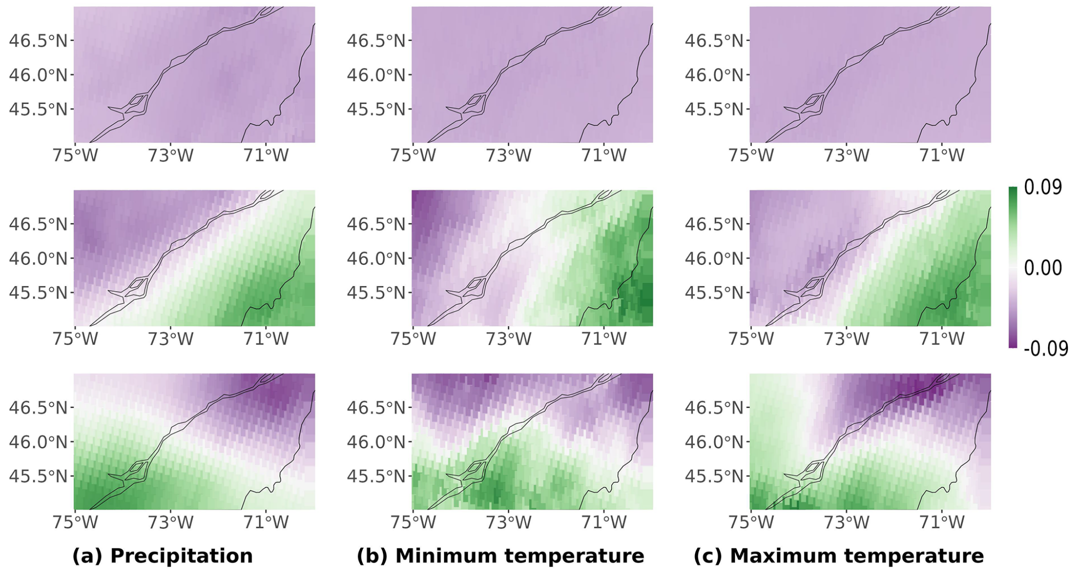

Figure 2. First three spatial patterns for each variable in the southern region (577 grid cells; see Fig. 1), ordered from top to bottom by decreasing explained variance. The first, second, and third columns show precipitation, minimum temperature, and maximum temperature, respectively. Spatial patterns are dimensionless.

ciated temporal coefficients. The vector $\boldsymbol{Z}_p^{\mathrm{grid}}$ represents the spatial mean field, T denotes the transpose operation, 1_n is a vector of ones of length n. More precisely, $\boldsymbol{Z}_p^{\mathrm{grid}}$ is the temporal column mean of the RCM data matrix over the auxiliary period, defined for each grid cell j as $Z_j^{\mathrm{grid}} = \frac{1}{n}\sum_{t=1}^{n} Z_{t,j}^{\mathrm{grid}}$. It is a single constant per grid cell, computed once over the auxiliary period, and represents the standard centering term of the SVD decomposition. It is distinct from monthly climatological means or seasonally varying backgrounds.

The spatial patterns $\mathbf{V}_{p,k}^{\mathrm{grid}}$ form a fixed, spatially complete basis that characterizes the dominant spatial structures of the variable. They are extracted once from the auxiliary period and do not vary from day to day. These patterns are assumed to be representative of the spatial variability encountered during the interpolation period and are subsequently used to reconstruct daily fields from sparse observations.

3.2.2 Spatio-temporal integration through regression

The second step of SPR estimates the temporal amplitudes of the spatial patterns by fitting a regression model to station observations for each day of the interpolation period. While spatial patterns are extracted from RCM simulations over the auxiliary period, station observations are only used at this regression step. Importantly, the station observations used here do not need to be temporally aligned with the auxiliary period used for pattern extraction – the two periods are entirely independent, which gives the method considerable flexibility in practical applications.

Let $\mathbf{Z}_d = \big(Z(s_1), \ldots, Z(s_d)\big)^T$ denote the vector of observed values at d gauged stations for a given day. To relate these observations to the spatial patterns, the full pattern matrix $\mathbf{V}_{p,k}^{\mathrm{grid}}$ is spatially restricted to the station locations, yielding the reduced matrix $\mathbf{V}_{d,k}^{\mathrm{grid}}$, $d < p$.

SPR assumes the following linear regression model:

$$\mathbf{Z}_d - \mathbf{Z}_d^{\mathrm{grid}} = \mathbf{V}_{d,k}^{\mathrm{grid}}\,\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_d, \quad (2)$$

where β_k is the vector of pattern coefficients for the day considered, $\mathbf{Z}_d^{\mathrm{grid}}$ is the mean RCM field restricted to the station locations, and $\boldsymbol{\varepsilon}_d$ is a vector of residual errors assumed to be independent and identically distributed with zero mean. The regression coefficients $\boldsymbol{\beta}_k$ are estimated using ordinary least squares independently for each day. Specifically, the coefficients are obtained by minimizing the squared difference between observed values and their reconstruction from the spatial patterns restricted to station locations.

This formulation is directly inspired by the PCA decomposition of the gridded fields (see Eq. 1). It allows the spatial structure of the field to be prescribed a priori, while the temporal evolution is constrained by the available observations. The estimated coefficients $\hat{\boldsymbol{\beta}}_k$ thus play the role of daily temporal scores, describing how the amplitude of each spatial pattern varies from one day to the next. In the idealized case where observations are available at all grid points, the coefficients $\boldsymbol{\beta}_k$ correspond to the PCA temporal scores. In the sparse-observation setting, they are estimated by least squares, ensuring consistency between observed values and the spatial structures imposed by the patterns.

The formulation in Eq. (2) can be equivalently rewritten in its uncentered form as:

$$\mathbf{Z}_d = \mathbf{Z}_d^{\mathrm{grid}} + \mathbf{V}_{d,k}^{\mathrm{grid}}\,\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_d, \quad (3)$$

which highlights that SPR can be interpreted as a linear model with a spatially varying intercept term given by the RCM-derived mean field, and a low-rank representation of spatial variability defined by the PCA-derived patterns. The centered formulation used in Eq. (2) is therefore a direct consequence of the PCA decomposition, in which spatial patterns represent variability around the mean field. In particular, this uncentered form makes clear that SPR is applied to raw observations Z_d, and that its output is on the same scale as the observations used by IDW, OK, and KED, ensuring a consistent basis for comparison.

The regression is performed independently for each day, yielding a sequence of coefficient vectors that capture the temporal evolution of the spatial patterns. The complete gridded field is then reconstructed using the full pattern matrix, as opposed to the restricted one in Eq. (2):

$$\hat{\boldsymbol{Z}}_p = \mathbf{V}^{\text{grid}}_{p,k}\,\hat{\boldsymbol{\beta}}_k + \boldsymbol{Z}^{\text{grid}}_p. \tag{4}$$

This reconstruction preserves the large-scale spatial coherence provided by the RCM-derived patterns while ensuring consistency with station observations at each time step. Rather than interpolating values directly in physical space, the method reconstructs fields as linear combinations of predefined spatial patterns, whose associated coefficients vary in time.

The method does not require a fixed station network over time, as the regression is performed independently for each day using the available observations. For a given day, the spatial pattern matrix is restricted to the station locations, and its dimension therefore adapts to the number of available observations.

3.3 Baseline interpolation methods

Three widely used spatial interpolation methods are employed as baselines against which the performance of SPR is compared. These methods reconstruct values at ungauged locations using weighted combinations of observations from neighbouring stations and are representative of both deterministic and geostatistical approaches commonly applied in hydrology and climatology (Li and Heap, 2014; Bokke, 2017).

Inverse Distance Weighting (IDW) is a deterministic interpolation method in which observations closer to the target location receive larger weights than more distant ones (Burhanuddin et al., 2015; Margaritidis, 2024; Li and Heap, 2011, 2014; Zimmerman et al., 1999). The method assumes that spatial proximity alone governs similarity, without explicitly modeling spatial correlation or uncertainty. Despite its simplicity, IDW is frequently used as a benchmark due to its low computational cost and minimal assumptions.

Ordinary Kriging (OK) is a geostatistical method that exploits the spatial autocorrelation structure of the variable through a variogram model (Snepvangers et al., 2003). Predictions are optimal in the sense of being unbiased and of minimum variance under the assumption of second-order stationarity and isotropy. In addition to point estimates, OK provides an estimate of prediction uncertainty, which makes it a common reference method in spatial interpolation studies.

Kriging with External Drift (KED) extends OK by incorporating an auxiliary variable to represent a spatial trend in the mean structure of the field (Varentsov et al., 2020; Hengl et al., 2003). In this study, the auxiliary information corresponds to monthly climatologies derived from RCM simulations on the auxiliary period. While KED allows part of the large-scale spatial variability to be explained by the external drift, it still relies on station-based variogram modeling and assumes that residuals are stationary.

These baseline methods fundamentally differ from SPR in how spatial information is exploited. IDW, OK, and KED operate directly in the geographical space using distances and local neighbourhoods, whereas SPR relies on a low-dimensional basis of spatial patterns extracted from high-resolution gridded data. By embedding physically consistent spatial structures into the interpolation process, SPR departs from local station-centred approaches and provides a complementary framework for reconstructing spatially coherent meteorological fields.

3.4 Experimental design and validation framework

3.4.1 Synthetic data experiments

These experiments rely on an idealized setting where a virtual station network is defined as a sub-sample of the RCM grid cells. By emulating observational networks of varying sizes and spatial configurations, the virtual stations yield pseudo-observations that are used in the same way as real station data in the regression step, allowing for a controlled evaluation of the method. For each of the two selected regions (see Fig. 1), we vary two key factors – the spatial extent of the region and the density of the virtual station network – to design a set of controlled experiments. Interpolation is then performed at the remaining grid cells, and the reconstructed fields are compared to the reference RCM fields, allowing a direct and controlled assessment of each method's sensitivity to network sparsity and spatial scale. The reference RCM field serves as ground truth in the synthetic experiments by design: the objective is to evaluate how well each method can reconstruct a known spatially complete field from sparse observations, rather than to assess realism with respect to actual meteorological observations.

The region size can take one of three values (large, medium, or small) corresponding to the nested rectangles shown in Fig. 1. The network density is defined as the proportion of grid cells retained as virtual stations and is set to 10 %, 30 %, 50 %, 70 %, or 90 %. For each variable, this design implies 30 distinct experiments, combining region location, spatial extent, and network density (see Table 1).

In addition to these experiments, a stress-test experiment is designed to mimic a common practical situation in spatial interpolation: extremely sparse observation networks, as often encountered in remote regions. We focus on a large northern Quebec region (Fig. 1) composed of 2970 grid cells, from which only three grid cells (approximately 0.1 %) are randomly selected as virtual stations.

3.4.2 Real data experiments

In these experiments, SPR is applied as intended for real-world data: station observations are provided by a real station network (see Sect. 2.2) while RCM simulations over the auxiliary period are used to construct the spatial patterns. Interpolation is then carried out over the interpolation period, day by day and for each meteorological variable independently, at grid cells where no station observations are available.

We focused on a single factor: the density of the station network. Specifically, we considered only 10 % and 30 % densities, representing data scarcity situations and corresponding to training sets of 2 to 4 stations over a 70 000 km^2 region. To assess the sensitivity of results to the specific choice of training stations, the station selection is repeated 100 times using independent random seeds for each density level and each variable. Performance metrics are averaged across repetitions, and their standard deviation is reported to quantify sampling uncertainty. This repeated subsampling strategy ensures that the conclusions are not driven by a particular station configuration and provides a statistically robust basis for comparison across methods.

3.5 Model evaluation

For each variable, spatial interpolation is performed independently for each day, and performance metrics (see Sect. 3.5.2) are aggregated over the interpolation period. For precipitation, a softplus transformation $f(x) = \log(\exp(x) - 1)$ is applied prior to interpolation to ensure positivity after bounding values by 10^{-5}. Interpolated values are then back-transformed to the original scale, and values below 0.5 mm are set to zero following Werner et al. (2019). The same softplus transformation and direct back-transformation procedure was applied uniformly to all interpolation methods (SPR, KED, OK, and IDW) to ensure fair comparison.

It should be noted that applying a direct inverse transformation after interpolation in transformed space may in general neglect the contribution of analysis-error variance (Fletcher and Zupanski, 2006). However, the inverse of the softplus transformation satisfies $f^{-1}(x) = \log(1 + e^x) \approx x$ for precipitation values of practical interest, implying that the back-transformation introduces negligible bias in this context. This transformation serves as a smooth link function ensuring that back-transformed interpolated values remain nonnegative. Unlike variance-stabilizing transformations such as the square-root (Schiemann et al., 2010; Erdin et al., 2012), the softplus transformation does not primarily aim to reduce skewness but rather to enforce the physical boundary of nonnegativity. Alternative transformations, such as the square-root transformation used by Schiemann et al. (2010), are also common for precipitation. Correction methods based on Taylor-series approximations (Fortin et al., 2015; van Hyfte et al., 2023) or quantile-based approaches (Erdin et al., 2012) are not considered here given the negligible bias of the softplus back-transformation, but represent a potential avenue for future improvements.

Since interpolation is conducted independently in time, model evaluation is formulated in the spatial domain. For synthetic experiments, virtual stations corresponding to a given network density define the training set, while the remaining grid cells form the test set. An additional validation set is obtained by randomly withholding the same proportion of grid cells from the training set, ensuring a consistent assessment of model generalization across network densities and region sizes. Validation set is used for hyperparameter selection, see Sect. 3.5.1.

For the real data experiments, a subset of stations corresponding to a given network density represent the training set while the remaining stations form the test set.

3.5.1 Hyperparameter selection

Hyperparameters are selected by maximizing average performance on the validation set over the full interpolation period. Although day-specific tuning is possible, a single globally optimal configuration is retained for each method to ensure robustness and comparability across days.

For the baseline methods (IDW, OK, and KED), the tuning parameters include the IDW distance power (1–5) and the variogram model (Gaussian, Spherical, or Exponential) for OK and KED. For SPR, the number of retained spatial patterns k is treated as a hyperparameter and selected as a percentage (10 %–90 %) of the maximum available patterns, in steps of 5 %, allowing a consistent search across regions of different sizes. The selected value of k is fixed once per experimental configuration – that is, for a given combination of region, region size, network density, and meteorological variable – and applied uniformly to all days of the interpolation period. This ensures that the reconstructed fields rely on a consistent spatial basis throughout, which is particularly important for climatological applications such as the estimation of long-term trends or return values, where consistency of the reconstruction framework across time steps is essential.

The optimal proportion of retained patterns varies across configurations and variables. For precipitation, it typically ranges from 20 %–60 %, with values most commonly around 40 %–50 % under sparse network conditions (10 %–30 % density). For temperature variables, the optimal proportion is generally lower, ranging from 20 %–50 % with most values around 25 %–35 %, reflecting the smoother and more

Table 1. Synthetic experiments framework: For each region location (south or north), three region sizes (S, M, or L) and five network density levels (from 90 % to 10 %) are considered. The number of grid cells where interpolation must be performed is given by: (100 − density) × total number of grid cells.

	Region South			Region North		
	Size L	Size M	Size S	Size L	Size M	Size S
Total # of grid cells	3025	1332	342	2970	1332	342
Network density	# of interpolated grid cells					
90 %	303	133	34	297	133	34
70 %	908	400	103	891	400	103
50 %	1523	666	171	1485	666	171
30 %	2118	932	239	2079	932	239
10 %	2723	1199	308	2673	1199	308

structured spatial variability of temperature fields compared to precipitation. These differences suggest that fewer spatial patterns are needed to capture the dominant spatial organization of temperature, while precipitation requires a richer basis to represent its more complex spatial structure.

Hyperparameter selection is performed exclusively within the synthetic experiment framework, where the ground truth is fully known and systematic validation is possible across a wide range of controlled configurations. Optimizing hyperparameters under very sparse network conditions is not feasible, as it would lead to unstable estimates and potential overfitting.

Consequently, for stress-test experiments, hyperparameters are fixed to the values obtained from synthetic experiments conducted at 10 % station density, which most closely matches the level of network sparsity considered in these tests. For the real station data experiments, hyperparameters are fixed to the values obtained from the synthetic experiments at the corresponding station densities (10 % and 30 %).

3.5.2 Performance evaluation metrics

Final model performance is evaluated on the test set using the optimal hyperparameters. Interpolated values are compared to the ground truth using two complementary metrics.

The Root Mean Squared Error (RMSE) quantifies pointwise interpolation accuracy and is computed daily over the test grid cells before being averaged over the interpolation period. In addition, the Structural Similarity Index Measure (SSIM) is used to assess the ability of each method to reproduce the spatial structure of the ground truth. SSIM values range from -1 to 1, with higher values indicating greater structural similarity.

While RMSE captures the average magnitude and bias of interpolation errors, SSIM explicitly accounts for spatial variance and structural correlation through its contrast and structure components respectively. Together, these two metrics provide a comprehensive evaluation covering the three aspects – bias, variability, and correlation – that are explicitly decomposed by metrics such as the Kling–Gupta Efficiency (Gupta et al., 2009).

4 Results

4.1 Results of synthetic data experiments

The performance of the three baseline interpolation methods and SPR is evaluated across all synthetic experiments using daily RMSE and SSIM, computed over the test grid cells and averaged over the interpolation period. Results are summarized in Fig. 3 for the southern region and Fig. 4 for the northern region. In these figures, RMSE is shown on the x-axis and $1-\mathrm{SSIM}$ on the y-axis, such that points closer to the origin indicate better performance.

Across the majority of experiments and for all three meteorological variables, SPR consistently achieves the best overall performance, combining low interpolation errors with a strong ability to preserve structural similarity. This behavior is particularly clear for medium and large regions in both the southern and northern domains, where SPR systematically outperforms all baseline methods.

At moderate to high station densities (50 % and above), SPR dominates the comparison regardless of region size or geographic location. At lower densities, performance differences mainly arise between SPR and KED, while OK and IDW consistently rank behind. In a limited number of low-density scenarios, primarily for precipitation and small regions, KED marginally outperforms SPR in terms of RMSE. In these cases, however, SPR generally maintains superior SSIM values, indicating a better reproduction of spatial patterns.

Overall, SPR demonstrates a favorable trade-off between pointwise accuracy and spatial coherence. Out of the 90 synthetic experiments considered, KED outperforms SPR in only six cases and yields comparable performance in three, while SPR remains superior in the vast majority of scenarios.

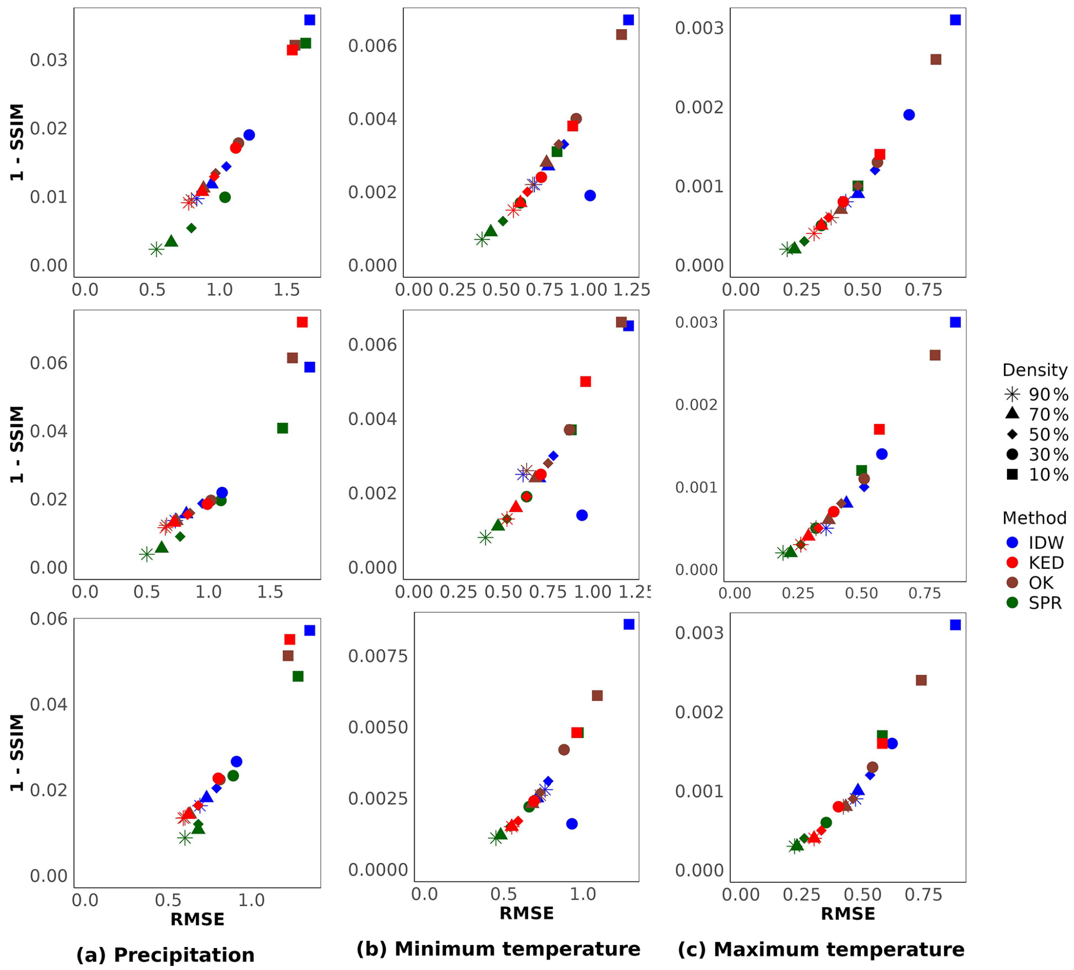

Figure 3. Synthetic data experiments: Comparison of the three baseline methods and SPR in the southern region, in terms of averaged RMSE (x-axis) and averaged 1-SSIM (y-axis). Each column represents a specific meteorological variable, while each row corresponds to a region size–ranging from the largest at the top to the smallest at the bottom. Each color represents a different interpolation method, and each plotting symbol corresponds to a specific network density. The closer a symbol is to the origin, the better the performance.

4.2 Results of synthetic stress-test experiments

Results (see details in Table 2) shows that SPR consistently yields lower average and lower median RMSE than KED for all three variables. For minimum and maximum temperature, SPR also achieves higher SSIM values, whereas for precipitation, KED shows a higher SSIM despite larger RMSE.

To further illustrate the differences between both methods, interpolated fields and corresponding ground truth are shown for three representative days for each variable (Fig. 5), along with the associated spatial RMSE fields (Fig. 6). KED fields are almost perfectly correlated with the climatological background (average Spearman correlation: 0.9997), while SPR fields exhibit weaker correlation (0.7910). However, SPR produces interpolated fields that are more structurally similar to the ground truth in terms of SSIM (average SSIM: 0.4151 for SPR versus 0.2504 for KED).

These results suggest that under extremely sparse observational constraints, SPR can better reproduce the spatial organization of the target fields, while KED remains strongly driven by the auxiliary climatological information used as external drift.

4.3 Results of real data experiments

Figure 7 summarizes the distribution of interpolation performance across 100 independent random station samplings, for each method, variable, and density level. The consistency of results across repetitions confirms that the conclusions are robust to the specific choice of training stations.

At the lowest density (10 %), SPR systematically achieves lower median RMSE than KED for precipitation and both temperature variables, while maintaining higher SSIM values. Notably, the interquartile ranges of SPR are more compact than those of KED, indicating that SPR is not only

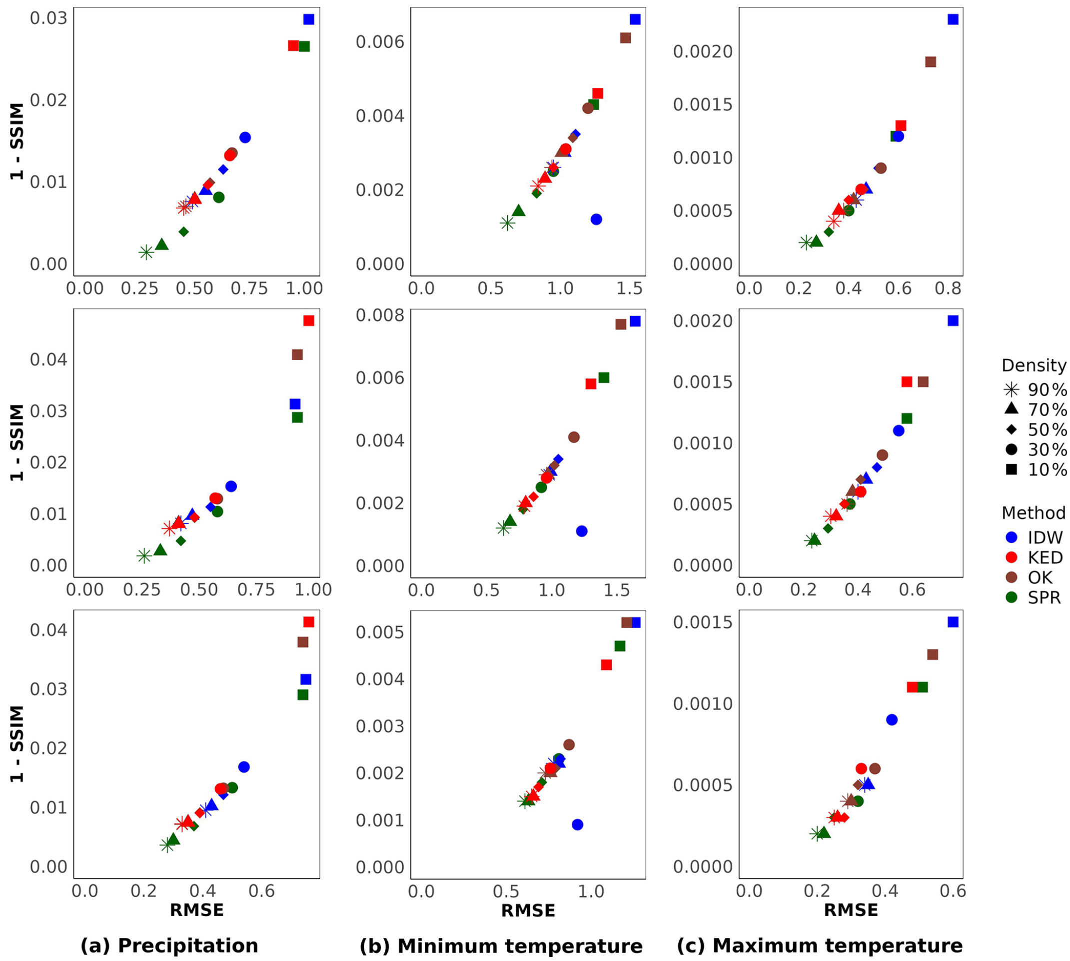

Figure 4. Synthetic data experiments: Comparison of the three baseline methods and SPR in the northern region, in terms of RMSE (x-axis) and 1-SSIM (y-axis). Each column represents a specific meteorological variable, while each row corresponds to a region size–ranging from the largest at the top to the smallest at the bottom. Each color represents a different interpolation method, and each plotting symbol corresponds to a specific network density. The closer a symbol is to the origin, the better the performance.

Table 2. Realistic stress-test experiments: daily RMSE statistics (mean, median, standard deviation (SD) and 2.5 % and 97.5 % quantiles) and average SSIM values for each variable and interpolation method. Lower RMSE and higher SSIM values indicate better performance. The best values for average and median RMSE, as well as SSIM, are shown in bold.

Variable	Method	RMSE					SSIM
		Mean	Median	SD	2.5 %	97.5 %	
PR	SPR	**3.18**	**1.97**	3.63	0.01	13.17	0.6654
	KED	3.39	2.16	4.06	0.04	14.02	**0.8044**
TMIN	SPR	**3.53**	**3.19**	1.8	1.08	7.79	**0.9636**
	KED	3.99	3.54	2.10	1.24	9.00	0.9541
TMAX	SPR	**2.59**	**2.29**	1.39	0.86	6.10	**0.9757**
	KED	2.97	2.59	1.61	1.01	7.09	0.9687

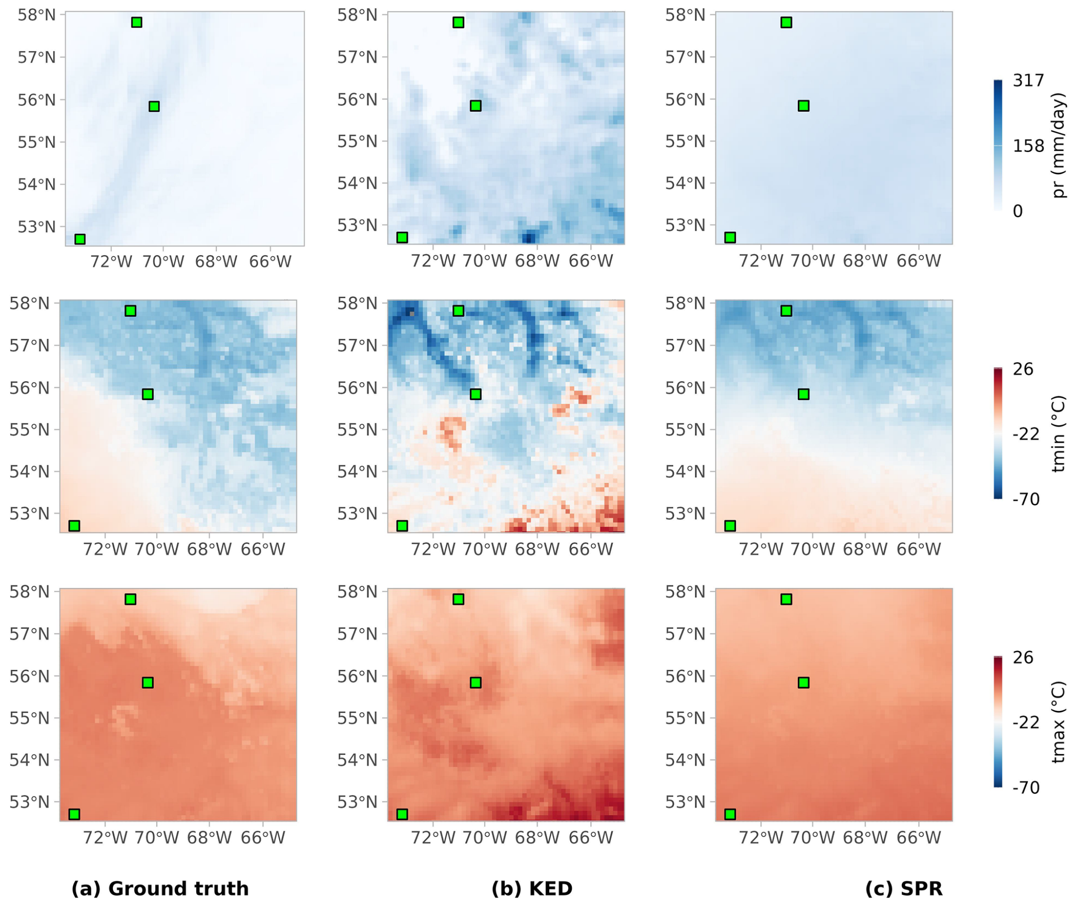

Figure 5. Realistic stress-test experiments: ground truth, interpolated fields of KED and SPR over the larger northern region. Green squares indicate the virtual station locations. Solid squares indicate that KED does not interpolate at virtual station locations. Rows correspond to precipitation, minimum temperature, and maximum temperature (top to bottom) on three different days.

more accurate on average but also more robust to the specific choice of training stations under very sparse network conditions.

At 30 % density, differences between SPR and KED become smaller and more variable across variables, with SPR yielding better accuracy only for precipitation. These results suggest that the primary advantage of SPR emerges in sparse-network settings, while its performance remains competitive as station density increases within the low-density regime considered here.

These findings are consistent with the synthetic experiments, which indicate that the added value of SPR is most pronounced under sparse observational coverage.

4.4 Sensitivity analysis of SPR performance

A factor-wise sensitivity analysis of SPR is conducted using synthetic experiments to assess the influence of station network density, region size, and region location across the three meteorological variables. The main conclusions are summarized below, while detailed results are not shown.

Station network density is the dominant factor controlling SPR performance: increasing density systematically reduces RMSE and uncertainty in both regions. Precipitation exhibits higher RMSE and larger uncertainty than temperature variables, reflecting its greater interpolation difficulty.

In contrast, region size has a limited impact on average performance, with RMSE remaining broadly stable across scales. Region location affects performance in a variable-dependent manner: precipitation errors are lower in the north, whereas temperature errors are higher, suggesting that differences in spatial structure affect interpolation accuracy.

5 Discussion and conclusions

In this work, we introduced Spatial Pattern Regression (SPR), a statistical framework for generating meteorological data that explicitly leverages the spatial structures embedded in RCM simulations. SPR formulates spatial interpolation as a low-rank reconstruction problem, in which daily meteorological fields are represented as linear combinations of spatial patterns extracted from auxiliary gridded data through

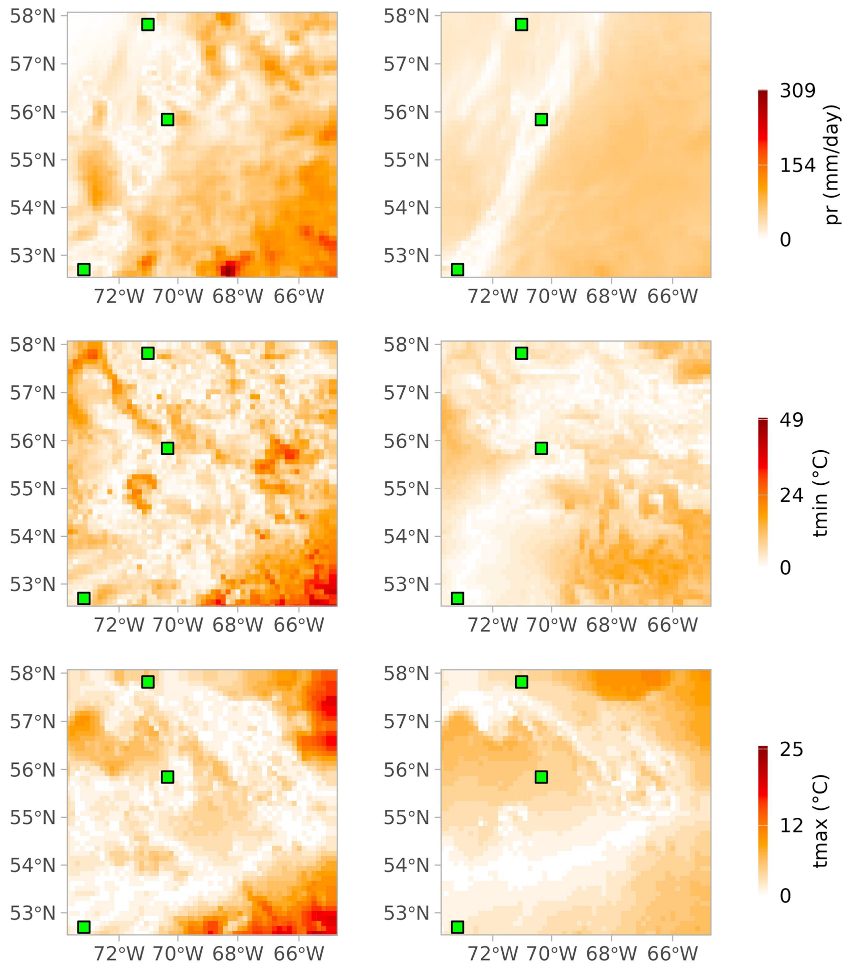

Figure 6. Realistic stress-test experiments: comparison of KED (right column) and SPR (left column) in the larger northern region, based on their spatial RMSE. Each row represents a specific variable: precipitation, minimum temperature and maximum temperature, from top to bottom, shown on the same three days as in Fig. 5. Green squares indicate the virtual station locations. Solid squares indicate that KED does not interpolate at virtual station locations.

SVD/PCA. Overall, SPR reframes interpolation by prioritizing physically consistent spatial organization rather than purely local spatial proximity.

More specifically, the proposed method addresses a key limitation of existing interpolation approaches, which typically incorporate auxiliary information – such as elevation or RCM-based climatologies – through predefined covariates or drift terms, thereby failing to fully leverage the rich spatial structures embedded in gridded datasets. By explicitly extracting spatial patterns from RCM simulations, SPR provides a principled mechanism for integrating physically consistent spatial organization into interpolation, without directly constraining interpolated values toward the auxiliary information. Instead, the auxiliary dataset serves to define the dominant modes of spatial variability rather than to impose absolute values. This conceptual shift – from using auxiliary information to constrain absolute values toward using it to define spatial organization – underpins the development of the proposed methodology and allows SPR to accommodate a wide range of auxiliary gridded products, including RCM outputs, elevation fields, radar products, or remote sensing data.

The experimental results demonstrate that SPR provides accurate and robust interpolations across a wide range of configurations. SPR performs competitively with established baseline methods, including KED, OK, and IDW, and shows a clear advantage under very sparse station network conditions. Among the baseline approaches, KED remains the strongest comparator, consistent with previous findings in spatial interpolation studies (Bishop and McBratney, 2001). Our results also confirm that station density and region location are first-order drivers of interpolation uncertainty, in

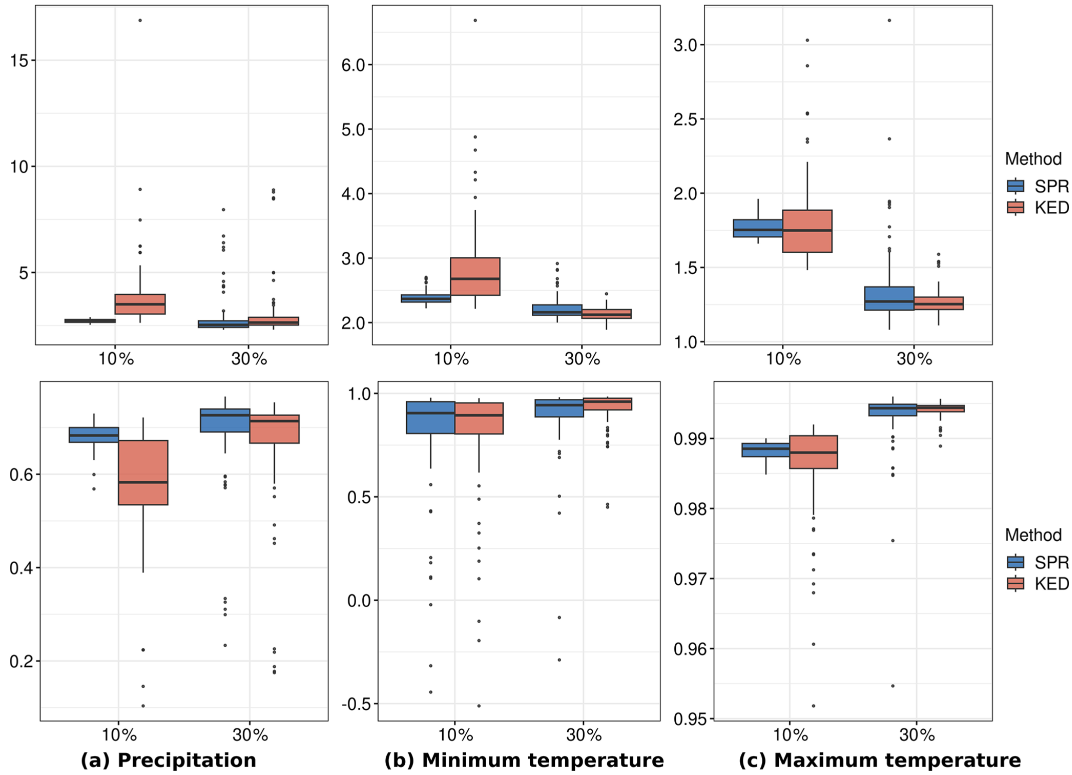

Figure 7. Validation with real observations: distribution of mean daily RMSE (first row) and SSIM (second row) across 100 independent random station samplings for SPR and KED, for each variable. Each box shows the median, interquartile range, and whiskers extending to 1.5 times the interquartile range. Lower RMSE and higher SSIM indicate better performance. Note that 10 % density corresponds to 2–4 stations in a region of 70 000 km^2.

agreement with earlier work (Stahl et al., 2006; Li and Heap, 2014; Wagner et al., 2012). Importantly, SPR does not uniformly outperform KED across all configurations, but its relative benefit becomes most apparent when observational information is severely limited.

Moreover, while all methods are evaluated on the same meteorological scale, they differ fundamentally in the structural information they incorporate. SPR explicitly separates a spatially varying mean structure from low-dimensional variability through the PCA-based decomposition, effectively embedding RCM-derived spatial organization into the interpolation framework. In contrast, IDW and OK rely solely on spatial proximity without incorporating any auxiliary structural information. KED occupies an intermediate position, using RCM-derived monthly climatologies as an external drift term to account for large-scale spatial variability. These differences in modeling philosophy – rather than differences in the scale or nature of the data – explain the contrasting behaviors observed across experimental configurations, particularly under sparse network conditions where the structural constraints provided by SPR become most valuable.

A key contribution of SPR lies in its ability to ensure spatial consistency between interpolated historical fields and future climate simulations. In climate change impact studies, interpolated meteorological data are commonly used to calibrate hydrological models over historical periods, while RCM simulations are employed to assess future changes. By reproducing spatial structures derived from RCM simulations in the interpolated fields, SPR helps reduce structural inconsistencies between calibration and projection phases. This property is particularly relevant for distributed hydrological modeling, where spatial coherence strongly influences simulated fluxes and states.

The extraction of spatial patterns relies on the representativeness of the auxiliary period, which must adequately capture the dominant spatial variability and characteristic meteorological structures of the target variable. Beyond this requirement, the auxiliary period can be selected flexibly and does not need to overlap temporally with the observation period. This flexibility distinguishes SPR from reanalysis-based approaches, which rely on complex data assimilation systems and strict temporal alignment between observations and model states (Gasset et al., 2021). SPR is not intended to replace reanalysis products, but rather to provide a lightweight and transparent alternative when dense observational networks or full assimilation frameworks are unavailable.

Several limitations and avenues for improvement remain. In its current form, SPR relies on a fixed number of leading spatial patterns. Allowing the adaptive selection of spatial patterns, potentially including lower-ranked modes that capture event-specific features, may further enhance performance. In addition, the present implementation estimates regression coefficients independently for each day, ignoring temporal dependence. Incorporating temporal regularization or joint spatio-temporal modeling of the coefficients represents a natural extension of the framework. While a fixed station network was assumed throughout the evaluation for simplicity, future work could consider temporally varying networks, which would better reflect operational conditions where station availability changes over time.

Furthermore, SPR's output may partly inherit the spatial smoothness of the RCM, since it reconstructs fields as low-rank combinations of RCM-derived spatial patterns – a consequence of the dimensionality reduction inherent in the PCA-based representation. Evaluating against the same RCM field may therefore introduce a bias in favor of SPR over methods that preserve finer local variability. However, KED, through its variogram calibration, is also capable of reproducing spatially smooth fields, and the extent to which this bias specifically affects the SPR–KED comparison may be limited. This limitation is specific to the synthetic experiments; in the real-data experiments, evaluation is performed against independent station observations, which provides a complementary assessment. More broadly, conventional RCMs at approximately 11 km resolution are known to smooth spatial variability, particularly for precipitation extremes. The use of convection-permitting regional climate models (CP-RCMs), typically operating at 1–4 km resolution, represents a promising direction for improving the realism of spatial structures used within SPR (Caillaud et al., 2021; Dura et al., 2025).

Notably, SPR leverages spatial patterns derived from RCM simulations, which encode potentially richer structural information than simple variogram-based representations of spatial dependence. However, this also introduces a dependency on the fidelity of the RCM: any systematic spatial bias may directly propagate into the interpolated fields. This sensitivity to RCM structural biases should be considered in operational applications. Recent work has explored alternative ways of exploiting high-resolution model simulations to improve interpolation, including the use of anisotropic variograms derived from CP-RCM simulations (Dura et al., 2025), radar-based ensemble analyses (Vernay et al., 2025), and high-resolution reanalysis products (Khedhaouiria et al., 2026), which represent complementary directions for improving the spatial consistency of interpolated meteorological fields.

In conclusion, SPR offers an efficient and conceptually coherent alternative for spatial interpolation of meteorological variables, particularly suited to poorly instrumented regions where station density is low and spatial consistency is critical. The current formulation provides a solid baseline for future developments aimed at integrating temporal dependence, nonlinearity, and additional sources of spatial information, with clear potential for operational hydrological and climate impact applications.

Code and data availability. All analyses were performed using the R programming language version 4.4.3. The analysis scripts developed for this study are archived at Zenodo (Houssou, 2026), under the MIT License. The primary data source consists of daily climate simulations from the ClimEx project (https://climex-data.srv.lrz.de/Public/, last access: 30 January 2026). We used one ensemble member (kdj) from the CanESM2-driven simulations. Real station observations were obtained from the national monitoring network operated by Environment and Climate Change Canada (ECCC). Data were downloaded automatically through the official bulk data API (https://climate.weather.gc.ca/climate_data/bulk_data_e.html, last access: 30 January 2026).

Author contributions. VH: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review and editing. JC: Conceptualization, Methodology, Supervision, Resources, Funding acquisition, Validation, Writing – review and editing.

Competing interests. The contact author has declared that neither of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. The authors would like to acknowledge funding by the Natural Sciences and Engineering Research Council of Canada (NSERC), by Fonds de Recherche du Québec Nature et Technologies (FRQNT) and by IVADO.

Review statement. This paper was edited by Elena Toth and reviewed by Valentin Dura and one anonymous referee.

References

Abémgnigni Njifon, M., and Schuhmacher, D.: Graph convolutional networks for spatial interpolation of correlated data, Spat. Stat., 60, 100822,0000000 https://doi.org/10.1016/j.spasta.2024.100822, 2024.

Amato, F., Guignard, F., Robert, S., and Kanevski, M.: A novel framework for spatio-temporal prediction of environmental data using deep learning, Sci. Rep., 10, 22243, https://doi.org/10.1038/s41598-020-79148-7, 2020.

Baxevani, A. and Lennartsson, J.: A spatiotemporal precipitation generator based on a censored latent Gaussian field, Water Resour. Res., 51, 4338–4358, https://doi.org/10.1002/2014WR016455, 2015.

Benoit, L., Allard, D., and Mariethoz, G.: Stochastic Rainfall Modeling at Sub-kilometer Scale, Water Resour. Res., 54, 4108–4130, https://doi.org/10.1029/2018WR022817, 2018.

Bishop, T. F. A. and McBratney, A. B.: A comparison of prediction methods for the creation of field-extent soil property maps, Geoderma, 103, 149–160, https://doi.org/10.1016/S0016-7061(01)00074-X, 2001.

Bokke, A.: Comparative Evaluation of Spatial Interpolation Methods for Estimation of Missing Meteorological Variables over Ethiopia, J. Water Resour. Prot., 9, 945–959, https://doi.org/10.4236/jwarp.2017.98063, 2017.

Burhanuddin, S. N. Z. A., Deni, S., and Mohamed Ramli, N.: Geometric median for missing rainfall data imputation, AIP Conf. Proc., 1643, 113–119, https://doi.org/10.1063/1.4907433, 2015.

Burrough, P., and McDonnell, R.: Principles of Geographic Information Systems, Oxford Univ. Press, Oxford, ISBN: 9780191547645, 1998.

Caillaud, C., Somot, S., Alias, A., Bernard-Bouissières, I., Fumière, Q., Laurantin, O., Seity, Y., and Ducrocq, V.: Modelling Mediterranean heavy precipitation events at climate scale: an object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model, Clim. Dynam., 56, 1717–1752, https://doi.org/10.1007/s00382-020-05558-y, 2021.

Caldera, H. P. G. M., Piyathisse, V., and Nandalal, K. D. W.: A Comparison of Methods of Estimating Missing Daily Rainfall Data, Engineer (Sri Lanka), 49, 1, https://doi.org/10.4038/engineer.v49i4.7232, 2016.

Carreau, J. and Guinot, V.: A PCA spatial pattern based artificial neural network downscaling model for urban flood hazard assessment, Adv. Water Resour., 147, 103821, https://doi.org/10.1016/j.advwatres.2020.103821, 2021.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., and Codron, F.: Winter 2010 in Europe: a cold extreme in a warming climate, Geophys. Res. Lett., 37, L20704, https://doi.org/10.1029/2010GL044613, 2010.

Chen, J., Brissette, F. P., Liu, P., and Xia, J.: Using raw regional climate model outputs for quantifying climate change impacts on hydrology, Hydrol. Process., 31, 4398–4413, https://doi.org/10.1002/hyp.11368, 2017.

Cheng, G. and Lu, L.: Comparison of spatial interpolation methods, Adv. Earth Sci., 15, 260–265, 2000.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., and Jones, P. D.: An ensemble version of the E-OBS temperature and precipitation data sets, J. Geophys. Res.-Atmos., 123, 9391–9409, https://doi.org/10.1029/2017JD028200, 2018.

Daly, C., Taylor, G. H., and Gibson, W. P.: The PRISM approach to mapping precipitation and temperature, in: Proceedings of the 10th AMS Conference on Applied Climatology, American Meteorological Society (AMS), 675, 1997.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), J. Adv. Model. Earth Syst., 12, e2019MS001916, https://doi.org/10.1029/2019MS001916, 2020.

Doersch, C.: Tutorial on variational autoencoders, arXiv preprint, arXiv:1606.05908, https://doi.org/10.48550/arXiv.1606.05908, 2016.

Dura, V., Evin, G., Favre, A.-C., and Penot, D.: Improving Precipitation Interpolation Using Anisotropic Variograms Derived from Convection-Permitting Regional Climate Model Simulations, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-1779, 2025.

Eklundh, L. and Pilesjö, P.: Regionalization and spatial estimation of Ethiopian mean annual rainfall, Int. J. Climatol., 10, 473–494, https://doi.org/10.1002/joc.3370100505, 1990.

Erdin, R., Frei, C., and Künsch, H. R.: Data transformation and uncertainty in geostatistical combination of radar and rain gauges, J. Hydrometeorol., 13, 1332–1346, https://doi.org/10.1175/JHM-D-11-096.1, 2012.

Faghih, M. and Brissette, F.: Temporal and Spatial Amplification of Extreme Rainfall and Extreme Floods in a Warmer Climate, J. Hydrometeorol., 24, 1331–1347, https://doi.org/10.1175/JHM-D-22-0224.1, 2023.

Falola, Y., Churilova, P., Liu, R., Huang, C.-K., Delgado, J. F., and Misra, S.: Generating extremely low-dimensional representation of subsurface earth models using vector quantization and deep Autoencoder, Petrol. Res., 2024, https://doi.org/10.1016/j.ptlrs.2024.07.001, 2024.

Flato, G. M.: Earth system models: an overview, WIREs Clim. Change, 2, 783–800, https://doi.org/10.1002/wcc.148, 2011.

Fletcher, S. J. and Zupanski, M.: A data assimilation method for log-normally distributed observational errors, Q. J. R. Meteorol. Soc., 132, 2505–2519, https://doi.org/10.1256/qj.05.222, 2006.

Fletcher, C. G., McNally, W., Virgin, J. G., and King, F.: Toward Efficient Calibration of Higher-Resolution Earth System Models, J. Adv. Model. Earth Syst., 14, e2021MS002836, https://doi.org/10.1029/2021MS002836, 2022.

Fortin, V., Roy, G., Donaldson, N., and Mahidjiba, A.: Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA), J. Hydrol., 531, 296–307, https://doi.org/10.1016/j.jhydrol.2015.08.003, 2015.

Frei, C. and Schär, C.: A precipitation climatology of the Alps from high-resolution rain-gauge observations, Int. J. Climatol., 18, 873–900, https://doi.org/10.1002/(SICI)1097-0088(19980630)18:8<873::AID-JOC255>3.0.CO;2-9, 1998.

Gasset, N., Fortin, V., Dimitrijevic, M., Carrera, M., Bilodeau, B., Muncaster, R., Gaborit, É., Roy, G., Pentcheva, N., Bulat, M., Wang, X., Pavlovic, R., Lespinas, F., Khedhaouiria, D., and Mai, J.: A 10 km North American precipitation and land-surface reanalysis based on the GEM atmospheric model, Hydrol. Earth Syst. Sci., 25, 4917–4945, https://doi.org/10.5194/hess-25-4917-2021, 2021.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria:

Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Hartkamp, A. D., de Beurs, K., Stein, A., and White, J.: Interpolation Techniques for Climate Variables, GIS Series 99-01, CIMMYT, Mexico, 1999.

Hasanpour Kashani, M. and Dinpashoh, Y.: Evaluation of efficiency of different estimation methods for missing climatological data, Stoch. Environ. Res. Risk Assess., 26, 59–71, https://doi.org/10.1007/s00477-011-0536-y, 2012.

Hengl, T.: A Practical Guide to Geostatistical Mapping of Environmental Variables, Geoderma, 140, 417–427, 2007.

Hengl, T., Heuvelink, G. B. M., and Stein, A.: Comparison of kriging with external drift and regression-kriging, Tech. Note, International Institute for Geo-Information Science and Earth Observation (ITC), Enschede, the Netherlands, https://research.utwente.nl/en/publications/comparison-of-kriging-with-external-drift-and-regression-kriging/ (last access: 25 March 2026), 2003.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol., 25, 1965–1978, https://doi.org/10.1002/joc.1276, 2005.

Houssou, V.: Vihotogbe/SPR_interpolation: v1.0.1 – Rinal release, Zenodo [code], https://doi.org/10.5281/zenodo.22236745, 2026.

Hutchinson, M. F.: Interpolating mean rainfall using thin plate smoothing splines, Int. J. Geogr. Inf. Syst., 9, 385–403, https://doi.org/10.1080/02693799508902045, 1995.

Hutchinson, M. F., McKenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E., and Papadopol, P.: Development and Testing of Canada-Wide Interpolated Spatial Models of Daily Minimum–Maximum Temperature and Precipitation for 1961–2003, J. Appl. Meteorol. Climatol., 48, 725–741, https://doi.org/10.1175/2008JAMC1979.1, 2009.

Hwang, Y., Clark, M., Rajagopalan, B., and Leavesley, G.: Spatial interpolation schemes of daily precipitation for hydrologic modeling, Stoch. Environ. Res. Risk Assess., 26, 295–320, https://doi.org/10.1007/s00477-011-0509-1, 2012.

IPCC: Annex VII: Glossary, in: Climate Change 2021: The Physical Science Basis, Cambridge Univ. Press, Cambridge and New York, 2215–2256, https://doi.org/10.1017/9781009157896.022, 2021.

Khedhaouiria, D., Gasset, N., Fortin, V., Dimitrijevic, M., Bulat, M., and Wang, X.: The Canadian Surface Reanalysis (CaSR) v3.2 precipitation dataset: A 45-year high-resolution analysis for North America (1980–2024), EGUsphere [preprint], https://doi.org/10.5194/egusphere-2026-620, 2026.

Lam, N.: Spatial Interpolation Methods: A Review, Cartogr. Geogr. Inf. Sci., 10, 129–150, https://doi.org/10.1559/152304083783914958, 1983.

Lauer, A., Pausata, F. S. R., Leroyer, S., and Argueso, D.: Effect of urban heat island mitigation strategies on precipitation and temperature in Montreal, Canada: Case studies, PLOS Clim., 2, e0000196, https://doi.org/10.1371/journal.pclm.0000196, 2023.

Leduc, M., Mailhot, A., Frigon, A., Martel, J., Ludwig, R., Brietzke, G. B., Giguère, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J.: The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5), J. Appl. Meteorol. Climatol., 58, 663–693, https://doi.org/10.1175/JAMC-D-18-0021.1, 2019.

Li, J. and Heap, A. D.: A review of spatial interpolation methods for environmental scientists, Record 2008/23, Geoscience Australia, ISBN 9781921498305, https://pid.geoscience.gov.au/dataset/ga/68229 (last access: 25 March 2026), 2008.

Li, J. and Heap, A. D.: A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors, Ecol. Inform., 6, 228–241, https://doi.org/10.1016/j.ecoinf.2010.12.003, 2011.

Li, J. and Heap, A. D.: Spatial interpolation methods applied in the environmental sciences: A review, Environ. Model. Softw., 53, 173–189, https://doi.org/10.1016/j.envsoft.2013.12.008, 2014.

Li, Y.: Review on spatial interpolation methods of temperature data from meteorological stations, Prog. Geogr., 33, 1019–1028, https://doi.org/10.11820/dlkxjz.2014.08.002, 2014.

Link, R., Snyder, A., Lynch, C., Hartin, C., Kravitz, B., and Bond-Lamberty, B.: Fldgen v1.0: an emulator with internal variability and space–time correlation for Earth system models, Geosci. Model Dev., 12, 1477–1489, https://doi.org/10.5194/gmd-12-1477-2019, 2019.

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D. R., and Brekke, L.: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013, Sci. Data, 2, 150042, https://doi.org/10.1038/sdata.2015.42, 2015.

Looper, J. P. and Vieux, B. E.: An assessment of distributed flash flood forecasting accuracy using radar and rain gauge input for a physics-based distributed hydrologic model, J. Hydrol., 412/413, 114–132, https://doi.org/10.1016/j.jhydrol.2011.05.046, 2012.

Lucas-Picher, P., Riboust, P., Somot, S., and Laprise, R.: Reconstruction of the Spring 2011 Richelieu River Flood by Two Regional Climate Models and a Hydrological Model, J. Hydrometeorol., 16, 36–54, https://doi.org/10.1175/JHM-D-14-0116.1, 2015.

Lucas-Picher, P., Arsenault, R., Poulin, A., Ricard, S., Lachance-Cloutier, S., and Turcotte, R.: Application of a High-Resolution Distributed Hydrological Model on a U.S.-Canada Transboundary Basin: Simulation of the Multiyear Mean Annual Hydrograph and 2011 Flood of the Richelieu River Basin, J. Adv. Model. Earth Syst., 12, e2019MS001709, https://doi.org/10.1029/2019MS001709, 2020.

Ly, S., Charles, C., and Degré, A.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: A review, Biotechnol. Agron. Soc. Environ., 17, 392–406, 2013.

Maduako, I., Ebinne, E., Idorenyin, U., and Ndukwu, R.: Accuracy Assessment and Comparative Analysis of IDW, Spline and Kriging in Spatial Interpolation of Landform (Topography): An Experimental Study, J. Geogr. Inf. Syst., 9, 354–371, https://doi.org/10.4236/jgis.2017.93022, 2017.

Margaritidis, A.: Comparison of Spatial Interpolation Methods of Precipitation Data in Central Macedonia, Greece, Comput. Water Energy Environ. Eng., 13, 13–37, https://doi.org/10.4236/cweee.2024.131002, 2024.

Mirbod, M., Rajabzadeh Ghatari, A., Saati, S., and Shoar, M.: Industrial parts change recognition model using machine vision, image processing in the framework of industrial information integration, J. Ind. Inf. Integr., 26, 100277, https://doi.org/10.1016/j.jii.2021.100277, 2022.

Mitas, L. and Mitasova, H.: Spatial interpolation, in: Geographical Information Systems: Principles, Techniques, Management and Applications, Wiley, Hoboken, Vol. 1, 481–492, ISBN: 9780471735458, 1999.

Monahan, A. H., Fyfe, J. C., Ambaum, M. H. P., Stephenson, D. B., and North, G. R.: Empirical Orthogonal Functions: The Medium is the Message, J. Clim., 22, 6501–6514, https://doi.org/10.1175/2009JCLI3062.1, 2009.

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 Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Pavão, C., França, G., Marotta, G., Mnezes, P. H. B., Neto, G., and Roig, H.: Spatial Interpolation Applied to Crustal Thickness in Brazil, J. Geogr. Inf. Syst., 4, 142–152, 2012.

Prinn, R. G.: Development and application of earth system models, P. Natl. Acad. Sci. USA, 110, 3673–3680, https://doi.org/10.1073/pnas.1107470109, 2013.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (last access: 25 August 2026), 2023.

Razafimaharo, C., Krähenmann, S., Höpp, S., Rauthe, M., and Deutschländer, T.: New high-resolution gridded dataset of daily mean, minimum, and maximum temperature and relative humidity for Central Europe (HYRAS), Theor. Appl. Climatol., 142, 1531–1553, https://doi.org/10.1007/s00704-020-03388-w, 2020.

Schiemann, R., Liniger, M. A., and Frei, C.: Reduced space optimal interpolation of daily rain gauge precipitation in Switzerland, J. Geophys. Res., 115, D14109, https://doi.org/10.1029/2009JD013047, 2010.

Segond, M.-L., Wheater, H. S., and Onof, C.: The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK, J. Hydrol., 347, 116–131, https://doi.org/10.1016/j.jhydrol.2007.09.040, 2007.

Snepvangers, J. J. J. C., Heuvelink, G. B. M., and Huisman, J. A.: Soil water content interpolation using spatio-temporal kriging with external drift, Geoderma, 112, 253–271, https://doi.org/10.1016/S0016-7061(02)00310-5, 2003.

Sokolchuk, K. and Sokac, M.: Comparison of spatial interpolation methods of hydrological data on example of the Pripyat river basin (within Ukraine), Acta Hydrol. Slovaca, 23, 226–233, https://doi.org/10.31577/ahs-2022-0023.02.0025, 2022.

Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G., and McKendry, I. G.: Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density, Agr. Forest Meteorol., 139, 224–236, https://doi.org/10.1016/j.agrformet.2006.07.004, 2006.

Tan, Q. and Xu, X.: Comparative Analysis of Spatial Interpolation Methods: an Experimental Study, Sens. Transducers, 165, 155–163, 2014.

Tang, G., Clark, M. P., Newman, A. J., Wood, A. W., Papalexiou, S. M., Vionnet, V., and Whitfield, P. H.: SCDNA: a serially complete precipitation and temperature dataset for North America from 1979 to 2018, Earth Syst. Sci. Data, 12, 2381–2409, https://doi.org/10.5194/essd-12-2381-2020, 2020.

Taylor, M. H., Losch, M., Wenzel, M., and Schröter, J.: On the Sensitivity of Field Reconstruction and Prediction Using Empirical Orthogonal Functions Derived from Gappy Data, J. Clim., 26, 9194–9205, https://doi.org/10.1175/JCLI-D-13-00089.1, 2013.

van Hyfte, S., Le Moigne, P., Bazile, E., Verrelle, A., and Boone, A.: High-resolution reanalysis of daily precipitation using AROME model over France, Tellus A, 75, 27–49, https://doi.org/10.16993/tellusa.95, 2023.

Varentsov, M., Esau, I., and Wolf, T.: High-resolution temperature mapping by geostatistical kriging with external drift from large-eddy simulations, Mon. Weather Rev., 148, 1029–1048, https://doi.org/10.1175/MWR-D-19-0196.1, 2020.

Vernay, M., Lafaysse, M., and Augros, C.: Radar-based high-resolution ensemble precipitation analyses over the French Alps, Atmos. Meas. Tech., 18, 1731–1755, https://doi.org/10.5194/amt-18-1731-2025, 2025.

Verworn, A. and Haberlandt, U.: Spatial interpolation of hourly rainfall – effect of additional information, variogram inference and storm properties, Hydrol. Earth Syst. Sci., 15, 569–584, https://doi.org/10.5194/hess-15-569-2011, 2011.

Wagner, P. D., Fiener, P., Wilken, F., Kumar, S., and Schneider, K.: Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions, J. Hydrol., 464/465, 388–400, https://doi.org/10.1016/j.jhydrol.2012.07.026, 2012.

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

Wang, W., Yin, S., Yu, B., and Wang, S.: CLIGEN parameter regionalization for mainland China, Earth Syst. Sci. Data, 13, 2945–2962, https://doi.org/10.5194/essd-13-2945-2021, 2021.

Warren, F., Lulham, N., and Lemmen, D. S.: Canada in a Changing Climate: Regional Perspectives Report, Government of Canada, Ottawa, ON, https://changingclimate.ca/regional-perspectives/ (last access: 25 March 2026), 2022.

Werner, A. T., Schnorbus, M. A., Shrestha, R. R., Cannon, A. J., Zwiers, F. W., Dayon, G., and Anslow, F.: A long-term, temporally consistent, gridded daily meteorological dataset for northwestern North America, Sci. Data, 6, 1–16, https://doi.org/10.1038/sdata.2018.299, 2019.

Zhang, T., Zhou, Y., Zhao, K., Zhu, Z., Chen, G., Hu, J., and Wang, L.: A global dataset of daily maximum and minimum near-surface air temperature at 1 km resolution over land (2003–2020), Earth Syst. Sci. Data, 14, 5637–5649, https://doi.org/10.5194/essd-14-5637-2022, 2022.

Zimmerman, D., Pavlik, C., Ruggles, A., and Armstrong, M. P.: An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting, Math. Geol., 31, 375–390, https://doi.org/10.1023/A:1007586507433, 1999.