



Data-driven discovery of mechanisms underlying present and near-future precipitation changes and variability in Brazil

Márcia Talita A. Marques¹, Maria Luiza Kovalski¹, Gabriel M. P. Perez¹, Thomas C. M. Martin¹,
Edson L. S. Y. Barbosa¹, Pedro Augusto S. M. Ribeiro¹, and Roilan H. Valdes²

¹MeteoIA, São Paulo, Brazil

²Engie Brazil, São Paulo, Brazil

Correspondence: Gabriel M. P. Perez (gabriel@meteoia.com)

Received: 8 January 2024 – Discussion started: 9 February 2024

Revised: 3 February 2025 – Accepted: 28 February 2025 – Published: 17 July 2025

Abstract. Untangling the complex network of physical processes driving regional precipitation regimes in the present (1979–2014) and near-future climates (2020–2050) is fundamental to supporting a more robust scientific basis for decision-making in the water–energy–food nexus. We propose a data-driven mechanistic approach to (Goal 1) identify changes in and the variability of regional precipitation mechanisms and (Goal 2) reduce the ensemble spread of future projections by weighting and filtering models that satisfactorily represent these drivers in the present climate. Goal 1 is achieved by applying the partial least squares (PLS) technique, a two-sided variant of principal component analysis (PCA), on a reanalysis dataset and 30 simulations of the future climate submitted to the Coupled Model Intercomparison Project Phase 6 (CMIP6) to discover the links between global sea surface temperature (SST) and precipitation in Brazil. Goal 2 is achieved by selecting and weighting the future climate simulations from climate models that better represent the dominant modes discovered by the PLS in the present climate; with this subset of climate simulations, we produce precipitation change maps following the Intergovernmental Panel on Climate Change (IPCC) Working Group I (WGI) methodology. The main mechanistic link discovered by the technique is that the generalised warming of the oceans promotes a suppression of precipitation in northeastern and southeastern Brazil, possibly mediated by the intensification of the Hadley circulation. We show that this pattern of precipitation suppression is stronger in the near-future precipitation change maps produced using our methodology. This demonstrates that a reduction in epistemic uncertainty is achieved after we select models that skilfully

represent these mechanisms in the present climate. Therefore, the approach is capable of supporting both a quantitative analysis of regional changes and the construction of storylines supported by mechanistic evidence.

1 Introduction

Information about near-future regional precipitation change is crucial for planning and managing critical infrastructure, such as hydropower plants, water reservoirs and city planning. Unpreparedness for changes and variations in regional precipitation regimes may lead to disruption in water–food–energy supply chains, as well as avoidable deaths and damages by flooding and landslides. Although there is a degree of certainty about global precipitation changes (Shepherd et al., 2018), such as the intensification of the hydrological cycle, a current major challenge in climate change science is informing planners and decision-makers about regional changes within the critical time frame of the next 3 decades.

Within this time frame, the two main sources of uncertainty in regional precipitation changes are model uncertainty and internal variability (Hawkins and Sutton, 2011). Uncertainty due to the internal variability of the climate system is impossible to reduce and is aleatoric and related to the chaotic nature of the system (Shepherd, 2019). Model uncertainty, on the other hand, is epistemic in nature and stems from our limited knowledge of Earth's climate system and from the challenges in translating this system into computer models (Shepherd, 2019). Currently, there are 131 available models on the Coupled Model Intercomparison Project Phase

6 (CMIP6) database, each representing Earth's climate with a range of parameterisations and numerical modelling strategies. From the full set of models, 29 are available for the relevant experiments in this study.

In this study, we aim to reduce the epistemic uncertainty in regional precipitation changes in Brazil through a data-driven process-based methodology of model selection and weighting. The method discovers the relationships between sea surface temperature and precipitation in Brazil and evaluates the capability of CMIP6 models to reproduce these precipitation mechanisms in the present climate. Later, the best models are selected and weighted to produce refined precipitation maps. Due to the process-based nature of the method, it is also possible to isolate mechanisms and draw storylines of plausible futures. The paper answers the following questions:

- What are the spatiotemporal links between global sea surface temperature (SST) and regional precipitation change and variability in Brazil? Many patterns have been identified in the literature (Grimm et al., 2000; Coelho et al., 2002), but here we choose to use a supervised machine learning (ML) approach to systematically identify and quantify their importance.
- Can we take advantage of these mechanisms to filter CMIP6 simulations and reduce the epistemic uncertainty in regional precipitation changes?
- What are the projections for precipitation in Brazil over the next 30 years based on a mechanistic filtering and weighting procedure?

2 Materials and methods

2.1 Data-driven discovery of precipitation mechanisms

To discover the underlying mechanisms linking SST spatiotemporal variability and regional precipitation in Brazil, we employ a data-driven dimensionality reduction method known as partial least squares (PLS), adapted to a latitude–longitude grid, which has recently been shown to successfully identify circulation mechanisms leading to precipitation (Perez et al., 2022).

The PLS method identifies pairs of latent variable vectors ξ and ω that maximise the information present in $X^T Y$. This means finding latent variables that represent the maximum covariance between X and Y , where X and Y represent two arrays of SST and precipitation, respectively; the rows of X and Y represent the monthly averaged temporal samples; and the columns represent the spatial lat–long grid points. The more familiar principal component analysis (PCA) can be seen as a special case where $X = Y$. The initial set (or mode) of latent variables is determined through the following covariance (Eq. 1):

$$\text{Cov}(\xi_1, \omega_1) = \max_{\|u\|=\|v\|=1} \text{Cov}(Xu, Yv), \quad (1)$$

where u and v are temporally invariant arrays of loadings. In contrast to PCA, PLS yields a pair of loading matrices per component rather than a single loading matrix; the first pair of loading matrices is the one in which the corresponding latent vectors ξ and ω are the most correlated. The following modes are found through repeating the process on the residuals of each preceding pair.

The interpretation of PLS results should always consider scores and loadings concurrently. A positive loading correlation, coupled with a positive trend in the scores, indicates an increase in signal strength over time. Conversely, when loadings exhibit the same signal but are associated with a negative trend in scores, this suggests a decrease in signal intensity. When evaluating the relationship between two loadings, we first observe how the loading patterns are linked through the score time series. This connection allows us to infer the response of the Y pattern from the X pattern. For instance, if the SST signal is negative at a particular location while the score is positive, this indicates a negative association with the corresponding precipitation loading pattern; the score may change sign with time, indicating a change in sign in the association between the two variables. For example, in the El Niño 3.4 region, positive SST may be positively associated with positive precipitation anomalies in southern Brazil, while negative SST anomalies may be associated with negative precipitation anomalies. A detailed explanation of the PLS method can be found in Wegelin (2000).

2.2 Present and future climate datasets

The PLS method was applied to two kinds of climate datasets: first, to present climate data from AMIP experiments and reanalysis and, second, to future climate simulations. In the AMIP experiments, atmospheric models are forced by prescribed sea surface temperatures, while radiative forcings and land use are kept constant. This approach helps identify model errors that arise from interactions within the atmosphere, the land surface, or between these components (Eyring et al., 2016). The subsections below describe the methodologies and data behind the present and future climate results.

2.2.1 Present climate (AMIP)

The first step was to establish a transfer function linking SST and precipitation month-to-month co-variability using the PLS technique for the reanalysis and atmosphere-only experiments. The goal is to identify models that accurately represent the transfer function identified in the reanalysis in the present climate. To achieve this, we employ precipitation data derived from the ERA5 reanalysis (Hersbach and Dee, 2016), in addition to precipitation data from 29 AMIP models from the Coupled Model Intercomparison Project Phase 6 (CMIP6), as outlined in Table 1. Before the PLS technique was employed, the ERA5 precipitation data un-

Table 1. CMIP6 simulations, their native resolutions, vertical levels and source institutions.

Model	Horizontal resolution	Vertical levels	Variant label	Institution
ACCESS-CM2	$1.875^{\circ} \times 1.25^{\circ}$	85	rlilplf1	CSIRO
ACCESS-ESM1-5	$1.875^{\circ} \times 1.25^{\circ}$	38	rlilplf1	CSIRO
BCC-CSM2-MR	$2.81^{\circ} \times 2.81^{\circ}$	46	rlilplf1	BCC
CAMS-CSM1-0	$1^{\circ} \times 1^{\circ}$	31	rlilplf1	CAMS
CanESM5	$2.81^{\circ} \times 2.81^{\circ}$	49	rlilplf1	CCCma
CESM2-WACCM	$0.9^{\circ} \times 1.25^{\circ}$	70	rlilplf1	NCAR
CIESM	$1^{\circ} \times 1^{\circ}$	30	rlilplf1	THU
CMCC-CM2-SR5	$1^{\circ} \times 1^{\circ}$	30	rlilplf1	CMCC
CNRM-CM6-1	$1.4^{\circ} \times 1.4^{\circ}$	91	rlilplf2	CNRM-CERFACS
CNRM-CM6-1-HR	$1.4^{\circ} \times 1.4^{\circ}$	91	rlilplf2	CNRM-CERFACS
CNRM-ESM2-1	$1.4^{\circ} \times 1.4^{\circ}$	91	rlilplf2	CNRM-CERFACS
EC-Earth3-CC	$0.7^{\circ} \times 0.7^{\circ}$	91	rlilplf1	EC-Earth-Consortium
EC-Earth3-Veg	$0.7^{\circ} \times 0.7^{\circ}$	91	rlilplf1	EC-Earth-Consortium
EC-Earth3-Veg-LR	$1.1^{\circ} \times 1.1^{\circ}$	62	rlilplf1	EC-Earth-Consortium
FGOALS-f3-L	$1^{\circ} \times 1^{\circ}$	32	rlilplf1	IAP/CAS
FGOALS-g3	$2^{\circ} \times 2^{\circ}$	26	rlilplf1	IAP/CAS
GFDL-CM4	$1^{\circ} \times 1^{\circ}$	33	rlilplf1	NOAA-GFDL
GFDL-ESM4	$1^{\circ} \times 1^{\circ}$	49	rlilplf1	NOAA-GFDL
IITM-ESM	$2^{\circ} \times 2^{\circ}$	64	rlilplf1	CCCR-IITM
INM-CM4-8	$2^{\circ} \times 1.5^{\circ}$	21	rlilplf1	INM
INM-CM5-0	$2^{\circ} \times 1.5^{\circ}$	73	rlilplf1	INM
IPSL-CM6A-LR	$2.5^{\circ} \times 1.3^{\circ}$	79	rlilplf1	IPSL
KACE-1-0-G	$1.9^{\circ} \times 1.3^{\circ}$	85	rlilplf1	NIMS-KMA
MIROC6	$1.4^{\circ} \times 1.4^{\circ}$	81	rlilplf1	MIROC
MPI-ESM1-2-HR	$0.93^{\circ} \times 0.93^{\circ}$	95	rlilplf1	MPI-M
MPI-ESM1-2-LR	$1.9^{\circ} \times 1.9^{\circ}$	47	rlilplf1	MPI-M
MRI-ESM2-0	$1.125^{\circ} \times 1.125^{\circ}$	80	rlilplf1	MRI
NESM3	$1.9^{\circ} \times 1.9^{\circ}$	47	rlilplf1	NUIST
NorESM2-LM	$2^{\circ} \times 2^{\circ}$	32	rlilplf1	NCC
TaiESM1	$1.25^{\circ} \times 0.9^{\circ}$	30	rlilplf1	AS-RCEC

derwent systematic error correction using observations from the Global Precipitation Climatology Project (GPCP; Adler et al., 2018) as a reference through the quantile-mapping method, which adjusts probability distributions by individually matching each quantile to the respective quantile of the reference dataset (Themeßl et al., 2011). GPCP data have a shorter time period (from 1979) when compared to ERA5 data, which have been available since 1950. Since we aimed to investigate interannual and interdecadal variability and climate change, we chose the longer dataset. Each precipitation dataset was conservatively gridded to a regular $1^{\circ} \times 1^{\circ}$ lat–long grid in a monthly temporal resolution between 1979 and 2014. SST data were obtained from the COBE dataset, produced by the Japan Meteorological Agency (Hirahara et al., 2014), which has long temporal availability and is based on observations.

The models listed above, through their computational representations of the atmosphere, choices of parameterisation, vertical levels, etc., provide unique numerical representations of the physical climate system. Each of these representations has a distinct level of skill in simulating the mechanisms of

precipitation variability and changes in Brazil. Therefore, we rank and select the models with higher performance to represent the SST–precipitation transfer function revealed by the PLS analysis.

This ranking is based on the normalised root mean square error (NRMSE), which is obtained by comparing the PLS scores and loadings between each model and those derived from the ERA5 reanalysis. Specifically, the RMSE is computed between the scores and loadings from each model and the corresponding scores and loading from ERA5. Then, these RMSEs are collectively normalised, yielding the NRMSEs, to ensure consistent scaling across all comparisons. Models with NRMSE values below 0.6 in at least two of the first four PLS components are then selected. This threshold was chosen so that a representative number of models is kept in the ensemble: 15 models compose the selected subset. A more strict NRMSE threshold would lead to a substantially smaller subset and reduce the statistical robustness of the results, while a more relaxed threshold would be less effective in reducing the epistemic uncertainty by keeping models that represent the desired mechanisms poorly. These selected

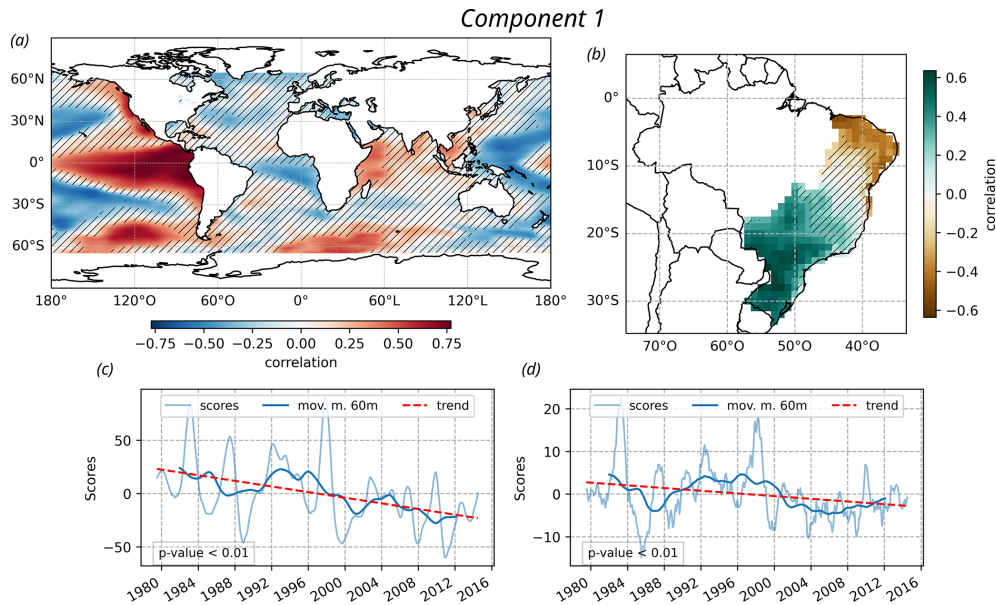


Figure 1. First component of the PLS methodology applied using monthly precipitation data from ERA5 and SST data from COBE between 1979 and 2014. **(a, b)** The spatial maps represent the loading matrices, where the hatching represents areas where the statistical confidence on the sign of the anomaly is lower than 95 %. **(c, d)** The time series represents the scores, and the p values indicate the statistical significance of the results.

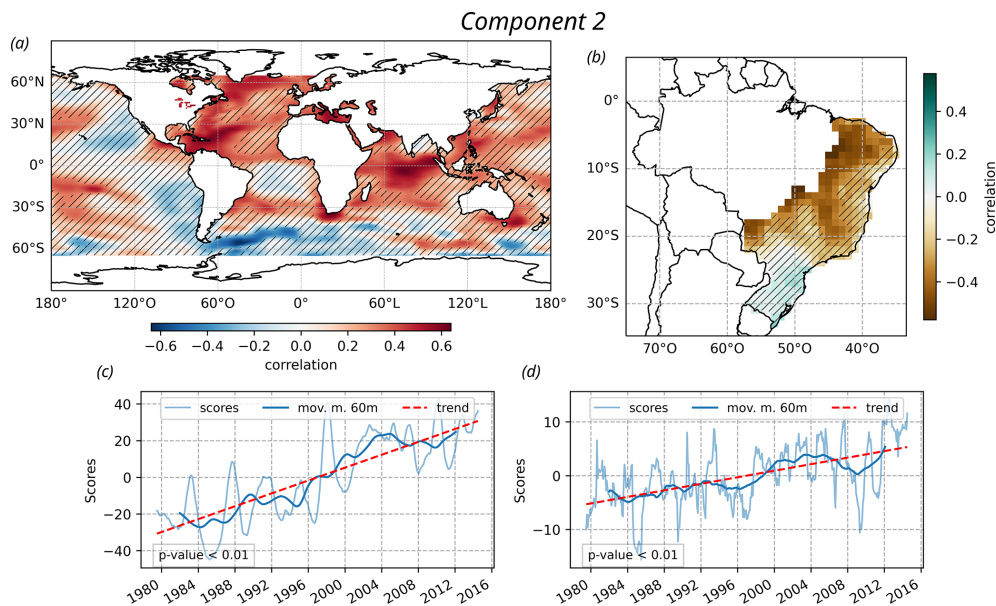


Figure 2. Second component of the PLS methodology applied using monthly precipitation data from ERA5 and SST data from COBE between 1979 and 2014. **(a, b)** The spatial maps represent the loading matrices, where the hatching represents areas where the statistical confidence on the sign of the anomaly is lower than 95 %. **(c, d)** The time series represents the scores, and the p values indicate the statistical significance of the results.

models are singled out as more reliably representing mechanisms that cause precipitation in Brazil, while the rest are discarded for the remaining analysis.

After the model ranking and selection step, we provide a set of weights that is used later for model averaging. This set

of weights is found by multiplying the inverse of the NRMSE by the importance of each PLS component; this is done so that models that perform well in representing more relevant mechanisms are favoured during the model-pooling step. The importance of each PLS component is quantified by the co-

Table 2. List of selected models and their weights represented as a percentage of their contribution to the ensemble mean.

Model	Components	Weight (%)
CAMS-CSM1-0	1, 2, 4	7.76
CNRM-ESM2-1	1, 3, 4	7.73
GFDL-ESM4	2, 4	7.59
BCC-CSM2-MR	1, 2, 4	7.37
EC-Earth3-CC	1, 2	7.11
EC-Earth3-Veg-LR	1, 2	7.08
EC-Earth3-Veg	2, 4	6.83
IPSL-CM6A-LR	2, 3, 4	6.69
KACE-1-0-G	1, 2	6.61
CNRM-CM6-1-HR	2, 3, 4	6.56
MPI-ESM1-2-HR	1, 4	6.28
CMCC-CM2-SR5	1, 2	6.19
FGOALS-f3-L	2, 3	6.18
MIROC6	1, 4	5.94
CESM2-WACCM	1, 4	4.08

efficient of determination (r^2) of the reconstructed precipitation using only that component and the original ERA5 precipitation.

2.2.2 Future climate

We employ the same PLS methodology on future climate simulations under the SSP2-2.45 scenario between 2020 and 2050; in this near-future temporal range, we do not expect the choice of scenario to influence the results because scenario uncertainty in regional precipitation changes only becomes relevant in later decades (Hawkins and Sutton, 2011).

Finally, the effectiveness of this methodology in reducing the uncertainty in near-future precipitation changes in the CMIP6 ensemble is assessed by comparing the uncertainty in all CMIP6 models listed in Table 1 with the uncertainty in the subset of models selected by our methodology. The climate change signal was computed for each grid cell by calculating the ratio (in percentage) between the anomaly of the ensemble mean climatologies of the SSP2-4.5 scenario for the years 2020–2050 and the historical period of 1979–2014, divided by the historical period. To assess the robustness of the models, we apply the procedure adopted by the Intergovernmental Panel on Climate Change (IPCC), as outlined in its Sixth Assessment Report, made available through the Interactive Atlas developed by Working Group I (WGI) (Gutiérrez et al., 2021). This approach determines the robustness of climate change signals based on a strong model consensus, highlighting where at least 80 % of the models agree on the sign of the predicted changes.

3 Results and discussion

In this section, we present the results of the analysis for the present and future climates, discussing the underlying precipitation mechanisms in reanalysis and model data. We also discuss the reduction in epistemic uncertainty in regional precipitation changes obtained through the selection of models that skilfully represent precipitation mechanisms in the present climate. In all PLS analyses, the Legal Amazon area was cropped out; this is because precipitation in the Amazon region presents significantly higher variability in magnitude, dominating the results and obscuring patterns in other areas that are also socioeconomically relevant.

3.1 Precipitation mechanisms in the present climate (1979–2014)

In the present climate, the first PLS loading matrix of the SST reveals a prominent positive pattern in the central Pacific Ocean that aligns with the region dominated by the El Niño–Southern Oscillation (ENSO) phenomenon (Fig. 1a). This ENSO-like pattern with high statistical significance (unhatched area) extends from the west coast of South America to the Maritime Continent in the equatorial region, surrounded by a pattern of opposite signal. There is a negative trend in Fig. 1c scores, indicating a change in sign in the patterns of Fig. 1a. This suggests that in the first half of the time series, El Niño conditions were dominant, while in the second half of the time series, La Niña conditions were dominant. The associated PLS loading matrix for precipitation shows a significant positive correlation in southern Brazil and a negative correlation in northeastern Brazil (Fig. 1b). The time series of the associated scores do not show a strong linear trend, reinforcing that this PLS mode is more closely associated with a natural variability mechanism like ENSO than with climate change (Fig. 1d).

The global warming trend can explain the mostly positive SST loading matrix and the increasingly positive score time series of the second PLS component (Fig. 2a, c). This warming oceanic pattern is linked to a precipitation reduction in most of southeastern and northeastern Brazil (Fig. 2b, d). A possible explanation for this precipitation suppression is the expansion of the Hadley cell under climate change (Lu et al., 2007; Grise and Davis, 2020) and, consequently, the restriction of the equatorward motion of extratropical cyclones and their fronts, which are important precipitation mechanisms in southeastern Brazil (Perez et al., 2021). Perez et al. (2022) have shown that a temporary intensification of the Hadley circulation during positive North Atlantic Oscillation (NAO) events leads to precipitation suppression in southeastern Brazil.

For the third and fourth components (Figs. S1 and S2 in the Supplement), we observe more divergence in the results. Despite this, we consider these components in our evaluation due to their assigned weights and the potential for provid-

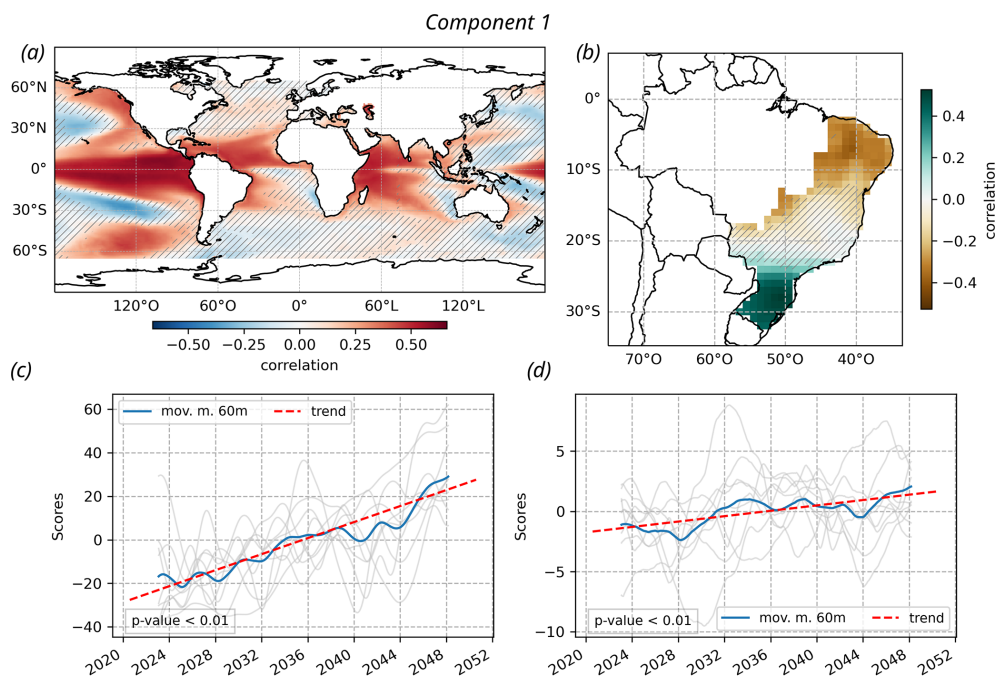


Figure 3. First component of the PLS methodology applied using monthly precipitation data from CMIP6 models under the SSP2-4.5 scenario, listed in Table 2, between 2020 and 2050. The spatial maps represent the loading matrices, and the time series represent the scores. The regions with hatching indicate areas of uncertainty with < 80 % agreement in the sign change among the models. The p values indicate the statistical significance of the results.

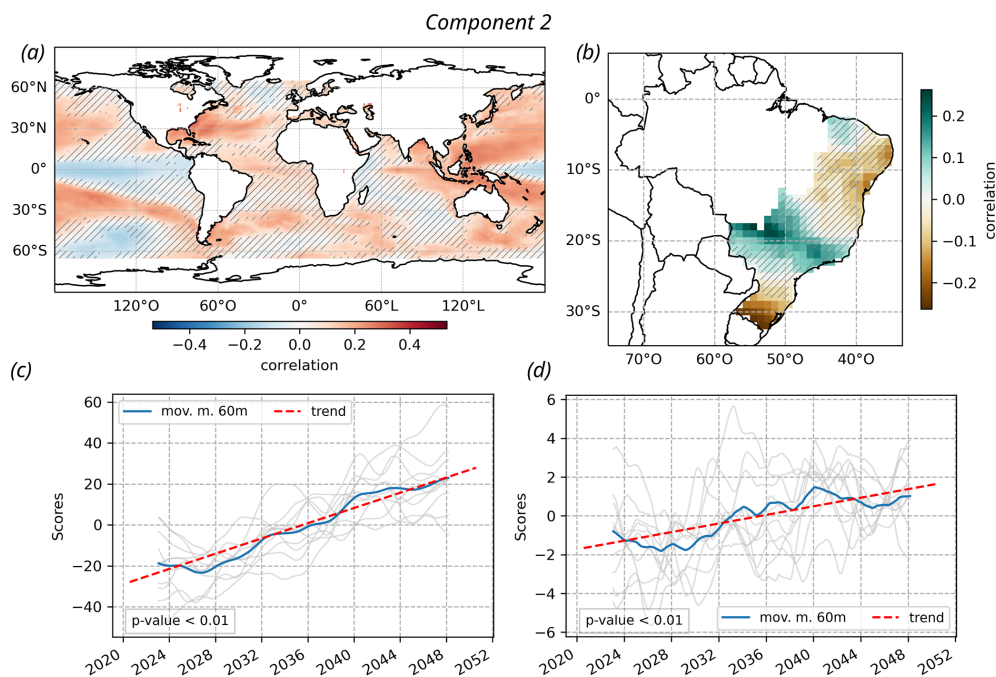


Figure 4. Second component of the PLS methodology applied using monthly precipitation data from CMIP6 models under the SSP2-4.5 scenario, listed in Table 2, between 2020 and 2050. The spatial maps represent the loading matrices, and the time series represent the scores. The regions with hatching indicate areas of uncertainty with < 80 % agreement in the sign change among the models. The p values indicate the statistical significance of the results.

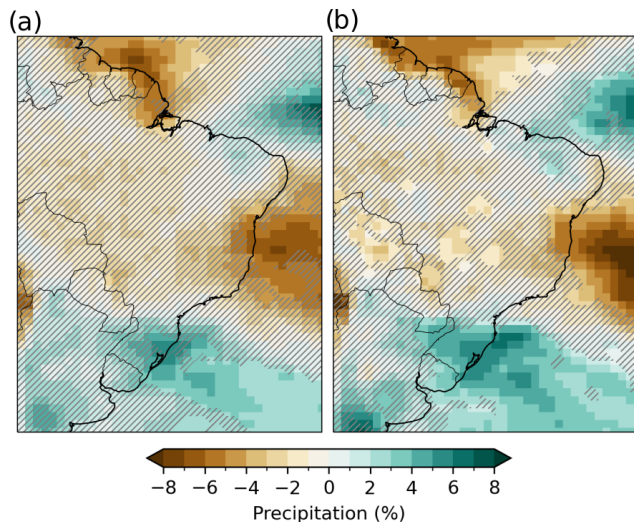


Figure 5. Percentage precipitation changes in 2020–2050 relative to 1979–2014 based on all assessed models, as listed in Table 1, (a) and the percentage changes based on the selected models listed in Table 2 (b) from CMIP6 under the SSP2-4.5 scenario. The regions with hatching indicate areas of uncertainty with < 80 % agreement in the sign change among the models.

ing additional representation of driving mechanisms beyond the first and second components. Moreover, these third and fourth components might reveal important patterns that affect precipitation in Brazil, especially the Atlantic SST variability (Hastenrath and Greischar, 1993; Yoon and Zeng, 2010; Perez et al., 2022).

Through the analysis of the PLS components in the present climate datasets, we are able to select and rank the models based on their performance in reproducing these components. The model selection is based on a threshold of $\text{NRMSE} < 0.6$, and the individual model weights are based on the inverse of the average NRMSE among the PLS components scaled by the importance of each component, as described in the “Materials and methods” section. Table 2 lists the selected models and their respective weights along with the components these models skilfully represent, later employed to construct the weighted ensemble mean in the future climate section. These models selected through our approach are those that showed better performance in the task of simulating the impacts of precipitation in Brazil. This way, for example, the high weight of GFDL-ESM4 indicates that this model performs well in representing the overall components more accurately when compared to other models. While it is true that component 1 relates to the ENSO dynamics, the overall evaluation takes into account components that represent other important forcings of the Brazilian precipitation regime. For instance, the Atlantic SST variability drives the Brazilian precipitation variability in the Amazon (Yoon and Zeng, 2010), northeastern Brazil (Hastenrath and Greischar, 1993) and subtropical regions (Perez et al., 2022). Further-

more, by including multiple components in the analysis, we acknowledge that climate dynamics are multifaceted, and a comprehensive evaluation should account for more than just the primary modes of variability like ENSO. This approach rests on the importance of a holistic evaluation of model performance across various components rather than focusing solely on the primary modes. The first and second PLS components of all models used are presented in the Supplement (Figs. S3–S62).

3.2 Precipitation mechanisms in the future climate (2020–2050)

The oceanic mechanisms driving precipitation in Brazil in the future climate (2020–2050) are discovered by applying the PLS methodology in CMIP6 future climate simulations (Figs. 3 and 4). Figure 3 shows the first PLS component and Fig. 4 the second PLS component; for each component, only models that performed well ($\text{NRMSE} < 0.6$) in the present climate are considered. The spatial maps show the average loading matrices of the model ensemble, where each model is weighed by its skill in the present climate (Table 2); the hatched areas represent regions where at least 80 % of the models disagree on the sign of the loading matrix.

The first component shows a strong El Niño-like pattern in the Central Pacific, similarly to what is found in the present climate (Fig. 3a). However, unlike the present climate analysis, this El Niño-like component shows a strong linear trend in the time series of scores (Fig. 3c), suggesting that the climate models mix the natural variability of the ENSO phenomenon and anthropogenic global warming; this warming trend can also be seen in the increasingly positive patterns in the tropical Atlantic and Indian oceans. The impact of this warming trend in Brazilian regional precipitation is a wetting pattern in southern Brazil and a drying pattern in northeastern Brazil, interfaced by a large region of uncertainty (Fig. 3b).

The second component illustrates a generalised warming trend in most regions of model agreement (Fig. 4a, c). This component impacts precipitation in Brazil through a drying trend in the southernmost border of the country and a wetting trend in the southeastern area. Some coastal areas in northeastern Brazil are significantly affected by a drying trend (Fig. 4b, d). Although the linear trend was observed in most models (Figs. 3c, d and 4c, d), it becomes clearer and more robust in the model subset; this reflects how sifting models in a mechanistic approach helps reduce the epistemic uncertainties.

3.3 Future climate precipitation changes and uncertainty reduction

While the analysis of individual PLS components may support storyline approaches and mechanistic understanding, a quantitative precipitation change map is often required by decision-making bodies. With that in mind, we provide an

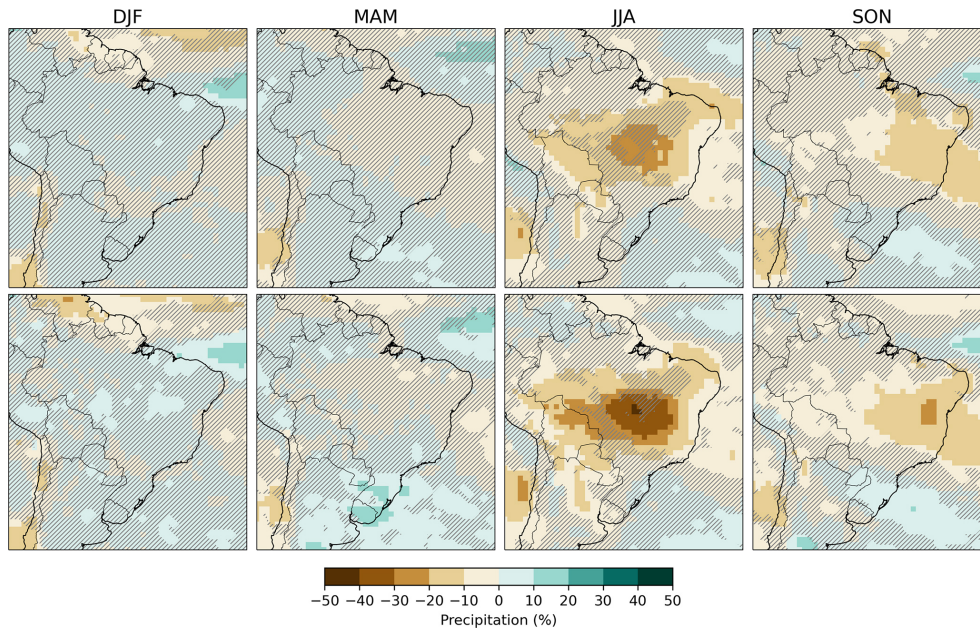


Figure 6. Seasonal percentage precipitation changes in 2020–2050 relative to 1979–2014 based on all assessed models, as listed in Table 1 (top row), and the percentage changes based on the selected models listed in Table 2 (bottom row) from CMIP6 under the SSP2-4.5 scenario. The regions with hatching indicate areas of uncertainty with < 80 % agreement in the sign change among the models.

uncertainty map based on the methodology employed by the IPCC in its Sixth Assessment Report (Fig. 5). Here, we focus on the percentage of projected changes in 2020–2050 relative to 1979–2014. The hatching highlights regions where there is a significant lack of consensus, with at least 80 % of the models analysed showing non-concordance, similar to the PLS uncertainty maps shown in the previous section.

Figure 5a shows the ensemble mean of the future precipitation changes using all CMIP6 models, listed in Table 1, while Fig. 5b uses the mean of the subset of models in Table 2 weighted by their skill in simulating precipitation mechanisms in the present climate (Fig. 5b). First, we notice that the reduction in epistemic uncertainty by the proposed methodology is revealed by stronger anomalies and fewer hatched areas. This is further supported by the computed standard deviation values, which show a reduction from 0.24 for the full ensemble set to 0.21 for the reduced subset – a decrease of 12.5 %. Particularly, the South Atlantic Subtropical High (SASH) shows stronger negative anomalies, suggesting a trend towards drier conditions in the region via an intensification of the Hadley cell descending branch. Moreover, the positive changes in southern Brazil have increased after the application of the methodology; this enhanced dipole between the SASH and southern Brazil is consistent with the mechanism of restriction of cold fronts revealed by the PLS in the present climate and is discussed in Sect. 3a. In other words, selecting and weighting models that reproduce important precipitation mechanisms in the present climate have increased the clarity of what may happen in the region in

the near-future climate. Additional uncertainty maps based on alternative model selections are provided in the Supplement (Figs. S63–S64).

Figure 6 shows the future precipitation changes broken down by season based on all models listed in Table 1 and only using the models selected by the methodology (Table 2). A noticeable reduction in uncertainty across all seasons is evident when comparing the hatched areas using all models versus only using the selected models, underscoring the success of our process-based model selection methodology in enhancing our confidence in regional climate projections. The period from December to May corresponds to the rainy season, characterised by a prevalence of uncertainties; this is in agreement with Bazzanella et al. (2023) and Firpo et al. (2022), who also indicate that CMIP6 models perform better in the dry season than in the wet season.

From June to November the central and northeastern regions exhibit a clear drying pattern. In JJA, in particular, precipitation in most of Brazil is largely driven by cold fronts, which, as previously discussed, can be restrained in higher latitudes if the SASH is intensified. In SON, we expect an intensified SASH to also contribute to a later onset of the rainy season. This drying pattern in JJA and SON is intensified in the subset of selected models. This is unsurprising, since the SASH subsidence associated with an intensification of the Hadley circulation is one of the mechanisms discovered by the PLS analysis in the present climate and is used to select the best-performing models.

4 Summary and conclusions

This study aims to reduce the epistemic uncertainty in regional precipitation changes in Brazil through a data-driven process-based methodology of model selection and weighting. To achieve this, we first employ the methodology to discover the main precipitation drivers in the present climate (1979–2014) in a reanalysis dataset (Sect. 3a), revealing that the El Niño and the generalised warming of the oceans are linked to significant precipitation impacts in Brazil (Figs. 1 and 2). A distinct positive linear trend in the global warming component is linked to drying across most of northeastern and southeastern Brazil. We propose that the linking mechanism between these SST and precipitation patterns is the intensification of the Hadley circulation (Hu and Fu, 2007) and, consequently, of the subsidence at the South Atlantic Subtropical High (Carvalho et al., 2011).

The same methodology is then applied to CMIP6 present climate simulations (Table 1) to evaluate the capability of CMIP6 models to simulate these precipitation drivers, thus creating a process-based model selection and weighting approach to underpin the future climate analysis. From a total of 30 models, we select 15 models that are capable of simulating at least two (Table 2) of the main regional precipitation drivers.

The mechanism discovery methodology is then applied to the near-future (2020–2050) climate simulations of the selected models. We find that an ENSO-like pattern, tied to a generalised warming of the tropical oceans, is linked to an increase in precipitation in southern Brazil and a decrease in northeastern Brazil (Figs. 3 and 4), consistent with the present climate indication of an intensification of the Hadley circulation. This mechanistic view of regional precipitation changes can underpin the development of storylines in future studies to support decision-making bodies in the water–energy–food nexus.

We go further to provide a quantitative view of regional precipitation changes based on the IPCC WG1 approach, contrasting the uncertainty in precipitation changes using 30 CMIP6 models versus using the 15 selected models. We show that the approach increased model agreement, particularly in southern Brazil and the SASH region. In the next 30 years (Fig. 6), a noticeable reduction in uncertainty across all seasons is evident mostly from June to November. This period is characterised by a clear drying pattern due to the strengthening of SASH, intensified within the subset of selected models, which leads to a suppression of precipitation in northeastern and southeastern Brazil, possibly delaying the rainy season in these regions.

Our methodology of model selection and weighting considers precipitation drivers rather than simply comparing CMIP6 model precipitation with observations. By selecting and weighting models mechanistically, we achieve a reduction in the epistemic uncertainty in precipitation changes in Brazil in the CMIP6 ensemble. The method is based on the

discovery of statistical relationships between SST patterns and precipitation through the PLS and the assumption that models with an accurate representation of these statistical relationships better represent atmospheric processes leading to precipitation. Considering that the atmospheric flow is the medium connecting SST and precipitation and the statistical significance of the PLS loadings, we believe this assumption to be robust. However, as with any data-driven methodology, there could be instances where confounding factors may influence the results; this highlights the need for other mechanistic approaches capable of isolating rainfall mechanisms individually, such as atmospheric rivers, convergence zones, hurricanes and fronts (Catto et al., 2015; Franco-Díaz et al., 2019; Perez et al., 2024). Moreover, studies addressing the sensitivity of the results to the thresholds of model selection could potentially increase the robustness of future precipitation change predictions (Figs. 5 and 6).

Code availability. The code supporting the findings of this study is not publicly available due to institutional confidentiality.

Data availability. The ERA5 dataset used in this study is provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) and is available through the Copernicus Climate Data Store (<https://cds.climate.copernicus.eu/>, ECMWF, 2024). The COBE Sea Surface Temperature data is provided by the NOAA Physical Sciences Laboratory (PSL), Boulder, Colorado, USA, and is available at <https://psl.noaa.gov/> (NOAA Physical Sciences Laboratory, 2024). The ensemble model simulations used in this study are part of the Coupled Model Intercomparison Project Phase 6 (CMIP6) and are available at https://esgf.ceda.ac.uk/thredds/catalog/esg_cmip6/CMIP6/catalog.html (World Climate Research Programme (WCRP), 2020).

Supplement. The supplement related to this article is available online at <https://doi.org/10.5194/wcd-6-757-2025-supplement>.

Author contributions. GMPP contributed with the conceptualisation of the method and led MTAM and MLK in performing the computations and analysing the data. TCMM, ELSYB and PASMR revised the computational code and performed additional analyses. RHV provided analytical feedback. MLK and GMPP wrote the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes ev-

ery effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. This research results from the R&D project developed by MeteoIA for Engie Brazil, funded by ANEEL under project no. PD-00403-0054/2022.

Financial support. This research has been supported by the Agência Nacional de Energia Elétrica (grant no. PD-00403-0054/2022).

Review statement. This paper was edited by Erich Fischer and reviewed by Peter Pfleiderer and Elena Saggioro.

References

- Adler, R. F., Sapiiano, M. R. P., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R., and Shin, D. B.: The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation, *Atmosphere*, 9, 138, <https://doi.org/10.3390/atmos9040138>, 2018.
- Bazzanella, A. C., Dereczynski, C., Luiz-Silva, W., and Regoto, P.: Performance of CMIP6 models over South America, *Clim. Dyn.*, 62, 1501–1516, <https://doi.org/10.1007/s00382-023-06979-1>, 2023.
- Carvalho, L. M. V., Jones, C., Silva, A. E., Liebmann, B., and Silva Dias, P. L.: The South American Monsoon System and the 1970s climate transition, *Int. J. Climatol.*, 31, 1248–1256, <https://doi.org/10.1002/joc.2147>, 2011.
- Catto, J., Jakob, C., and Nicholls, N.: Can the CMIP5 models represent winter frontal precipitation?, *Geophys. Res. Lett.*, 42, 8596–8604, <https://doi.org/10.1002/2015gl066015>, 2015.
- Coelho, C. A. S., Uvo, C. B., and Ambrizzi, T.: Exploring the impacts of the tropical Pacific SST on the precipitation patterns over South America during ENSO periods, *Theor. Appl. Climatol.* 71, 185–197, <https://doi.org/10.1007/s007040200004>, 2002.
- ECMWF: ERA5 Reanalysis Data, Copernicus Climate Data Store, <https://cds.climate.copernicus.eu/> (last access: 13 June 2025), 2024.
- World Climate Research Programme (WCRP): CMIP6 Ensemble Climate Model Simulations, Earth System Grid Federation (ESGF) [data set], https://esgf.ceda.ac.uk/thredds/catalog/esg_cmip6/CMIP6/catalog.html (last access: 13 June 2025), 2020.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- Firpo, M.Â. F., Guimarães, B. S., Dantas, L. G., Silva, M. G. B., Alves, L. M., Chadwick, R., Llopart, M. P., and Oliveira, G. S.: Assessment of CMIP6 models' performance in simulating present-day climate in Brazil, *Front. Clim.*, 4, 948499, <https://doi.org/10.3389/fclim.2022.948499>, 2022.
- Franco-Díaz, A., Klingaman, N. P., Vidale, P. L., Guo, L., and Demory, M. E.: The contribution of tropical cyclones to the atmospheric branch of Middle America's hydrological cycle using observed and reanalysis tracks, *Clim. Dyn.*, 53, 6145–6158, <https://doi.org/10.1007/s00382-019-04920-z>, 2019.
- Grimm, A. M., Vicente, R. B., and Doyle, M. E.: Climate variability in southern South America associated with El Niño and La Niña events, *J. Climate*, 13, 35–58, [https://doi.org/10.1175/1520-0442\(2000\)013<0035:CVISSA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<0035:CVISSA>2.0.CO;2), 2000.
- Grise, K. M. and Davis, S. M.: Hadley cell expansion in CMIP6 models, *Atmos. Chem. Phys.*, 20, 5249–5268, <https://doi.org/10.5194/acp-20-5249-2020>, 2020.
- Gutiérrez, J. M., Jones, R. G., Narisma, G. T., Alves, L. M., Amjad, M., Gorodetskaya, I. V., Grose, M., Klutse, N. A. B., Krakovska, S., Li, J., Martínez-Castro, D., Mearns, L. O., Mernild, S. H., Ngo-Duc, T., van den Hurk, B., and Yoon, J.-H.: Atlas. In *Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: 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., Cambridge University Press, Interactive Atlas, <http://interactive-atlas.ipcc.ch/> (last access: 13 June 2025), in press, 2021.
- Hastenrath, S. and Greischar, L.: Circulation mechanisms related to northeast Brazil rainfall anomalies, *J. Geophys. Res.-Atmos.*, 98, 5093–5102, <https://doi.org/10.1029/92JD02646>, 1993.
- Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, *Clim. Dyn.*, 37, 407–418, <https://doi.org/10.1007/s00382-010-0810-6>, 2011.
- Hersbach, H. and Dee D.: ERA5 reanalysis is in production, *ECMWF Newsl.*, 147, 5–6, 2016.
- Hirahara, S., Ishii, M., and Fukuda, Y.: Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty, *J. Climate*, 27, 57–75, <https://doi.org/10.1175/JCLI-D-12-00837.1>, 2014.
- Hu, Y. and Fu, Q.: Observed poleward expansion of the Hadley circulation since 1979, *Atmos. Chem. Phys.*, 7, 5229–5236, <https://doi.org/10.5194/acp-7-5229-2007>, 2007.
- Lu, J., Vecchi, G. A., and Reichler, T.: Expansion of the Hadley cell under global warming, *Geophys. Res. Lett.*, 34, L06805, <https://doi.org/10.1029/2006GL028443>, 2007.
- NOAA Physical Sciences Laboratory: COBE Sea Surface Temperature Data, NOAA PSL, Boulder, CO, USA, <https://psl.noaa.gov/> (last access: 13 June 2025), 2024.
- Perez, G. M. P., Vidale, P. L., Klingaman, N. P., and Martin, T. C. M.: Atmospheric convergence zones stemming from large-scale mixing, *Weather Clim. Dynam.*, 2, 475–488, <https://doi.org/10.5194/wcd-2-475-2021>, 2021.
- Perez, G. M. P., Vidale, P. L., Dacre, H., and García-Franco, J. L.: Using a Synoptic-Scale Mixing Diagnostic to Explain Global Precipitation Variability from Weekly to Interannual Time Scales, *J. Climate*, 35, 8225–8243, <https://doi.org/10.1175/JCLI-D-22-0110.1>, 2022.
- Perez, G. M. P., Vidale, P. L., Dacre, H., and Martin, T. C. M.: What Is the Contribution of Convergence Zones to Global Precipitation? Assessing Observations and Climate Models Biases, *J. Geophys. Res.-Atmos.*, 129, e2023JD039635, <https://doi.org/10.1029/2023JD039635>, 2024.

- Shepherd, T. G.: Storyline approach to the construction of regional climate change information, *Proc. R. Soc. A*, 475, 20190013, <https://doi.org/10.1098/rspa.2019.0013>, 2019.
- Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Desai, S., Dima-West, I. M., Fowler, H. J., James, R., Maraun, D., Martius, O., Senior, C. A., Sobel, A. H., Stainforth, D. A., Tett, S. F. B., Trenberth, K. E., van den Hurk, B. J. J. M., Watkins, N. W., Wilby, R. L., and Zenghelis, D. A.: Storylines: an alternative approach to representing uncertainty in physical aspects of climate change, *Clim. Change*, 151, 555–571, <https://doi.org/10.1007/s10584-018-2317-9>, 2018.
- Themeßl, M. J., Gobiet, A., and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, *Int. J. Climatol.*, 31, 1530–1544, <https://doi.org/10.1002/joc.2168>, 2011.
- Wegelin, J. A.: A Survey of Partial Least Squares (PLS) Methods, with Emphasis on the Two-Blok Case, *Rel. téc.* Seattle: University of Washington, <https://stat.uw.edu/sites/default/files/files/reports/2000/tr371.pdf> (last access: 13 June 2025), 2000.
- Yoon, J. H. and Zeng, N.: An Atlantic influence on Amazon rainfall, *Clim. Dyn.*, 34, 249–264, <https://doi.org/10.1007/s00382-009-0551-6>, 2010.