Response to the Referee's Comments

June 9, 2026

Vihotogbé Houssou & Julie Carreau

Response to Referee 1

We would like to thank Referee 1 for the careful reading of the manuscript and for the constructive and insightful comments. We appreciate the positive assessment of the relevance and clarity of the work. The suggestions provided have helped us improve the manuscript, both in terms of methodological clarification and presentation.

Below, we address each comment in detail. Referee comments are reproduced in a sans serif font, and our responses are provided below each comment. Changes made in the manuscript are indicated where relevant.

Comment 1 :

The SPR method described in the manuscript is not original. It was introduced years ago and is called Reduced Space Optimal Interpolation (RSOI, Schiemann et al. 2010). Please adjust your manuscript to acknowledge the RSOI reference and eventually point out the elements of originality of your work with respect to RSOI.

Response :

- We thank the Referee for this important comment and for pointing out the connection to Reduced Space Optimal Interpolation (RSOI) as introduced by Schiemann et al. (2010). We acknowledge that SPR and RSOI share conceptual similarities: spatial patterns are extracted via EOF/PCA from a high-resolution gridded dataset, and sparse observations are used to estimate pattern coefficients from which the full field is reconstructed. We were not aware of this prior work, and we have revised the manuscript to explicitly acknowledge RSOI and position SPR with respect to it.

- The revised manuscript now includes a detailed discussion of the relationship between SPR and RSOI, as well as the key differences between the two approaches. In particular, while RSOI is formulated as an optimal interpolation method requiring explicit background and observation error covariance matrices estimated from historical observations, SPR adopts a regression-based formulation in which the amplitudes of spatial patterns are estimated through ordinary least squares on a day-by-day basis. This avoids the need for specifying error statistics and results in a simpler and more flexible framework.

- In addition, an important distinction lies in the origin of the reduced-space basis. In RSOI, this basis is typically derived from gridded datasets obtained through prior interpolation of station observations (specifically, approximately 420 rain gauges interpolated via the SYMAP algorithm of Frei and Schär (1998) onto a 13,216-point grid), meaning that the inferred spatial structures may reflect assumptions introduced by these preliminary interpolation steps. In contrast, SPR derives its spatial patterns exclusively from regional climate model simulations, which are physically based and spatially complete. This avoids potential circularity associated with learning spatial structures from previously interpolated fields and allows SPR to rely on physically consistent spatial variability.

- Furthermore, SPR is evaluated in a substantially more comprehensive experimental framework than RSOI. Schiemann et al. (2010) explicitly acknowledged that systematic analyses of sensitivity to region size, geographic setting, and network characteristics were missing and left as open questions. The present work directly addresses these questions through a full factorial design varying network density

(10%–90%), region size (small, medium, large), and geographic location (north, south), plus a stress-test experiment across three meteorological variables and against three baseline methods.

- Finally, SPR does not require historical observational archives to construct its reduced space, and can therefore be applied in sparsely observed regions. SPR is explicitly motivated by climate change impact studies, with the objective of ensuring structural consistency between interpolated historical fields and future RCM projections. This broader applicability constitutes a key motivation and contribution of the present work.
- A further distinction concerns adaptability to climate change. In RSOI, spatial patterns are derived from historical observation-based gridded fields, which by definition only exist for past periods. If precipitation patterns shift under climate change, RSOI has no mechanism to update its reduced space, as no future dense observational archive exists from which to learn new patterns. SPR does not share this limitation: since its patterns are derived from RCM simulations, which span both historical and future periods (1950–2099 in the ClimEx dataset used here), the auxiliary period can be chosen freely. Patterns representative of a future climate can therefore be extracted directly from future RCM simulations, making SPR genuinely adaptable to changing climate conditions.

The following text has been added to the Introduction (lines 56–67):

> *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 regions.*

The following text has been added to Section 3.2 (lines 146–150):

> *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 interpolated observations, allowing the method to rely on physically consistent spatial structures.*

Comment 2 :

The direct back transformation proposed at the beginning of section 3.5 is prone to introducing systematic errors in the final reconstructed field. When the background is a deterministic model, applying inverse transformations directly to the analysis can lead to systematic underestimation of precipitation, as described by Fletcher and Zupanski (2006). This occurs because the analysis-error variance at grid points must be considered in the inverse transformation. A correction method can be implemented using a Taylor series decomposition of the inverse transformation, as outlined in Fortin et al. (2015) and van Hyfte et al. (2023) for Box-Cox transformations. Alternatively, a more computationally intensive approach proposed by Erdin et al. (2012) involves applying the inverse transformation to 399 quantiles, equidistant in probability.

Response :

- We thank the Referee for highlighting this important point. We agree that applying a direct inverse transformation after performing interpolation in transformed space may in general introduce systematic biases, as it neglects the contribution of analysis-error variance. This issue has been discussed in the data assimilation literature, notably by Fletcher and Zupanski (2006), and in the context of precipitation reconstruction by Fortin et al. (2015) and van Hyfte et al. (2023).

- We wish to clarify, however, that the transformation used in this study is not a log transformation but a **softplus** transformation, defined as $f(x) = \log(\exp(x) - 1)$, whose inverse is $f^{-1}(x) = \log(1 + e^x)$. For precipitation values of practical interest, which are typically large enough that $e^x \gg 1$, this inverse satisfies $f^{-1}(x) = \log(1 + e^x) \approx x$. This means that the back-transformation is approximately the identity function for the range of values encountered in practice, and therefore introduces negligible bias even without correction for analysis-error variance. The softplus transformation thus provides the variance-stabilizing and positivity-preserving benefits of a log-type transformation while avoiding the systematic underestimation that would arise from a pure log back-transformation.

- We note that this consideration applies broadly to any interpolation method that operates on transformed precipitation data. 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. The potential bias therefore affects all methods equally and does not differentially disadvantage any particular approach in the comparative evaluation. More sophisticated correction approaches, such as Taylor-series-based corrections (Fortin et al., 2015; van Hyfte et al., 2023) or quantile-based back-transformations (Erdin et al., 2012), are left for future work.

The following text has been added to Section 3.5 (lines 267–279):

> *The same softplus transformation* $f(x) = \log(\exp(x) - 1)$ *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 non-negative. 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 non-negativity. 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.*

Comment 3 :

Lines 84-88. This part is not clear. Does your method require a fixed station network over time?

Response :

- We thank the Referee for this question and for highlighting the need for clarification. The SPR framework does not require a fixed station network over time. The regression step is performed independently for each day using the set of observations available at that time. In practice, for a given day, the spatial pattern matrix is restricted to the locations of the available stations, and the regression coefficients are estimated accordingly. This allows the method to naturally accommodate temporally varying observation networks.

- In the real-data validation experiments presented in the manuscript, a fixed subset of stations is used over the entire evaluation period for consistency and comparability of results. However, this is a design choice specific to the validation framework and does not reflect a limitation of the method itself.

- It is also worth noting that station observations used in the regression step do not need to be from the same period as the RCM simulations used to extract spatial patterns. The auxiliary period (for PCA) and the interpolation period (for regression) are entirely independent, which gives the method considerable practical flexibility.

The following text has been added to Section 3.2.2 (lines 206–208):

> *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.*

The following clarification has also been added to Section 3.2.2, (lines 172–175):

> *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.*

Comment 4 :

Lines 89-92. This part is not clear. It seems that you are not using station data in your interpolation. Please rephrase this part.

Response :

- We agree that the original wording was unclear. Station observations are explicitly used in the regression step to estimate the amplitudes of the spatial patterns for each day. The method relies on these observations as the primary source of information, while spatial patterns derived from RCM simulations provide structural constraints.
- In the synthetic experiments, a subset of RCM grid cells is used as pseudo-observations (referred to as virtual stations) in order to emulate station networks under controlled conditions. In the real-data validation, actual station observations are used in the same role. In both cases, the observations and pseudo-observations are used identically within the regression framework, and their role is to constrain the temporal amplitudes of the spatial patterns.
- The relevant parts of the manuscript have been clarified to explicitly distinguish between these two experimental settings and to better describe the role of observations in the interpolation process.

The following text has been added to Section 3.4.1 (lines 236–237):

> *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.*

Comment 5 :

It is not clear what procedure was used to select the data for the computation of PCA for daily data. Do you extract the PCA considering all daily data over the whole period? Do you consider only the data belonging to that specific day of the year? Please elaborate more on this point.

Response :

- We thank the Referee for highlighting this missing detail. The PCA used to extract spatial patterns is computed using all daily fields from the auxiliary period (10950 days = 365 days x 30 years, as the simulations use a standard 365-day calendar with no leap years), rather than being conditioned on specific days of the year. This approach ensures that the resulting spatial patterns capture the dominant modes of spatial variability across a wide range of meteorological conditions and remain robust for application throughout the interpolation period. This clarification has been added to the revised manuscript.

The following text has been added to Section 3.1 (lines 138–139):

> *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.*

Comment 6 :

Can you elaborate a bit more on which strategy you used to fit the linear regression model of Eqs. (2)-(3)?

Response :

- We thank the Referee for this suggestion and agree that additional details on the regression fitting procedure improve clarity and reproducibility. For each day, the regression coefficients in Eqs. (2)–(3) are estimated using ordinary least squares (OLS), based on the observations available at that time. Specifically, the coefficients are obtained by minimizing the squared difference between the observed values and their reconstruction from the spatial patterns restricted to station locations.
- The regression is performed independently for each day, using the set of available observations.
- The estimated coefficients $\hat{\beta}_k$ play the role of daily temporal scores: they describe 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, these coefficients correspond exactly to the PCA temporal scores from Eq. (1). In the sparse-observation setting, they are the OLS-based approximation of these scores constrained by the available stations.

The following text has been added to Section 3.2.2 (lines 182–185):

> *The regression coefficients β_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.*

The following text has been added to Section 3.2.2 (lines 187–189):

> *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.*

References

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.

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.

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.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

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.

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.