https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

A new fast multivariate bias correction technique, EMBCCA-UNSEEN (v1.1.0): a case study of compound events in Hunan Province, China, using the UNSEEN approach

Yiwei Yang[1], James Bacon[2], Lanpeng Ji[1], Nadhir Ben Rached[1], Catherine D. Bradshaw[2,3], Jemma C.S. Davie[2], Thomas Crocker[2], and Laurence Bradshaw[4]

[1]School of Mathematics, University of Leeds, Woodhouse, Leeds LS2 9JT, UK
[2]Met Office Hadley Centre, Fitzroy Road, Exeter EX1 3PB, UK
[3]The Global Systems Institute, University of Exeter, North Park Road, Exeter, EX4 4QE, UK
[4]Department of Computer Science, University of Exeter, North Park Road, Exeter EX4 4QF, UK

Correspondence: James Bacon (james.bacon@metoffice.gov.uk)

Abstract. The frequency of extreme climate events such as floods, droughts and heatwaves has been increasing due to anthropogenic climate change. However, the limited availability of historical data poses a significant challenge in accurately assessing these events and their likelihood. This study addresses this issue by employing the UNSEEN approach (UNprecedented Simulated Extremes using ENsembles) in combination with the UK Met Office's DePreSys4 initialized hindcast model to simulate a large ensemble of climate data. To enhance the reliability of the model data, we develop a multivariate bias and correlation correction method, Eigenvalue-based Multivariate Bias Correction with Correlation Alignment (EMBCCA-UNSEEN), which combines mean-shift correction and linear transformations to correct inter-variable correlations, ensuring consistency across multiple dimensions of the model data with the observational data. EMBCCA-UNSEEN v1.1.0 is distributed as an installable Python package (Yang et al., 2026). An evaluation of the corrected model data is conducted using two multivariate fidelity testing methods, including statistical feature assessment and a support vector machine-based testing approach. The test results demonstrate that applying the EMBCCA-UNSEEN correction method can significantly improve the consistency between the model data and the observational data, and is significantly more computationally efficient than alternative approaches.

Using a case study for Hunan Province, China, we demonstrate how the methodology improves the representation of extreme events and quantify the annual risk of extreme compound hot and dry events. The results indicate that the new bias correction method effectively captures inter-variable correlations while preserving variance, thereby enhancing the accuracy of simulations for compound extreme climate events.

1 Introduction

Climate change has become a major environmental concern worldwide, leading to a marked increase in the frequency and intensity of extreme climate events (Masson-Delmotte et al., 2021). These events not only are becoming more frequent, but are also reaching scales of magnitude and duration that are unprecedented in historical records. The observational record is very short, however, which makes quantification of the risks of record-breaking climate extremes and understanding of the

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

conditions that led to them difficult. To tackle this challenge, statistical techniques such as Extreme Value Analysis have been developed to extend the climatological distribution beyond the tails. However, as extreme events are rare in the observational record because of the short length, extrapolating beyond the observations in the tails can often give large errors (Alaya et al., 2020; Wehner et al., 2024; Clarkson et al., 2023). An alternative method is to use the UNSEEN (UNprecedented Simulated Extremes using ENsembles) approach (Thompson et al., 2017) with a dynamical model. This approach is defined as using a large ensemble of initialised climate model data to estimate physically plausible but unprecedented climate events not seen in the historical record. By initialising a climate model with atmospheric, oceanic, and sea-ice observations at regular timesteps and generating many realisations of the possible climate evolution at later timesteps the uncertainty in the observational record of what actually happened can be assessed because the very large ensemble size provides adequate exploration of internal variability (i.e. the observational record can be considered one ensemble member of the possible outcomes from the same initial conditions).

Several studies have used the UNSEEN approach to investigate the likelihood and impacts of extreme climate events. For example, this approach has been used to quantify a 7% chance per year of unprecedented rainfall totals in a given UK winter (Thompson et al., 2017), a 4% chance per year of unprecedented high temperatures in South Africa (Bradshaw et al., 2022), and an 11% chance per year of the UK exceeding the high temperatures of summer 2018 (Kay et al., 2020). In 2022, Hunan Province in China was one of several to experience a record-breaking heatwave whilst at the same time the Yangtze was experiencing a severe drought (Zhu et al., 2024; Zhang et al., 2023). Application of the UNSEEN approach adapted for multivariate data to this region of China would therefore be extremely beneficial in terms of understanding how likely such compound events are in the current climate and how extreme they could be. However, previous studies have highlighted that the credibility of UNSEEN analyses is sensitive to methodological choices, including bias correction and fidelity testing (Kelder et al., 2022a), motivating further methodological development.

The core assumption in the UNSEEN approach is that the climate model simulations are realistic representations of the observations. As the model forecasts drift away from the initial climate states, in the early stages each simulation will still be close to what actually happened and unprecedented extremes are unlikely to occur. As the simulations progress, the forecasts drift further away from the observations but the ensemble distributions are still consistent with them and so plausible unprecedented events can be identified. If the model ensemble develops a distribution inconsistent with the observations then the model bias may have become too great for the data to be used in the UNSEEN approach. The core limitation of the UNSEEN approach is that if the climate model data are not deemed to be realistic representations of the observations then the data cannot be used. Effective climate analysis therefore depends not only on data generation but also on accuracy testing to ensure model outputs reflect real-world climate behaviour, hereafter known as fidelity testing. Systematic biases often occur in climate model simulations because observations used for the initialisation are very scarce in some regions of the world and so model reanalysis data is used instead (which may itself be systematically biased). Systematic biases also occur because parameterizations used for unresolved processes may systematically misrepresent the true process, or because low model resolution results in the blurring of orographic features such as mountains making them wider and lower, which can systematically affect rainfall patterns, and lower resolution models may also systematically under- or over-estimate the strength or frequency of climate phenomena influ-

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

enced by finer-scale dynamics (e.g. tropical cyclones). Unless these systematic biases in model outputs are corrected robustly, application of the data to real-world problems will be flawed (Dosio and Paruolo, 2011).

There are numerous bias-correction techniques available for climate model output. Common univariate methods include mean-shift correction, climatological difference correction, and empirical quantile mapping, with quantile mapping often proving effective but sensitive to the chosen calibration period (Lafon et al., 2013; Kelder et al., 2022b). While such methods can improve agreement between simulated and observed marginal distributions, they do not account for dependencies between variables, limiting their applicability in multivariate settings (François et al., 2020). To address this limitation, a range of multivariate bias-correction methods has been proposed. For example, Li et al. (2014) introduced the Joint Bias Correction (JBC) method, which sequentially corrects temperature and precipitation in order to preserve their inter-variable relationship. The Rank Resampling for Distributions and Dependences (R2D2) approach (Vrac, 2018) extends this concept through a stochastic resampling framework that aims to adjust both marginal distributions and inter-variable and inter-site dependence structures. The MBCn method (Cannon, 2018b) further generalises multivariate bias correction by iteratively adjusting the full N-dimensional distribution of model output to align with observations. Although effective in reproducing observed dependence structures, MBCn can become computationally intensive and costly when applied to high-dimensional data or large spatial domains. More generally, achieving both statistical fidelity and computational efficiency at large scales remains challenging. Related approaches such as MRec (Cannon, 2018b) implement a direct linear recorrelation of model output to match observed correlation structures following marginal correction, providing a computationally efficient alternative that targets second-order dependence without reshaping the full joint distribution. Alternative recorrelation approaches, including matrix-based and sequential regression methods, have been proposed to correct spatial cross-correlation structures prior to applying univariate quantile mapping (Bárdossy and Pegram, 2012). However, these methods are recursive: each variable is corrected conditionally on previously adjusted variables, making the outcome dependent on variable ordering and progressively reducing the effective sample size available for correction at each step. More recently, optimal transport theory has been used to develop stochastic multivariate bias-correction techniques, such as Dynamical Optimal Transport Correction (dOTC) (Robin et al., 2019), although these approaches can also be computationally demanding. Importantly, most existing multivariate bias-correction methods are designed to reproduce the observed joint distribution. While this is well suited to many impacts-modelling applications, it can suppress variance and weaken extremes, making such methods less appropriate for UNSEEN-type analyses that seek to retain ensemble-derived variability in order to explore plausible unprecedented events.

Research indicates that while univariate bias correction methods are well-developed, multivariate methods still face challenges and require further research (Cannon, 2018a). In particular, there is significant scope for investigating how to computationally correct the means and intervariable correlations of multivariate data efficiently without distorting the distribution of extreme values (Wang and Tian, 2022). Advancing this area is essential to improve the accuracy and reliability of climate model data used for risk assessment and other impact studies. This study proposes a new, simple, and computationally efficient multivariate bias correction method that uses a correlation-alignment strategy to correct model bias whilst preserving the relationship between climate variables. We term this method the Eigenvalue-based Multivariate Bias Correction with Correlation

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

Alignment (EMBCCA). When optionally preserving model variance to retain extremes, thereby extending the utility of the UNSEEN approach for multivariate analyses, we refer to the method as EMBCCA-UNSEEN.

Our study also proposes two additional multivariate fidelity testing methods suitable for an UNSEEN approach with multivariate data to verify the consistency of the corrected model data and the observational data. Using a case study for Hunan Province, China, we demonstrate how the methodology improves the representation of extreme events and quantify the annual risk of extreme compound hot and dry events in the region similar to those experienced in 2022. This study provides new theoretical and methodological support for assessment of the risk of extreme climate events in the context of climate warming.

2 Methods

2.1 Data description

For our case study of the UNSEEN approach in China, we use data from the fourth version of the UK Met Office's Decadal Prediction Systems (DePreSys4) (Hermanson, 2023). For the observational reference for bias correction and fidelity testing we use ERA5-Land reanalysis (Muñoz-Sabater et al., 2021). ERA5-Land provides a spatially complete and temporally consistent estimate of land-surface climate variables and is widely used as a benchmark dataset for regional climate model evaluation and UNSEEN-type analyses. We do not assume ERA5-Land to be free of biases in an absolute sense. Instead, ERA5-Land is treated as a pragmatic reference dataset whose large-scale statistical structure and regional climate signals are considered more reliable than those of the initialised climate model. Previous evaluations indicate that ERA5-Land performs particularly well for near-surface temperature over humid subtropical regions of China, while precipitation biases tend to be systematic and regionally coherent rather than random (Wu et al., 2024; Zhang et al., 2025). Bias correction is applied exclusively to DePreSys4, with ERA5-Land serving as the reference distribution for aligning means and inter-variable relationships.

For both simulated and observational data, we extract the average summer (June to August; JJA) temperature and total summer precipitation in Hunan Province from 1992 to 2021. The DePreSys4 data consist of simulations initialised in November of each year and ten 10-year realisations generated for each of those starting years. This results in 100 samples per year with lead times of 1 to 10 years, and provides a total of 3,000 data points over the 30 year period of our analysis for each variable, for use in subsequent statistical analysis and model validation (Figure 1). For our case study, we spatially aggregate the data and consider only the area mean summertime temperature and precipitation metrics.

2.2 Study framework

With a model generated raw dataset and an observational dataset at hand, Figure 2 illustrates the study framework of the UNSEEN approach for multivariate data analysis, demonstrated using the Hunan Province data described in the last subsection. First, we apply multivariate fidelity tests to assess the consistency between the raw model data and the observational data using bootstrapping (details given in Subsection 2.3). If the fidelity tests are passed, then these data can be used for climate analysis. Otherwise, bias correction methods (details discussed in Subsection 2.4) are applied to the raw model data to derive corrected

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

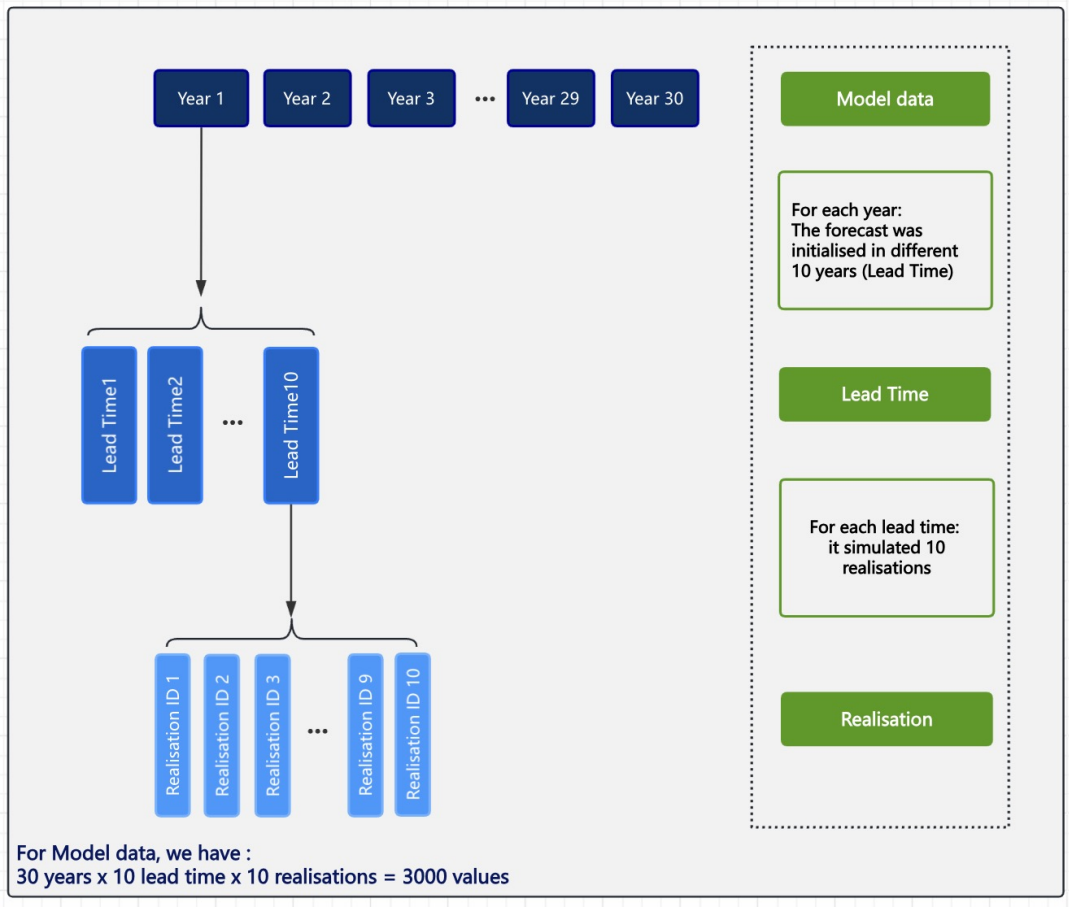

Figure 1. DePreSys4 data description.

model data which can hopefully pass the fidelity test in the next stage. Only when the fidelity tests are passed, can these corrected model data be used in the final stage to assess the likelihood of extreme climate events.

2.3 Fidelity testing

Studies using the UNSEEN approach rely on fidelity testing to identify model biases and to verify the consistency of statistical features between model data and observational data. For example, (Thompson et al., 2017) assesses the mean, standard deviation, skewness, and kurtosis of both model data and observational data distributions to evaluate model accuracy, using a bootstrapping approach (this method we hereafter refer to as traditional fidelity testing). In this approach, a model is deemed accurate if these four features of the observational data fall within the 95% confidence interval of the corresponding model data feature distributions. Other studies employing the UNSEEN approach have similarly used this traditional fidelity test methodology (Thompson et al., 2019; Jain et al., 2020; Bradshaw et al., 2022), though some expand the confidence interval to 97.5 %

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

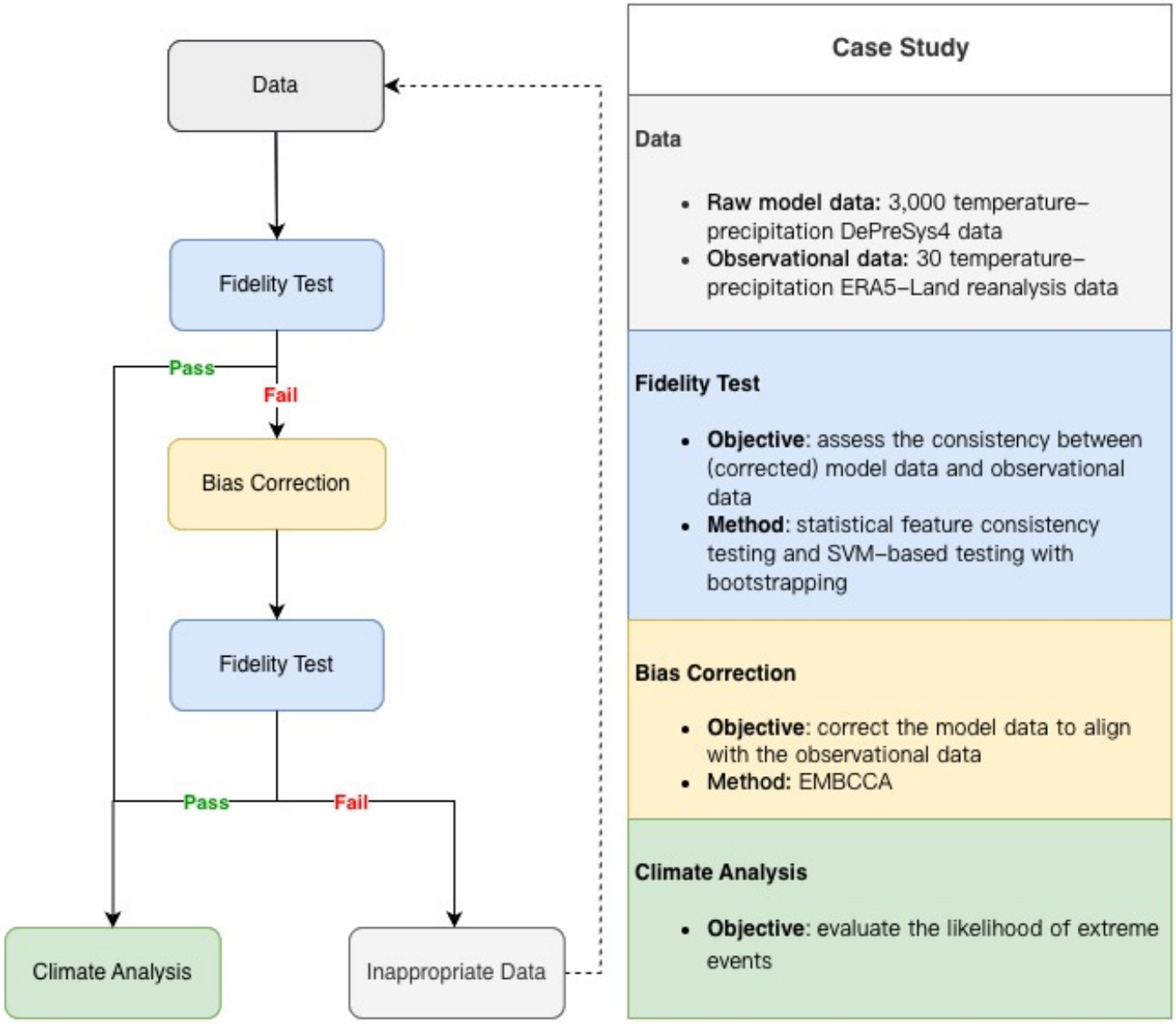

Figure 2. Study framework; case study of Hunan Province demonstrated.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

(Wang et al., 2020). Additionally, some studies have applied maximum likelihood estimation to compute the shape, location, and scale parameters of a Generalised Extreme Value distribution for a more comprehensive analysis (Kelder et al., 2020; Kent et al., 2022). The traditional fidelity testing most commonly used in UNSEEN studies evaluates marginal moments (mean, variance, skewness, kurtosis) independently for each variable, typically using a bootstrap-based percentile test. While this approach is useful for detecting large univariate biases, it is not sufficient for compound-event analysis, where impacts arise from joint behaviour across variables rather than marginal properties alone. In the context of this study, temperature–precipitation dependence is central to hot–dry compound events, yet traditional fidelity testing provides no formal assessment of inter-variable relationships. A model can therefore pass standard UNSEEN fidelity tests while still misrepresenting the joint structure that governs compound extremes, leading to potentially misleading risk estimates.

We therefore introduce two primary multivariate testing methods: multivariate statistical feature consistency (SFC) testing and Support Vector Machine (SVM)-based testing. For the multivariate SFC test, in addition to the evaluation of marginal means and standard deviations using the traditional UNSEEN statistical fidelity tests, we also evaluate inter-variable correlations and multivariate skewness and kurtosis as introduced in (Mardia and Zemroch, 1975). Moment-based tests, even when extended to multivariate settings, still assess only a limited set of summary statistics and statistical approaches can become misleading after bias correction. Two datasets can have similar means, variances, and correlations yet remain distinguishable in a higher-dimensional sense. The SVM-based test is therefore introduced as a distribution-level diagnostic, asking a different question: Can a classifier reliably distinguish model data from observations? By framing fidelity as a separability problem, this approach provides an integrated assessment of multivariate consistency without relying on distributional assumptions or predefined summary statistics. The SVM-based test uses the SVM supervised learning algorithm widely used for classification tasks (Cortes, 1995), which is applied here in a novel way to evaluate the consistency between (corrected) model data and observational data. Specifically, we leverage the SVM's ability to define classification boundaries in high-dimensional spaces to measure the separability of these datasets. Importantly, within this framework, high classification skill is interpreted as low fidelity. Unlike conventional uses of SVM, this approach interprets higher classification accuracy and area under the curve (AUC) values as indicators of greater divergence between the distributions of the model and observational data. In other words, the more accurately the SVM distinguishes between the two datasets, the less consistent their distributions are. This SVM-based testing algorithm, illustrated using the Hunan Province data, is described in Algorithm 1.

The data we use for the fidelity testing uses a bootstrapping approach as follows: a single value is randomly sampled from each year's simulated dataset, representing that year's summer-averaged temperature and precipitation data. This process is carried out across the entire 30-year dataset, ensuring that each newly created sample consists of 30 data points, with each data point corresponding to a specific year's summer-averaged temperature and precipitation. This resampling procedure is repeated 10,000 times, resulting in 10,000 independent bootstrapped samples. Each sample thus comprises a 30-year dataset of paired summer-averaged temperature and precipitation values, which are subsequently used for statistical features testing and evaluation of the model data's fidelity.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

Algorithm 1 SVM-based testing illustrated using the Hunan Province data.

Input: 100 samples of 30-year corrected model data and the observational data.

for each of the 100 samples **do**

Label the observational data as 0 and the model data as 1.

Combine the observational and model data.

Apply `StandardScaler` to eliminate differences in scale.

Train the combined data with an SVM classifier.

Calculate the confusion matrix, accuracy, and area under the curve (AUC).

end for

Output: Average accuracy and average AUC.

2.4 Bias correction

Climate models often display systematic biases in their simulations, and thus bias correction methods are often required in interpreting results. When analysing simulated climate extremes, which often exceed historical ranges, ensuring the plausibility of these simulations is essential. Given the aim of using model data to analyse compound climate events, and that compound climate events are often driven by multiple variables acting together, we propose a new correction method (EMBCCA) for multivariate data. The proposed method differs from existing multivariate bias correction approaches in that it explicitly decouples univariate distributional correction from dependence correction, and addresses the latter through a single deterministic linear transformation acting in standardized space. In contrast to copula-based methods such as R2D2, which aim at reproducing the full empirical multivariate dependence structure through rank-based resampling, the proposed approach does not seek to modify higher-order or nonlinear dependence features. Instead, it focuses on an exact correction of second-order dependence as quantified by the correlation matrix. Unlike iterative methods such as MBCn (Cannon, 2018b), which approximate the target multivariate distribution through successive random rotations and univariate corrections and may suffer from convergence and variance drift in high dimensions, the proposed correction is obtained in closed form, without stochastic elements or iteration, and scales naturally to large-dimensional settings. The method is most closely related to matrix recorrelation techniques, but differs fundamentally in that it targets a bias-correction of original data correlation rather than Gaussian transformed data correlation, thereby ensuring that the marginal variances resulting from the univariate bias correction step are exactly preserved. As such, the method occupies an intermediate position between purely marginal bias correction and fully multivariate distribution-matching approaches, providing a transparent, robust, and computationally efficient correction of linear dependence structures

Since the UNSEEN approach relies on an implicit assumption that the large model ensemble provides a more robust estimate of the true variance than the limited observational data, our application of this method is designed to transform the model data to match the correlations and means of the observational data, while preserving the variances of the model data. EMBCCA-UNSEEN operates by aligning the covariance structure of model-simulated variables with that of observations through an eigenvalue-space transformation. While the underlying linear algebraic operations are well established, their formulation here as a multivariate bias-correction method framework that combines mean correction, explicit correlation align-

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

ment, optional variance preservation, and computational efficiency within a single linear transformation is, to our knowledge, new. Specifically, for each ensemble member, EMBCCA-UNSEEN performs a whitening–recoloring operation based on the eigen-decomposition of the model and observed covariance matrices. This approach adjusts the multivariate relationships between variables while preserving the general shape and dynamics of the model distribution. Unlike existing multivariate bias correction methods such as MBC, MBCp, and MBCR (Cannon, 2018b), which employ iterative quantile mapping to match both marginal distributions and intervariable dependence structures, EMBCCA-UNSEEN provides a computationally efficient, linear alternative that maintains closer fidelity to the original model output. By focusing on covariance alignment rather than full distributional transformation, EMBCCA-UNSEEN minimises distortion of the model's internal variability while still improving its multivariate consistency with observations. The main steps of the EMBCCA-UNSEEN procedure are listed in Algorithm 2, with all calculation details deferred to Appendix A.

Algorithm 2 The EMBCCA-UNSEEN correction method.

Input: model data matrix $\mathbf{Y}$ and observational data matrix $\mathbf{X}$.

Standardise both $\mathbf{Y}$ and $\mathbf{X}$.

Perform orthogonal decomposition for the correlation matrices of $\mathbf{Y}$ and $\mathbf{X}$.

Transform the model data $\mathbf{Y}$ to align its correlation with that of $\mathbf{X}$.

Adjust the mean of the transformed model data.

Output: Corrected model data $\mathbf{Y}^{\mathrm{cor}}$.

To assess the computational performance of the EMBCCA-UNSEEN method, we compare it against a selection of established multivariate bias-correction methodologies implemented in the Statistical Bias Correction Kit (SBCK; version 1.4.2) Python package. Specifically, we consider the Dynamical Optimal Transport Correction (dOTC) method (Robin et al., 2019), the multivariate recorrelation approach (MRec; Cannon 2018b), and the Rank Resampling for Distributions and Dependences (R2D2) method (Vrac, 2018). These methods span a range of conceptual approaches, from deterministic recorrelation to stochastic full-distribution matching, and provide a representative benchmark for evaluating both performance and computational efficiency. Joint Bias Correction (JBC; Li et al. 2014) was not included in the comparison. Although effective for correcting the joint behaviour of temperature and precipitation, JBC is intrinsically formulated as a pairwise, sequential correction framework, in which one variable is adjusted conditionally on another. Consequently, the method is order-dependent and does not generalise naturally to higher-dimensional or scalable settings, including spatially aggregated and potentially multi-variable UNSEEN applications. For consistency and generality, we therefore restrict our comparison to methods that operate in a fully multivariate framework without requiring an explicit variable ordering. The MBCn method (Cannon, 2018b) was also excluded from the final evaluation. While well established and widely used, MBCn aims to reproduce the full observed joint distribution through an iterative sequence of random rotations and univariate corrections. In the present application, this resulted in substantial distortions of the joint temperature–precipitation structure, yielding distributions that were physically implausible and inconsistent with observed patterns.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

For the purpose of evaluating EMBCCA-UNSEEN against other existing multivariate bias correction methods, the exact same bootstrapped samples are used in each case, but we additionally extend our focus to the whole of China in order to get a better estimation of performance and speed.

3 Results

In this section, using the case study for Hunan Province, we demonstrate the proposed methodology and finally quantify the annual risk of extreme compound hot and dry events in the region. We shall follow the flow chart displayed in Figure 2 in our analysis.

3.1 Fidelity testing of raw model data

Following the methodology outlined in Figure 2, we first test the fidelity of the marginal model data using the traditional univariate fidelity testing methodology. In Figure 3, panel a) presents the fidelity test results for temperature data, revealing the significant bias between raw model and observational data, as the observational mean lies far outside the mean of the simulated distribution. This indicates that all simulated sample means are higher than the observational mean, and confirm that the data warrants bias correction of the mean. In contrast, for precipitation data, the observational mean, standard deviation, skewness, and kurtosis fall within the model simulated distributions at positions of 45%, 79%, 85%, and 6%, respectively (not shown). This result suggests that the original DePreSys4 precipitation data reasonably approximates the observational data and may not require bias correction. A simple univariate mean-shift bias correction to adjust the mean of the model temperature data can be applied as shown in panel b) of Figure 3 which successfully corrects the mean temperature bias. However, this will not address the relationship between the temperature and precipitation variables. The correlation between mean JJA temperature and total JJA precipitation in Hunan Province in the ERA5-Land data is -0.69 (see top-left panel in Figure 4). The equivalent correlation in the raw DePreSys4 data is -0.21.

3.2 Bias correction

In this section, we apply the aforementioned bias-correction methods to our bivariate model data, including a marginal correction method where the univariate mean-shift bias correction is applied to the marginals of the raw model data. Figure 4 compares the bivariate summer temperature–precipitation relationship before and after bias corrections. The DePreSys4 output (orange points) in its raw state span a wide temperature range (approximately 25–30°C) and precipitation range (200–1400mm). This distribution deviates substantially from the observational data (blue points), which exhibit a clearer and more coherent temperature–precipitation relationship. In particular, the raw model data fail to reproduce the strength and orientation of the observed inter-variable correlation. A univariate mean-shift correction applied to temperature only improves the alignment of the model and observed temperature means but leaves the joint structure largely unchanged; as a result, the corrected model data continue to display weak and unrealistic temperature–precipitation dependence relative to the observations (top-centre panel in Figure 4). In contrast to both the raw and marginally corrected data, the model points after applying the EMBCCA-UNSEEN

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

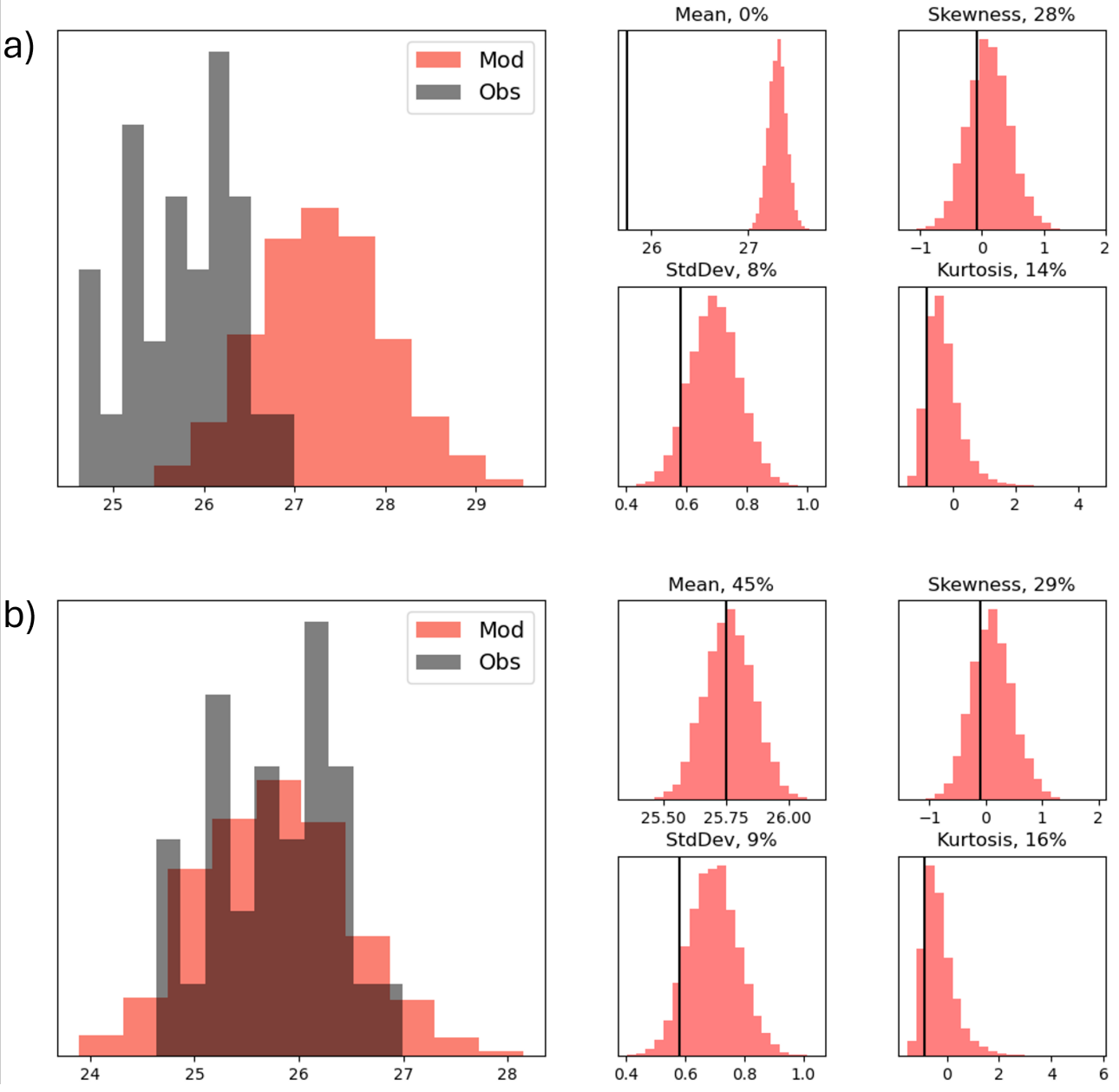

Figure 3. Traditional univariate fidelity testing for DePreSys4 summer mean temperature. a) Raw DePreSys4 model output, b) DePreSys4 model output after applying a univariate mean-shift bias correction. For each subplot, the percentage shown in the title indicates the percentile of the model-derived distribution at which the corresponding observed statistic (ERA5-Land) falls, based on the bootstrap sampling used in the fidelity test.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

correction exhibit a markedly improved alignment with the observed joint structure (top-right). The clustering and orientation of points indicate that the method simultaneously corrects mean biases and reproduces the observed inter-variable correlation, bringing the model distribution into much closer agreement with the observational pattern. The dOTC correction improves the alignment of the marginal distributions relative to the raw model output and introduces a clearer temperature–precipitation relationship. However, the resulting joint distribution remains more diffuse than the observations, with residual discrepancies in both the orientation and spread of the data cloud (bottom-left). This indicates that, while dOTC partially corrects inter-variable dependence, it also reshapes the joint distribution in a way that departs from the observed structure. When using the MRec approach (bottom-centre), the temperature–precipitation relationship more closely follows the observed correlation, reflecting the explicit recorrelation step applied by the method. However, the corrected model data are confined to a noticeably narrower range, with values restricted to the range of the observations. As a result, the spread of the model ensemble is reduced compared to the raw simulations, indicating that the method limits model variability rather than preserving the full range available for exploring extremes with the UNSEEN approach. R2D2 reproduces the general orientation and overall structure of the observed temperature–precipitation relationship, yielding a joint distribution very similar to that obtained with EMBCCA-UNSEEN (bottom-right). The scatter is consistent with the rank-based resampling strategy of the method, which matches the empirical joint distribution while allowing stochastic variation in the placement of individual ensemble members.

Figure 5a shows the spatial pattern of correlation between JJA mean temperature and JJA total precipitation across China for the ERA5-Land observations, the raw DePreSys4 output, and the DePreSys4 data after applying different multivariate bias-correction methods. The observed ERA5-Land data exhibit a coherent and spatially heterogeneous temperature–precipitation relationship, with predominantly negative correlations across much of eastern and southern China, reflecting the coupling between warm conditions and reduced summer rainfall in these regions. The raw, uncorrected DePreSys4 model output reproduces the broad sign of the temperature–precipitation relationship in some regions but shows substantial spatial biases in both the magnitude and geographical structure of the correlation. In particular, the raw model correlations are systematically weaker and more spatially uniform than those derived from observations, indicating a misrepresentation of inter-variable dependence at regional scales. Applying multivariate bias-correction methods leads to varying degrees of improvement in the spatial representation of correlation. Figure 5b presents the anomaly in correlation relative to ERA5-Land for each correction method, highlighting how effectively each approach reduces model–observation discrepancies. Differences between the methods are apparent in both the magnitude and spatial coherence of the remaining anomalies, reflecting their contrasting strategies for correcting dependence structure. Across all methods, EMBCCA-UNSEEN shows the strongest and most spatially coherent improvement in reproducing the observed temperature–precipitation correlation structure. The corrected correlations closely resemble the ERA5-Land pattern in both magnitude and geographical organisation, while the corresponding anomaly map indicates consistently small residual biases across most regions of China. This suggests that explicitly aligning the correlation matrix in standardised space is effective at correcting large-scale inter-variable dependence without introducing additional spatial artefacts. The dOTC method also reduces model–observation discrepancies relative to the raw DePreSys4 output, but noticeable regional biases remain in both the magnitude and spatial structure of the correlation anomalies. While dOTC improves the overall correspondence with observations, the residual patterns indicate that the stochastic optimal-transport adjustment

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

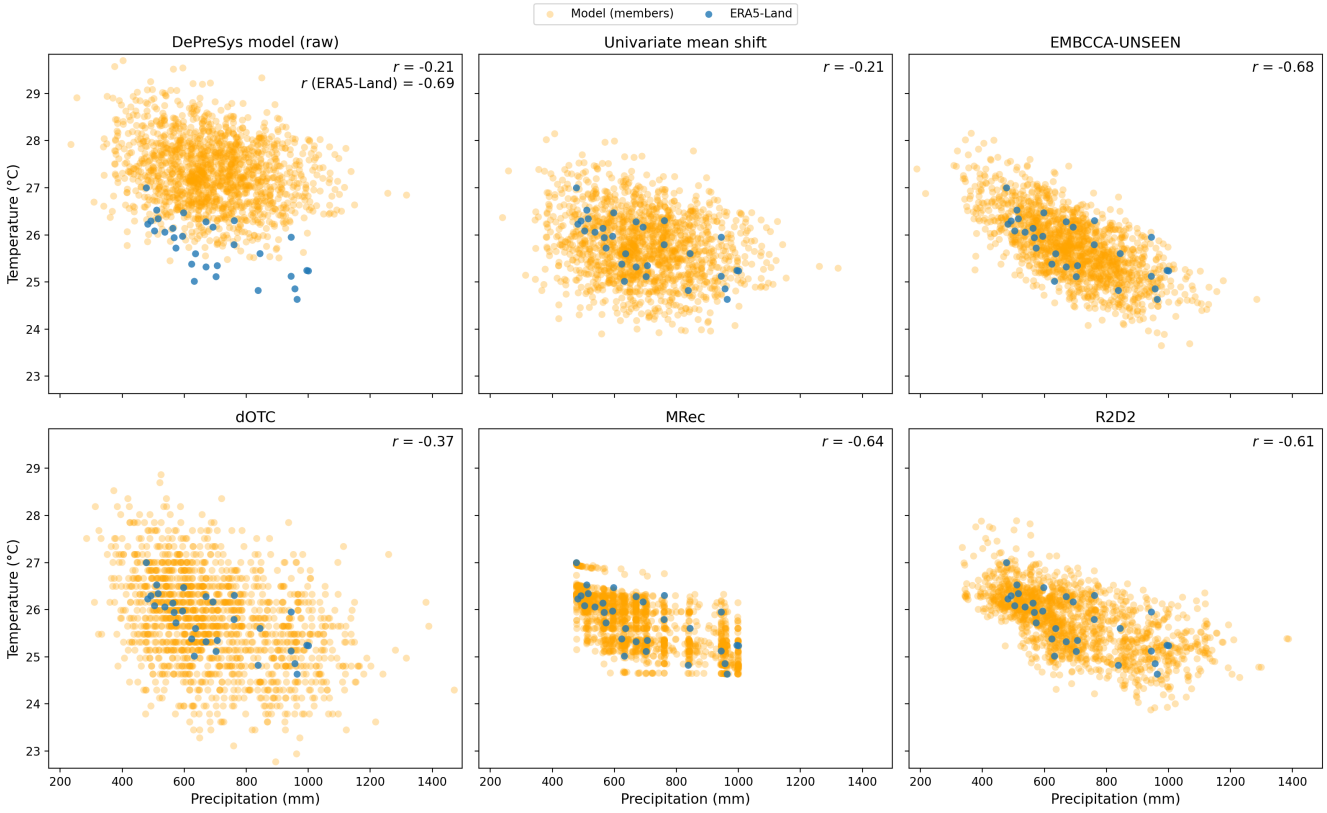

Figure 4. Bivariate summer mean temperature–precipitation distributions for Hunan Province comparing observations (blue points) with DePreSys4 model output before and after bias correction (orange points). The top panels show the raw, uncorrected DePreSys4 model data, the DePreSys4 model data after a univariate mean-shift correction has been applied to temperature only, and the DePreSys4 model data after the EMBCCA-UNSEEN correction has been applied. The bottom panels show the DePreSys4 model data after the dOTC correction, the MRec correction, and the R2D2 correction have been applied.

does not fully recover the observed temperature–precipitation coupling at continental scales. For the MRec approach, the corrected correlations show improved agreement in sign and broad spatial distribution, reflecting the explicit recorrelation step employed by the method. However, the anomaly maps reveal regions where correlations are over- or under-corrected, consistent with the restriction of the corrected model data to the observational range and the associated reduction in ensemble spread.
285 The R2D2 method produces correlation patterns that are broadly consistent with observations and visually similar to those obtained with EMBCCA-UNSEEN. Minor residual differences reflect the stochastic rank-resampling strategy used by R2D2, which introduces local variability in the corrected ensemble while successfully matching the empirical joint structure. Overall, both methods perform well in reproducing the large-scale temperature–precipitation relationship, with EMBCCA-UNSEEN achieving comparable spatial fidelity through a deterministic and computationally efficient correction. The time taken for each

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

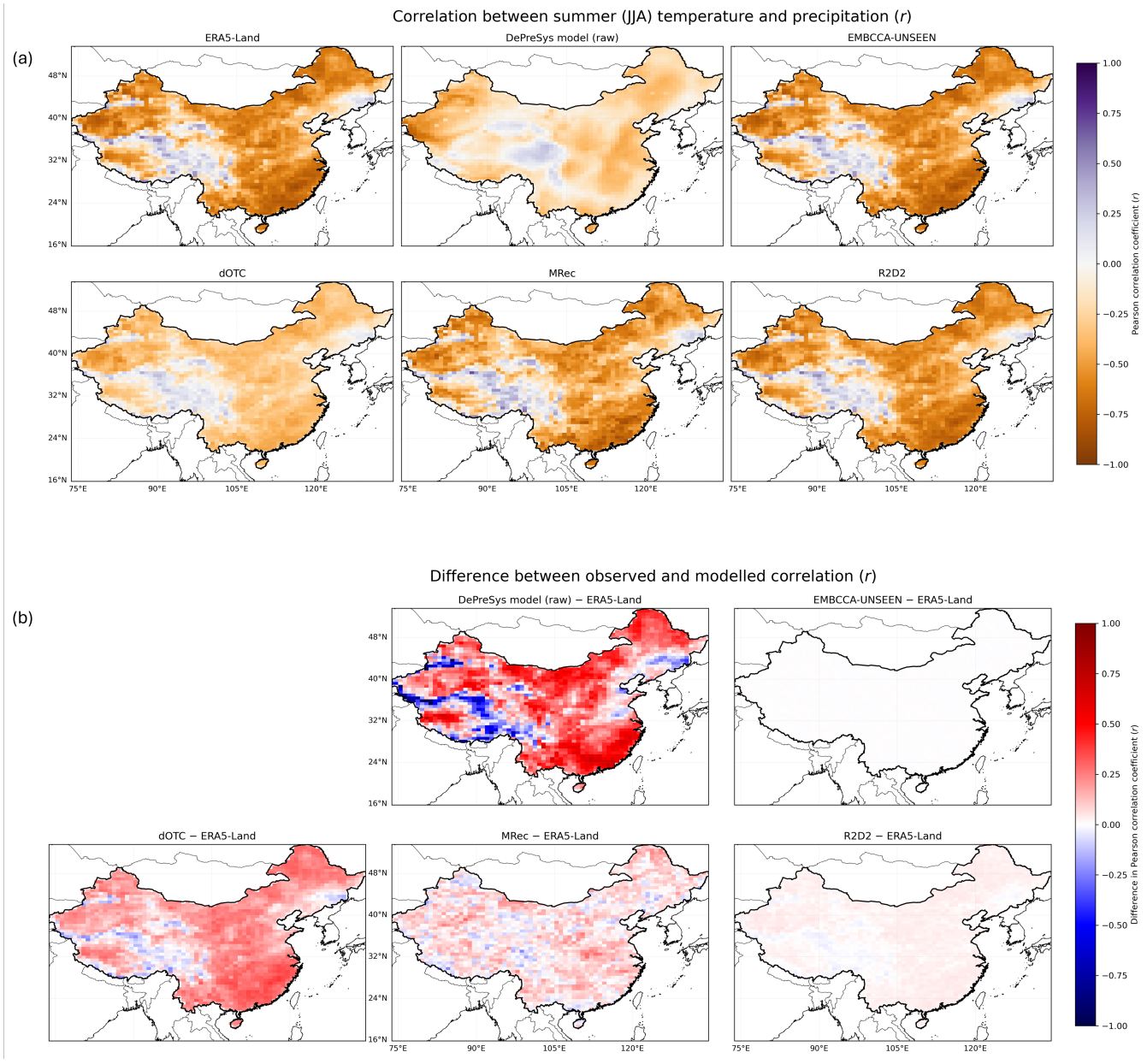

Figure 5. Correlation over China between JJA mean temperature and JJA total precipitation for the ERA5-Land data, the raw uncorrected DePreSys4 data, and the DePreSys4 data bias corrected using different multivariate bias correction methods, including the proposed EMBCCA-UNSEEN correction method. Top set of plots (a) shows the Pearson correlation coefficients between monthly mean temperature and precipitation, while the bottom set (b) shows the anomaly between the Pearson correlation coefficient for ERA5-Land and the DePreSys4 data under the various bias correction states.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

method to correct the biases over China was: EMBCCA-UNSEEN (38 seconds), dOTC (320 seconds), MRec (686 seconds) and R2D2 (642 seconds).

This spatial assessment provides an important first indication of how well each method reproduces observed temperature–precipitation coupling across China. However, while the maps illustrate geographic patterns of agreement and disagreement, they do not directly address whether the corrected correlations are statistically consistent with observations in the sense required by the UNSEEN framework. We therefore complement this spatial analysis with formal fidelity testing of correlation in the following section.

3.3 Fidelity testing of the bias-corrected data

We next assess the fidelity of the bias-corrected model data using the two multivariate testing approaches introduced in Section 2.3: multivariate statistical feature consistency (SFC) testing and SVM-based testing. For the multivariate SFC test, we evaluate consistency in the mean, standard deviation (variance), correlation, skewness, and kurtosis. The skewness and kurtosis are calculated using the multivariate methods of (Mardia and Zemroch, 1975). The EMBCCA-UNSEEN corrected model data successfully pass all components of this fidelity assessment, with each corresponding observational statistic falling within the central 95% confidence interval of the bootstrapped model-derived distributions (see the Supplement). To illustrate the effectiveness of the correction, we focus here on inter-variable correlation.

Figure 6 compares the distribution of temperature–precipitation correlations across the raw, marginally corrected, and multivariate-corrected model data. For both the raw DePreSys4 output and the marginally corrected data, the central correlation values (approximately -0.2) deviate substantially from the observed ERA5-Land correlation (-0.69). In contrast, for the EMBCCA-UNSEEN corrected data, the observed correlation lies close to the centre of the model distribution (47th percentile). This demonstrates that EMBCCA-UNSEEN effectively corrects inter-variable dependence while retaining ensemble variability, a critical requirement for assessing compound climate events driven jointly by temperature and precipitation, such as concurrent heat and drought extremes. For the dOTC-corrected data, the correlation distribution is shifted towards more negative values relative to the raw model output, indicating partial improvement in inter-variable dependence. However, the observed correlation still lies towards the edge of the corrected distribution (0.8th percentile), suggesting that the strength of the temperature–precipitation relationship remains underestimated. The R2D2 method yields a correlation distribution that is more strongly centred around the observed value, with the ERA5-Land correlation falling within the central portion of the model distribution (25th percentile). This indicates that R2D2 effectively captures the observed inter-variable dependence, consistent with its rank-based resampling approach aimed at matching the empirical joint distribution. For the MRec approach, the observed correlation also lies within the central range of the corrected model distribution (32nd percentile), reflecting the explicit recorrelation step applied by the method. However, as discussed previously, this agreement is achieved alongside a restriction of the corrected model values to the observational range, which reduces ensemble spread and limits the representation of more extreme combinations.

To further evaluate the robustness of the bias corrections, we apply the SVM-based fidelity test. Figure 7 summarises the resulting ROC curves for the raw DePreSys4 data and for data corrected using different multivariate bias-correction methods.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

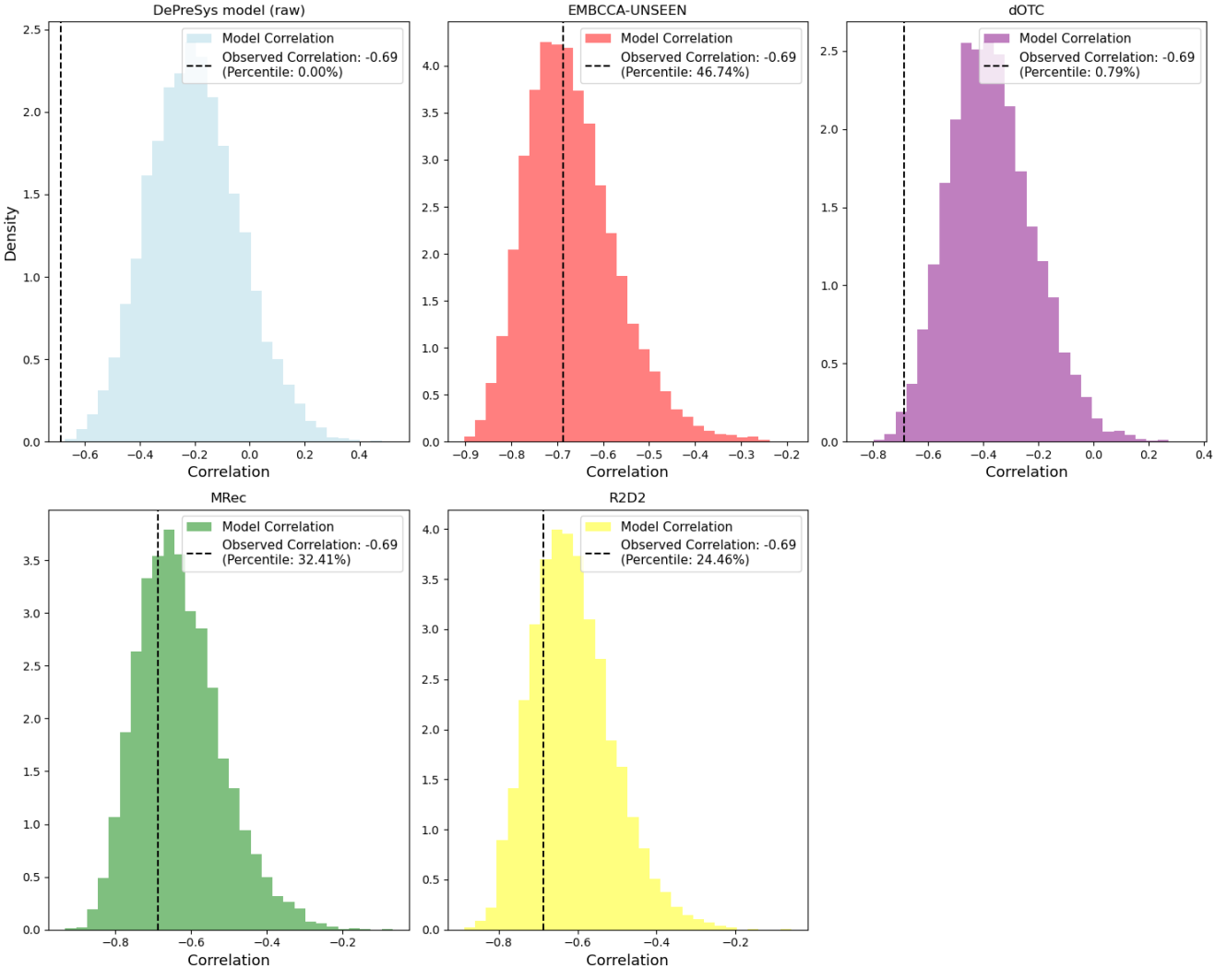

Figure 6. Multivariate fidelity testing for temperature–precipitation correlation: raw, uncorrected DePreSys4 data (top-left) and data bias corrected using different multivariate bias-correction methods: EMBCCA-UNSEEN, dOTC, MRec and R2D2. Percentile positions indicate where the observed ERA5-Land correlation falls within the bootstrapped model correlation distributions.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

For the raw model output, the SVM achieves very high classification skill, with a mean AUC of 0.97, indicating that the modelled and observational data are readily distinguishable. After applying the EMBCCA-UNSEEN correction, the ROC curve lies close to the diagonal, with a mean AUC of 0.56. This indicates that the SVM is only marginally better than random guessing at separating model data from observations, implying a high degree of statistical similarity between the corrected model ensemble and the observed joint temperature–precipitation behaviour. The alternative multivariate methods show more variable performance. The dOTC-corrected data yield an intermediate AUC of 0.72, indicating partial improvement but continued separability between model and observations. For MRec, the AUC drops below random (0.41), consistent with over-correction and restriction of the corrected data to the observed range. The R2D2 method produces an AUC of 0.50, comparable to that obtained with EMBCCA-UNSEEN, reflecting similarly low discriminative power and good overall agreement with observations.

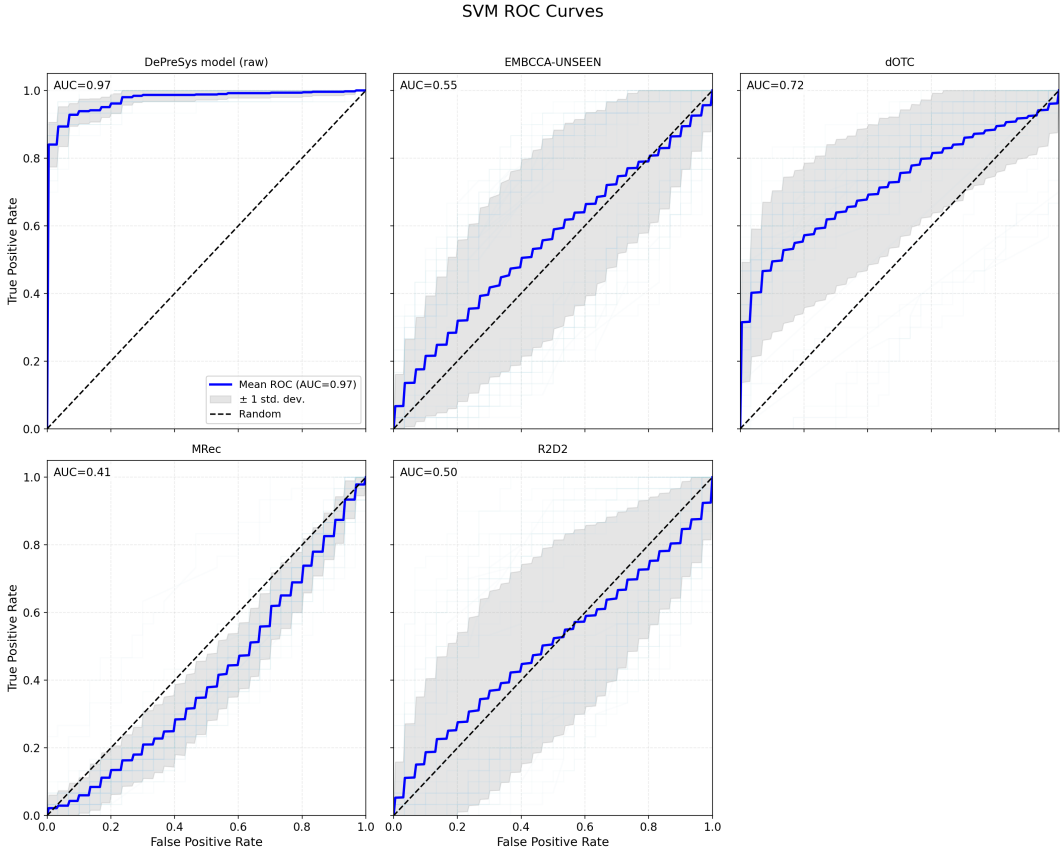

Figure 7. SVM-based fidelity testing using ROC curves for bivariate summer temperature–precipitation data in Hunan Province. Panels show classification performance compared against ERA5-Land observations for the raw uncorrected DePreSys4 output and for data bias-corrected using different multivariate methods: EMBCCA-UNSEEN, dOTC, MRec and R2D2. Reduced classification skill (ROC curves closer to the diagonal) indicates increased statistical similarity between model and observational data.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

3.4 Evaluation of extreme hot and dry events

As discussed in the Introduction, Hunan Province experienced a record-breaking heatwave in 2022, coinciding with severe drought conditions along the Yangtze River basin. These events raise important questions about how likely such compound hot and dry events are in the current climate and how extreme they could plausibly become. Here, we assess the risk of hot, dry, and compound hot–dry extremes using the EMBCCA-UNSEEN corrected ensemble, with comparisons to estimates derived from the raw model output and other bias-corrected data.

Figure 8 shows the estimated probabilities of events exceeding the observed historical records. For reference, equivalent estimates based on the raw DePreSys4 ensemble and marginal mean-shift correction are also shown. A corresponding probability estimate based directly on observations is not feasible due to the limited length of the observational record.

The left panel of Figure 8 presents the individual probabilities of unprecedented dry and hot events. An unprecedented dry event is defined as having summer precipitation below the observed minimum, while an unprecedented hot event exceeds the observed maximum summer temperature. The raw model output produces an estimate for an unprecedented dry event of 9.7% per year. Similar individual dry-event probabilities (around 8–10% per year) are obtained across most methods, reflecting the weaker mean bias in precipitation relative to temperature. In contrast, the raw model output substantially overestimates the risk of an unprecedented hot event, with a probability of 67.1% per year. This overestimation arises from the strong mean bias in the raw temperature simulations, which causes a large fraction of ensemble members to exceed the observed record. Following a marginal correction using EMBCCA-UNSEEN, the probability of an unprecedented hot event is significantly reduced to 4.2%. The estimates obtained using the other bias correction methods range from 1.7% to 12.2%. MRec, which by definition restricts the corrected data range to that of the observed data range, does not produce any unprecedented events and so both unprecedented dry and hot extremes have a probability of 0% per year using this method.

The right panel of Figure 8 shows the estimated joint probability of simultaneous hot and dry conditions. Using the EMBCCA-UNSEEN corrected ensemble, the joint probability is 2.5%, corresponding to a return period of approximately 40–45 years. This estimate lies between those obtained from other approaches. The raw DePreSys4 output substantially overestimates the joint probability at 8.4%, reflecting the combined effect of mean biases and a misrepresentation of temperature–precipitation dependence, which leads to an unrealistically large number of compound extremes. In contrast, methods that either incompletely correct dependence or restrict ensemble behaviour show markedly different outcomes. R2D2 yields a substantially lower joint exceedance probability (0.9%) than EMBCCA-UNSEEN (2.5%) and dOTC (3.0%). This difference is primarily driven by a lower estimated probability of record-exceeding hot events under R2D2 (1.5%), which reduces the frequency of joint hot–dry exceedances. As record-exceedance probabilities are governed by the upper and lower tails of the joint distribution, they are especially sensitive to how bias-correction methods modify variance and tail structure; even when central moments and correlations are well-reproduced, this fact means that small differences in how extremes are represented can lead to large differences in estimated event likelihood.

To further examine the risk of events beyond the observed record, Figure 9 explores how the joint probability of hot and dry conditions varies as the precipitation threshold is decreased below the historical minimum and as the temperature threshold

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

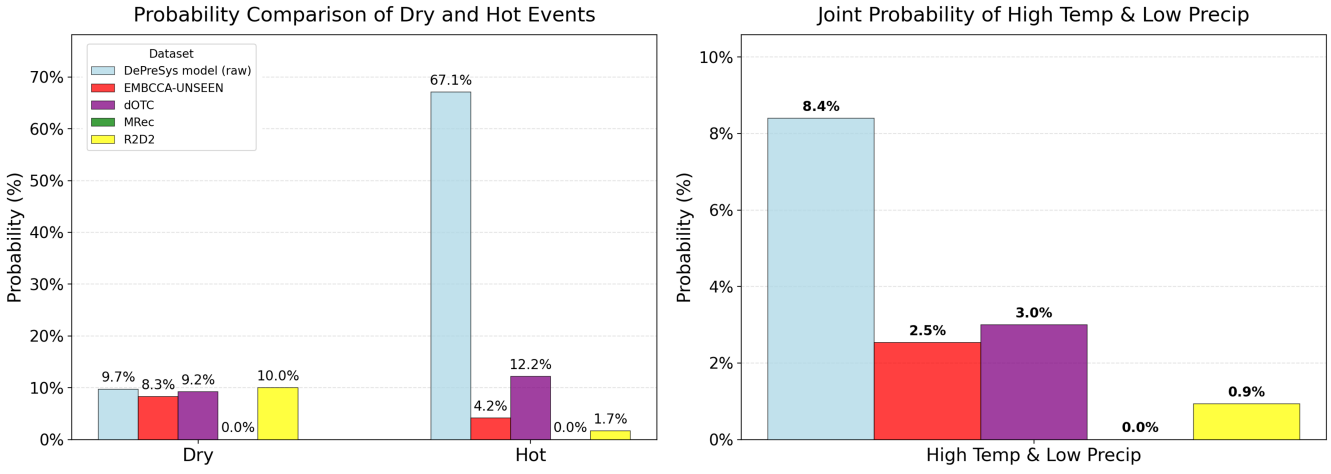

Figure 8. Estimated probabilities of events exceeding the historical record for summer temperature and precipitation in Hunan Province. Left: individual probabilities of unprecedented dry (low precipitation) and unprecedented hot (high temperature) events. Right: joint probability of concurrent hot and dry events. Estimates are shown for raw DePreSys4 output and for data corrected using different bias-correction methods: EMBCCA-UNSEEN, dOTC, MRec and R2D2.

is increased above the historical maximum. The reference point at zero corresponds to the record-breaking joint event with a probability of 2.53%, estimated using the EMBCCA-UNSEEN bias correction method.

As precipitation thresholds become progressively lower, the probability of a compound event declines non-linearly, exhibiting an approximately quadratic decrease. In contrast, increasing the temperature threshold leads to a more gradual, near-linear reduction in joint probability. The dashed horizontal line at 1% indicates a 100-year return period. For example, the results indicate that an event with temperatures at least as high as the historical maximum and precipitation approximately 80-90 mm below the observed minimum would be expected to occur once every 100 years.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

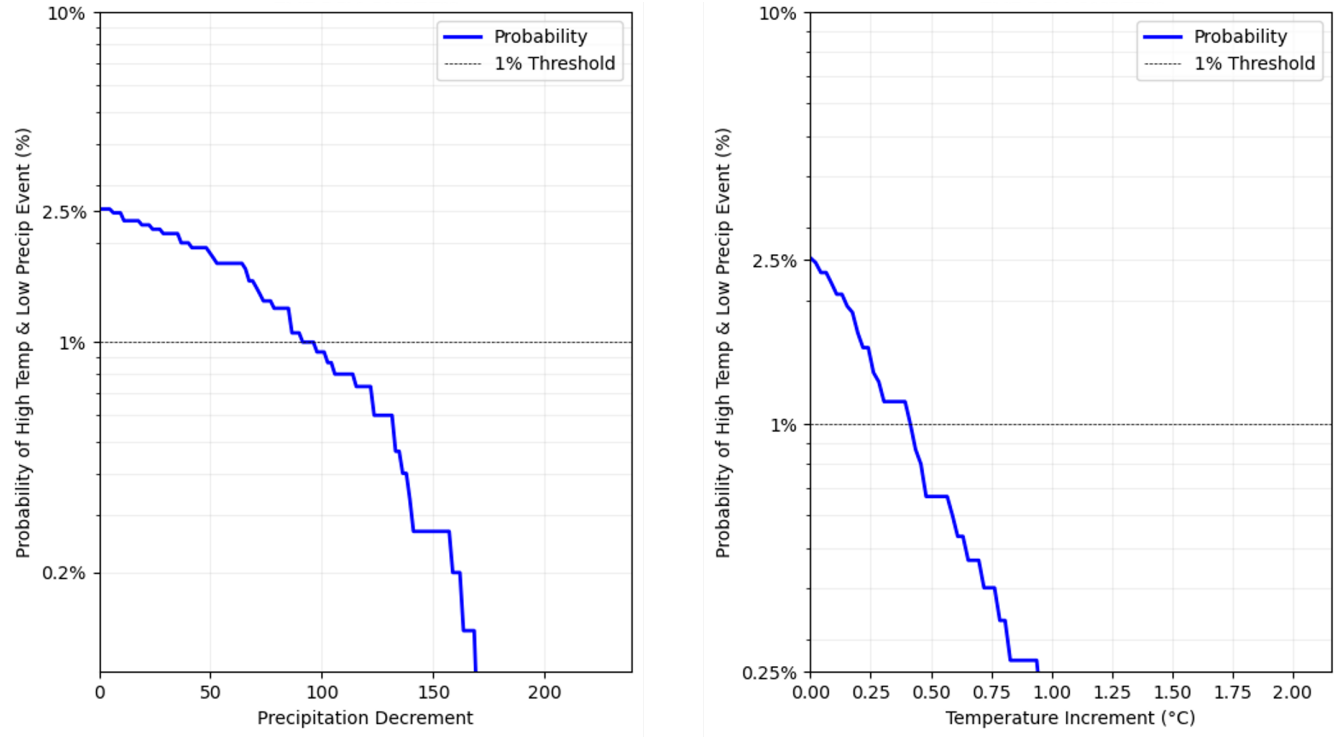

Figure 9. Joint probability of hot and dry event (annual percentage) according to progressively lower precipitation thresholds (left) and incrementally higher temperature thresholds (right). Initial temperature and precipitation values (i.e. where x-axis = 0) are those associated with the compound record-breaking probability of 2.53%.

4 Conclusions

This study introduces EMBCCA-UNSEEN, a fast and computationally efficient multivariate bias-correction method developed to support UNSEEN-type analyses of compound climate extremes. The method corrects mean biases and explicitly aligns inter-variable correlations using a single deterministic linear transformation, while optionally preserving the variance of the climate model ensemble. This design addresses a key limitation of many existing multivariate bias-correction approaches, which are often optimised for impacts-modelling applications and can unintentionally suppress variability and extremes that are central to UNSEEN-based risk assessment. Using a case study of summer temperature and precipitation in Hunan Province, China, we demonstrate that EMBCCA-UNSEEN substantially improves the representation of joint temperature–precipitation behaviour relative to raw model output and marginal bias correction alone. The corrected data successfully pass a comprehensive set of multivariate fidelity tests, including consistency in marginal moments, inter-variable correlation, and multivariate skewness and kurtosis. In addition, an independent SVM-based fidelity test shows that, following correction, the model data become statistically difficult to distinguish from observations, indicating a level of joint consistency that is appropriate for UNSEEN applications.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

Comparison with alternative multivariate bias-correction methods highlights important conceptual trade-offs. Stochastic full-distribution matching approaches such as dOTC and R2D2 can substantially improve dependence structure but may reshape variability and tail behaviour in ways that strongly influence record-exceedance probabilities. Conversely, recorrelation-based approaches such as MRec restrict corrected model values to the observed range, by construction eliminating unprecedented events. EMBCCA-UNSEEN occupies an intermediate position: it corrects second-order dependence effectively while retaining ensemble-derived variability, thereby avoiding both the strong overestimation of compound extremes seen in the raw model output and the suppression of extremes associated with methods that constrain the corrected data to the observational range.

Applied within the UNSEEN framework, EMBCCA-UNSEEN yields an estimated annual probability of a compound hot and dry event comparable to the 2022 conditions in Hunan Province of order 2–3%, corresponding to a return period of several decades. Extending the analysis beyond the observed record shows how compound-event probabilities decline as temperature and precipitation thresholds are pushed further into the tails of the joint distribution, highlighting the strong sensitivity of record-exceedance probabilities to how bias-correction methods modify variance and tail behaviour. This reinforces the importance of preserving physically plausible variability when assessing compound extreme risks.

A further key strength of EMBCCA is its computational efficiency. When applied across China, the time required to perform bias correction is approximately 38 seconds for EMBCCA-UNSEEN, compared with 642 seconds for the next most appropriate multivariate bias correction methodology for UNSEEN in our study region: R2D2. This substantial reduction in computational cost reflects the closed-form, non-iterative nature of the method and makes EMBCCA-UNSEEN particularly well suited to large-ensemble, spatially extensive, or operationally constrained applications.

Importantly, EMBCCA can also be applied without variance preservation as a general multivariate bias-correction method, making it suitable for non-UNSEEN applications such as impacts modelling where matching observed variability may be desired. In this configuration, EMBCCA provides a deterministic and interpretable alternative to computationally expensive multivariate distribution-matching approaches, while retaining the ability to efficiently correct inter-variable dependence.

While this study focuses on a bivariate case with predominantly linear dependence, several avenues for future research remain. These include application to higher-dimensional variable sets, spatially resolved fields, and other climate regimes, as well as extensions to settings where non-linear dependence structures are important. Beyond climate science, the general framework may also be applicable to multivariate bias-correction problems in other domains where dependence structure and variability both play a critical role such as financial modelling (bias correction for multivariate time-series data) or to medical research (analysis of multivariate genomic datasets). EMBCCA and EMBCCA-UNSEEN therefore provide a versatile, efficient, and robust addition to the toolkit for multivariate bias correction and compound-event risk assessment.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

Code and data availability. The EMBCCA-UNSEEN package (v1.1.0) is archived through Zenodo (https://doi.org/10.5281/zenodo.21102535) and developed openly at https://github.com/MetOffice/embcca-unseen. The repository also contains the manuscript analysis scripts, preprocessing workflows, and example datasets. The software is released under the BSD 3-Clause Licence. Data sources:

- DePreSys4 CMIP6 DCPP hindcast data (1960–2018): https://data.ceda.ac.uk/badc/cmip6/data/CMIP6/DCPP/MOHC/HadGEM3-GC31-MM/dcppA-hindcast
- DePreSys4 CMIP6 DCPP forecast data (2019–2024): https://data.ceda.ac.uk/badc/cmip6/data/CMIP6/DCPP/MOHC/HadGEM3-GC31-MM/dcppB-forecast
- ERA5-Land reanalysis: https://cds.climate.copernicus.eu
- Country and administrative boundary shapefiles: https://www.naturalearthdata.com/downloads/

The repository includes preprocessing workflows required to reproduce the gridded manuscript input datasets from these source data.

Author contributions. YY developed the EMBCCA and EMBCCA-UNSEEN methods and performed the analysis. JB led the evaluation and comparison against existing multivariate bias-correction methods and, along with CDB, JCSD, and TC, contributed to study design, interpretation, software development, and manuscript development. LB developed the software into a publishable package, implemented testing and documentation infrastructure, and contributed bug fixes and code streamlining. LJ and NBR contributed to methodological development, mathematical formulation, and interpretation. All authors contributed to writing, review, and editing of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work and some of its contributors (CDB, JB, JCSD, TC) were partially funded by the Met Office Climate Science for Service Partnership (CSSP) China project under the International Science Partnerships Fund (ISPF).

Appendix A: Multivariate bias correction theory

This section provides detailed derivations and proofs for the proposed EMBCCA-UNSEEN method. Suppose we have two datasets: an observational data matrix $\mathbf{X}$ of dimension $n \times d$ and a model data matrix $\mathbf{Y}$ of dimension $k \times d$. We aim to transform the model data $\mathbf{Y}$ so that it has the same correlation matrix and mean vector across columns as the observational data $\mathbf{X}$, while preserving its variances.

We denote $\mu_o \in \mathbb{R}^d$ and $\sigma_o \in \mathbb{R}^d$ to be the mean and standard deviation row vectors of the observational data $\mathbf{X}$, and $\mu_m \in \mathbb{R}^d$ and $\sigma_m \in \mathbb{R}^d$ to be the mean and standard deviation row vectors of the model data $\mathbf{Y}$. In the sequel, $\mathbf{1}_n = (1, \ldots, 1) \in \mathbb{R}^n$, and $\top$ denotes the transpose sign. The sum (or difference) of a matrix and a row vector is understood columnwise. For a vector $\mathbf{a}$, $\mathrm{diag}(\mathbf{a})$ represents a diagonal matrix with vector $\mathbf{a}$ on its diagonal.

The multivariate bias correction procedure using a linear transformation is as follows.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

1. **Standardising the data.**

We standardise both $\mathbf{X}$ and $\mathbf{Y}$ to obtain zero-mean, unit-variance versions of each variable:

$$\mathbf{X}^{\text{st}} = (\mathbf{X} - \mathbf{1}_n^\top \mu_o)\,\text{diag}\left(\frac{1}{\sigma_o}\right), \quad \mathbf{Y}^{\text{st}} = (\mathbf{Y} - \mathbf{1}_k^\top \mu_m)\,\text{diag}\left(\frac{1}{\sigma_m}\right). \tag{A1}$$

2. **Computing correlation matrices and performing orthogonal decomposition.**

We calculate the correlation matrices for both the standardised observational data and the standardised model data, and perform an orthogonal decomposition:

$$C_o = \frac{1}{n-1}(\mathbf{X}^{\text{st}})^\top \mathbf{X}^{\text{st}} = \mathbf{W}\Lambda\mathbf{W}^\top, \tag{A2}$$

$$C_m = \frac{1}{k-1}(\mathbf{Y}^{\text{st}})^\top \mathbf{Y}^{\text{st}} = \mathbf{V}\Gamma\mathbf{V}^\top. \tag{A3}$$

where C_o and C_m are the correlation matrices of the observational and model data, respectively. We assume both C_o and C_m are invertible, so the diagonal entries of Λ and Γ are positive. Consequently, we can define $\Gamma^{1/2}$ and $\Gamma^{-1/2}$ as element-wise square root and inverse square root matrices, respectively, with similar definitions for Λ.

3. **Correcting the correlation of model data.**

We adjust the model data so that it has the same correlation structure as the observational data by applying the following transformation:

$$\mathbf{Z}_m = \mathbf{Y}^{\text{st}}\mathbf{V}\Gamma^{-1/2}\Lambda^{1/2}\mathbf{W}^\top. \tag{A4}$$

Next, we demonstrate that the correlation matrix of $\mathbf{Z}_m$ equals C_o. We have, by definition,

$$C_{Z_m} = \frac{1}{k-1}\mathbf{Z}_m^\top \mathbf{Z}_m. \tag{A5}$$

By substituting $\mathbf{Z}_m = \mathbf{Y}^{\text{st}}\mathbf{V}\Gamma^{-1/2}\Lambda^{1/2}\mathbf{W}^\top$, we get

$$C_{Z_m} = \mathbf{W}\Lambda^{1/2}\Gamma^{-1/2}\mathbf{V}^\top \left(\frac{1}{k-1}(\mathbf{Y}^{\text{st}})^\top \mathbf{Y}^{\text{st}}\right)\mathbf{V}\Gamma^{-1/2}\Lambda^{1/2}\mathbf{W}^\top. \tag{A6}$$

Since $C_m = \frac{1}{k-1}(\mathbf{Y}^{\text{st}})^\top \mathbf{Y}^{\text{st}} = \mathbf{V}\Gamma\mathbf{V}^\top$ and $\mathbf{V}^\top\mathbf{V} = \mathbf{I}$, we have

$$C_{Z_m} = \mathbf{W}\Lambda^{1/2}\Gamma^{-1/2}\mathbf{V}^\top\mathbf{V}\Gamma\mathbf{V}^\top\mathbf{V}\Gamma^{-1/2}\Lambda^{1/2}\mathbf{W}^\top = \mathbf{W}\Lambda\mathbf{W}^\top = C_o. \tag{A7}$$

4. **Correcting the mean while preserving variances.**

We adjust the mean of $\mathbf{Z}_m$ to match the observational data mean, while preserving the variances of the model data. The final transformed model data $\mathbf{Y}^{\text{cor}}$ are given by:

$$\mathbf{Y}^{\text{cor}} = \mathbf{Z}_m\,\text{diag}(\sigma_m) + \mathbf{1}_k^\top \mu_o. \tag{A8}$$

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

References

Alaya, M. B., Zwiers, F., and Zhang, X.: An evaluation of block-maximum-based estimation of very long return period precipitation extremes with a large ensemble climate simulation, Journal of Climate, 33, 6957–6970, https://doi.org/10.1175/JCLI-D-19-0011.1, 2020.

Bárdossy, A. and Pegram, G.: Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small, Water Resources Research, 48, https://doi.org/10.1029/2011WR011524, 2012.

Bradshaw, C. D., Pope, E., Kay, G., Davie, J. C., Cottrell, A., Bacon, J., Cosse, A., Dunstone, N., Jennings, S., Challinor, A., et al.: Unprecedented climate extremes in South Africa and implications for maize production, Environmental Research Letters, 17, 084 028, https://doi.org/10.1088/1748-9326/ac816d, 2022.

Cannon, A. J.: Multivariate Bias Correction of Climate Model Output: Matching Marginal Distributions and Intervariable Dependence Structure, Journal of Climate, 31, 8285–8307, https://doi.org/10.1175/JCLI-D-15-0679.1, 2018a.

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Climate dynamics, 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2018b.

Clarkson, D., Eastoe, E., and Leeson, A.: The importance of context in extreme value analysis with application to extreme temperatures in the US and Greenland, Journal of the Royal Statistical Society Series C: Applied Statistics, 72, 829–843, https://doi.org/10.1093/jrsssc/qlad020, 2023.

Cortes, C.: Support-Vector Networks, Machine Learning, 20, 273–297, https://doi.org/10.1007/BF00994018, 1995.

Dosio, A. and Paruolo, P.: Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/10.1029/2011JD015934, 2011.

François, B., Vrac, M., Cannon, A. J., Robin, Y., and Allard, D.: Multivariate bias corrections of climate simulations: which benefits for which losses?, Earth System Dynamics, 11, 537–562, https://doi.org/10.5194/esd-11-537-2020, 2020.

Hermanson, L.: MOHC HadGEM3-GC31-MM model output prepared for CMIP6 DCPP dcppA-hindcast, https://www.wdc-climate.de/ui/entry?acronym=C6_4595253, 2023.

Jain, S., Scaife, A. A., Dunstone, N., Smith, D., and Mishra, S. K.: Current chance of unprecedented monsoon rainfall over India using dynamical ensemble simulations, Environmental Research Letters, 15, 094 095, https://doi.org/10.1088/1748-9326/ab7b98, 2020.

Kay, G., Dunstone, N., Smith, D., Dunbar, T., Eade, R., and Scaife, A.: Current likelihood and dynamics of hot summers in the UK, Environmental Research Letters, 15, 094 099, https://doi.org/10.1088/1748-9326/abab32, 2020.

Kelder, T., Müller, M., Slater, L., Marjoribanks, T., Wilby, R., Prudhomme, C., Bohlinger, P., Ferranti, L., and Nipen, T.: Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes, Climate and Atmospheric Science, 3, 47, https://doi.org/10.1038/s41612-020-00149-4, 2020.

Kelder, T., Marjoribanks, T., Slater, L., Prudhomme, C., Wilby, R., Wagemann, J., and Dunstone, N.: An open workflow to gain insights about low-likelihood high-impact weather events from initialized predictions, Meteorological Applications, 29, e2065, https://doi.org/10.1002/met.2065, 2022a.

Kelder, T., Wanders, N., van der Wiel, K., Marjoribanks, T., Slater, L. J., l Wilby, R., and Prudhomme, C.: Interpreting extreme climate impacts from large ensemble simulations—are they unseen or unrealistic?, Environmental Research Letters, 17, 044 052, https://doi.org/10.1088/1748-9326/ac5cf4, 2022b.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

Kent, C., Dunstone, N., Tucker, S., Scaife, A. A., Brown, S., Kendon, E. J., Smith, D., McLean, L., and Greenwood, S.: Estimating unprecedented extremes in UK summer daily rainfall, Environmental Research Letters, 17, 014 041, https://doi.org/10.1088/1748-9326/ac42fb, 2022.

Lafon, T., Dadson, S., Buys, G., and Prudhomme, C.: Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods, International journal of climatology, 33, 1367–1381, https://doi.org/10.1002/joc.3518, 2013.

Li, C., Sinha, E., Horton, D. E., Diffenbaugh, N. S., and Michalak, A. M.: Joint bias correction of temperature and precipitation in climate model simulations, Journal of Geophysical Research: Atmospheres, 119, 13–153, https://doi.org/10.1002/2014JD022514, 2014.

Mardia, K. and Zemroch, P.: Measures of multivariate skewness and kurtosis, Journal of the Royal Statistical Society: Series C (Applied Statistics), 24, 262–265, https://doi.org/10.1007/978-981-19-3525-1_10, 1975.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B.: Climate Change 2021: The Physical Science Basis, https://doi.org/10.1017/9781009157896, contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2021.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, Earth system science data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Robin, Y., Vrac, M., Naveau, P., and Yiou, P.: Multivariate stochastic bias corrections with optimal transport, Hydrology and Earth System Sciences, 23, 773–786, https://doi.org/10.5194/hess-23-773-2019, 2019.

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., and Belcher, S. E.: High risk of unprecedented UK rainfall in the current climate, Nature communications, 8, 1–6, https://doi.org/10.1038/s41467-017-00275-3, 2017.

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Hardiman, S. C., Ren, H.-L., Lu, B., and Belcher, S. E.: Risk and dynamics of unprecedented hot months in South East China, Climate Dynamics, 52, 2585–2596, https://doi.org/10.1007/s00382-018-4281-5, 2019.

Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R 2 D 2) bias correction, Hydrology and Earth System Sciences, 22, 3175–3196, https://doi.org/10.5194/hess-22-3175-2018, 2018.

Wang, F. and Tian, D.: On deep learning-based bias correction and downscaling of multiple climate models simulations, Climate Dynamics, 59, 3451–3468, https://doi.org/10.1007/s00382-022-06277-2, 2022.

Wang, L., Hardiman, S., Bett, P., Comer, R., Kent, C., and Scaife, A.: What chance of a sudden stratospheric warming in the southern hemisphere?, Environmental Research Letters, 15, 104 038, https://doi.org/10.1088/1748-9326/aba8c1, 2020.

Wehner, M. F., Duffy, M. L., Risser, M., Paciorek, C. J., Stone, D. A., and Pall, P.: On the uncertainty of long-period return values of extreme daily precipitation, Frontiers in Climate, 6, 1343 072, https://doi.org/10.3389/fclim.2024.1343072, 2024.

Wu, G., Lv, P., Mao, Y., and Wang, K.: ERA5 precipitation over China: Better relative hourly and daily distribution than absolute values, Journal of Climate, 37, 1581–1596, https://doi.org/10.1175/JCLI-D-23-0302.1, 2024.

Yang, Y., Bacon, J., Ji, L., Ben Rached, N., Bradshaw, C. D., Davie, J. C. S., Crocker, T., and Bradshaw, L.: EMBCCA-UNSEEN, https://doi.org/10.5281/zenodo.21102535, 2026.

Zhang, L., Yu, X., Zhou, T., Zhang, W., Hu, S., and Clark, R.: Understanding and attribution of extreme heat and drought events in 2022: current situation and future challenges, Advances in Atmospheric Sciences, 40, 1941–1951, https://doi.org/10.1007/s00376-023-3171-x, 2023.

https://doi.org/10.5194/egusphere-2026-3932
Preprint. Discussion started: 21 September 2026

Zhang, L., Yan, Z., Huang, K., and Zhang, W.: Evaluation and statistical bias correction of ERA5-Land meteorological variables for a humid river basin in Southwest China, Scientific Reports, 15, 41 101, https://doi.org/10.1038/s41598-025-24942-4, 2025.

Zhu, Y., Li, Y., Zhou, X., Feng, W., Gao, G., Li, M., and Zheng, G.: Causes of the severe drought in Southwest China during the summer of 2022, Atmospheric Research, 303, 107 320, https://doi.org/10.1016/j.atmosres.2024.107320, 2024.