



RESEARCH ARTICLE

10.1029/2024MS004905

MERCURY: A Fast and Versatile Multi-Resolution Based Global Emulator of Compound Climate Hazards

Key Points:

- Climate model emulators traditionally generate global fields amounting to Peta-Bytes of data
- We introduce the emulator MERCURY, that employs a lifting scheme based on image compression techniques to emulate region-specific fields whilst conserving global characteristics
- The lifting scheme further reduces data dimensionality allowing versatile extension to additional climate variables at minimal cost

Shruti Nath¹ , Julie Carreau² , Kai Kornhuber^{3,4} , Peter Pfleiderer⁵ , Carl-Friedrich Schuessner^{4,6}, and Philippe Naveau⁷

¹Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK, ²Department of Mathematics and Industrial Engineering, Polytechnique Montreal, Montreal, QC, Canada, ³Lamont-Doherty Earth Observatory, Columbia Climate School, New York, NY, USA, ⁴International Institute for Applied Systems Analysis, Vienna, Austria, ⁵Institut für Meteorologie, Universität Leipzig, Leipzig, Germany, ⁶IRI THESys and Geography Faculty, Humboldt-Universität zu Berlin, Berlin, Germany, ⁷Laboratory of Climate and Environmental Sciences, Institute Pierre-Simone-Laplace, Gif-sur-Yvette, France

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Nath,
shruti.nath@physics.ox.ac.uk

Citation:

Nath, S., Carreau, J., Kornhuber, K., Pfleiderer, P., Schuessner, C.-F., & Naveau, P. (2025). MERCURY: A fast and versatile multi-resolution based global emulator of compound climate hazards. *Journal of Advances in Modeling Earth Systems*, 17, e2024MS004905. <https://doi.org/10.1029/2024MS004905>

Received 22 DEC 2024

Accepted 10 OCT 2025

Author Contributions:

Conceptualization: Shruti Nath, Julie Carreau, Kai Kornhuber, Carl-Friedrich Schuessner, Philippe Naveau

Formal analysis: Shruti Nath, Julie Carreau, Kai Kornhuber, Peter Pfleiderer, Philippe Naveau

Funding acquisition: Shruti Nath, Philippe Naveau

Methodology: Shruti Nath, Julie Carreau, Kai Kornhuber, Philippe Naveau

Software: Shruti Nath

Abstract High-impact climate damages are often driven by compounding conditions, such as elevated heat stress arising from combined high humidity and temperatures. To explore future changes in compounding hazards under several climate scenarios, climate emulators can provide light-weight, data-driven complements to Earth System Models (ESMs). Yet, only a few existing emulators jointly emulate multiple climate variables. We introduce MERCURY (Multi-resolution EmulatoR for CompoUnd climate Risk analYsis), a spatio-temporal, multi-resolution emulator designed for compound climate risk analysis. MERCURY employs image-compression-based techniques for memory-efficient emulation and consists of two main modules. The regional module represents the monthly, regional response of a given variable to yearly Global Mean Temperature using a probabilistic additive model, resolving regional cross-correlations. The resulting regional values are then jointly disaggregated to grid-cell level values using a lifting-scheme operator, founded on principles of Discrete Wavelet Transforms. We demonstrate MERCURY on the humid-heat metric, wet bulb globe temperature (WBGT), as derived from temperature and relative humidity emulations. The emulated WBGT spatial correlations correspond well to those of ESMs and the 95% and 97.5% quantiles of WBGT distributions are well captured, with an average of 5% deviation. MERCURY's setup allows for region-specific emulations from which one can efficiently “zoom” into the grid-cell level across multiple variables by means of the reverse lifting-scheme operator. This circumvents the traditional problem of having to emulate complete, global-fields of climate data and resulting storage requirements.

Plain Language Summary Climate model emulators are approximations of climate models that provide a quick and low-cost alternative to exploring future climate scenarios. Traditional emulators generate large amounts of data covering the whole world, which still need to be condensed when exploring local and regional impacts. In this paper, we propose a new emulator based off image compression techniques. The setup allows one to “zoom” in and out from global to regional to local levels, providing user-relevant information across scales. It furthermore conserves both large-scale and local features and can be run in minutes. Given its versatile framework, the approach is easily extendable to new variables, and in this paper we demonstrate its ability to jointly capture temperature and relative humidity.

1. Introduction

High impact climatic events are often driven by a combination of physical variables acting together (Lesk et al., 2022; Raymond et al., 2020). For example, consequences for human health from extraordinary temperatures are most severe when they coincide with high levels of humidity (Baldwin et al., 2023). Multivariate events have been classified as one type of compound event in which correlated or entirely independent variables (or hazards) occur at the same location at the same time, leading to amplification of an impact (Zscheischler et al., 2020). With the aggravated risk of high-impact climate events, there is urgent demand for climate information that allows agile exploration of future climate risks.

State-of-the-art Earth System Models (ESMs) provide the basis for climate impact studies. However, they are computationally expensive to run and require large storage costs. Moreover, only a small number of their outputs (often post-processed e.g. to specific mean and spread values) are deployed for climate risk assessments, leading

Supervision: Julie Carreau, Carl-Friedrich Schuessner, Philippe Naveau
Validation: Shruti Nath
Visualization: Shruti Nath
Writing – original draft: Shruti Nath, Julie Carreau, Philippe Naveau
Writing – review & editing: Julie Carreau, Kai Kornhuber, Peter Pfleiderer, Carl-Friedrich Schuessner, Philippe Naveau

to information redundancy. This provides an entry point for low-cost statistical emulators focusing on application relevant outputs. To date, a wide range of emulators for different applications have been developed, providing yearly to monthly spatially resolved fields of climate variables such as temperature and precipitation, to allow for real-time impact assessments (Alexeeff et al., 2018; Beusch et al., 2020; Nath et al., 2022). Most emulators however focus on one to two variables at a time (Bassetti et al., 2024; Liu et al., 2023; Schöngart et al., 2024; Snyder et al., 2019), or jointly sample multivariate fields based on time sampling approaches. Time sampling approaches however, are hindered under scenarios where not many analog samples exist leading to repeated sampling of the same fields (Tebaldi et al., 2022). Moreover, problems of data storage costs and information redundancy, especially when moving to more climate variables, are still present.

In this paper we introduce the Multi-resolutional EmulatoR for CompoUnd climate Risk AnalYsis (MERCURY). MERCURY follows a multi-resolutional approach thus representing spatio-temporal climate fields as a series of closed sub-spaces and allowing for easy extension to multiple climate variables. This approach condenses key information of large-scale, low-frequency responses under climate change, whilst preserving small-scale, high-frequency features. It furthermore allows model reduction when representing the overall mean climate responses across multiple variables—which has previously been emphasized for climate model emulation (Kitsios et al., 2023). Our approach is built on discrete wavelet-transform methods to compress, estimate and recover our target climate fields of interest (Daubechies, 1992). Through a novel lifting scheme, discrete wavelet analysis has furthermore been adapted for irregularly shaped domains (Park & Oh, 2022; Sweldens, 1996) that are more common in climate and geographical spaces (e.g., continents). The lifting scheme is based on a local regression, to split irregularly shaped domains into subspaces and extract their low-frequency average values whilst storing the high-frequency residuals, known as second-generation, “Lazy wavelets” (Sweldens, 1995). Recently, Carreau and Naveau (2023) demonstrated the use of a lifting scheme based framework to simulate flood wave propagation by proposing an extension to represent spatio-temporal physics-based phenomena.

The emulator framework proposed in this study is composed of two components. The first component focuses on representing the mean response of impact-relevant regions to Global Mean Temperature (GMT). The second component then uses the lifting scheme to “zoom” into and generate the higher resolution, grid-cell level fields for each region. This allows compression of the emulation problem and efficient extension to multivariate representation, conserving both cross-variable and spatial correlations whilst circumventing data storage issues by only rendering regional, grid-cell level fields when needed. The structure of this paper is as follows: Section 2 describes the emulator framework of MERCURY and its evaluation procedure. Following this emulator evaluation results are provided in Section 3, after which we demonstrate the emulator output in Section 4 and proceed to final discussion in Section 5.

2. Methods

We are interested in jointly modeling monthly spatially resolved fields for a given set of variables conditional on the yearly GMT (GMT_y). MERCURY’s framework is summarized in Figure 1, and is composed of a regional, component generating monthly (m), regional (r) mean responses of a climate variable (generally referred to as V) to GMT_y (described in Section 2.1). Grid-cell level (gc) values are then reconstructed from given regional, monthly values using a month-specific lifting scheme (described in Section 2.2). Data used for training and evaluating MERCURY are described in Section S1 in Supporting Information S1.

2.1. Representing Mean Regional Monthly Responses

V is represented as a function of GMT_y using a classical Additive Model (AM) for each month m and AR6 region r , that is

$$V_{r,m} = f_{r,m}(GMT_y) + \epsilon_{r,m}, \quad (1)$$

where $m = 1, \dots, 12$, $r = 1, \dots, 44$. As the relationship between $V_{r,m}$ and GMT_y may not be necessarily linear, $f_{r,m}(\cdot)$ denotes a smooth function modeled by a classical cubic spline. The random vector $\epsilon_{r,m}$ corresponds to a zero-mean Gaussian vector with covariance matrix Σ_m constructed across all regions for a given month from the residuals of $f_{r,m}$. Equation 1 assumes that $f_{r,m}$ sufficiently captures the mean, region- and month-specific response of V , such that $\epsilon_{r,m}$ can be sampled on top as the residual, region- and month-specific variability from a

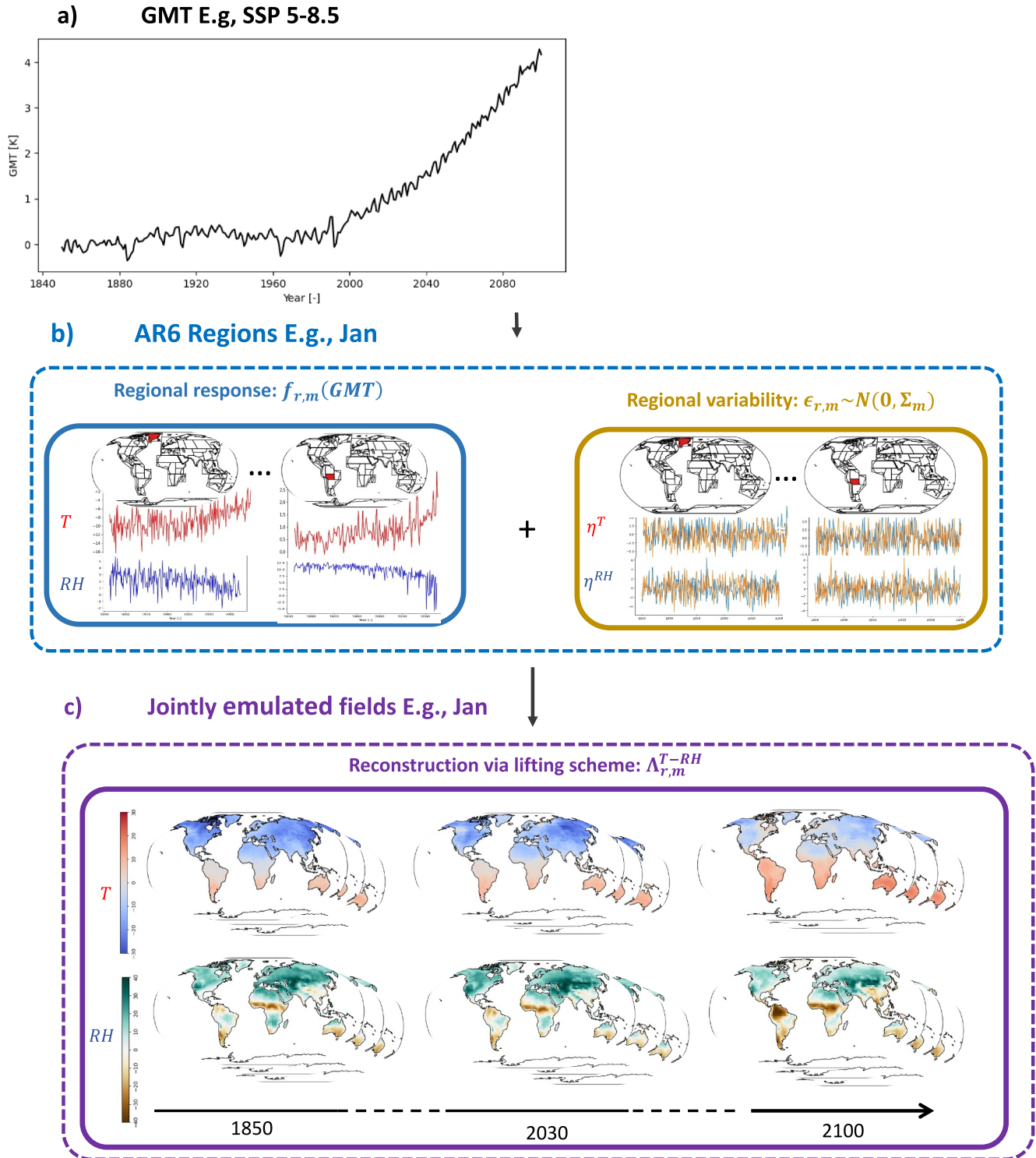


Figure 1. MERCURY’s framework for generating monthly spatially multivariate climate fields. Yearly global mean temperature values are used as inputs (panel a). The monthly, regional mean response and regional variability for each climate variable is first calculated (panel b). The lifting scheme is then employed to provide monthly spatially resolved, multivariate fields at the grid-cell level (panel c).

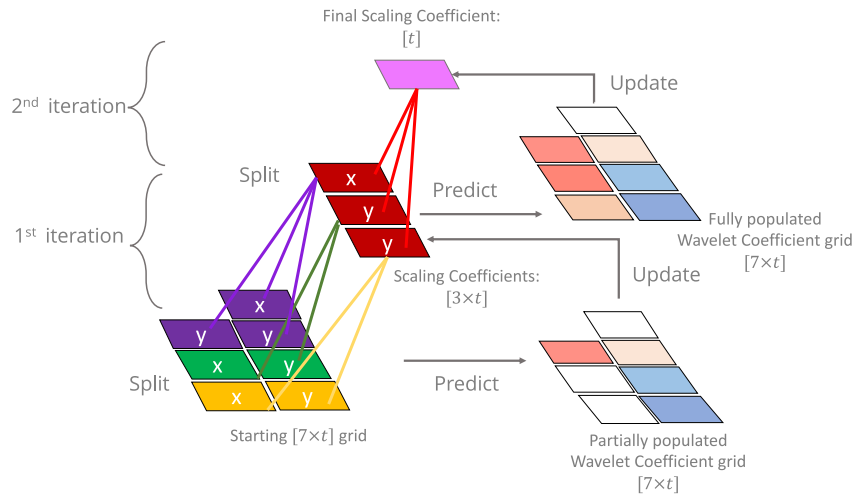


Figure 2. Toy example of the lifting scheme applied on a grid consisting of seven cells with values going from time 1 to t . At each iteration of the lifting scheme, the grid is split into groups of two (unless there is an odd number of cells in which case one group of three exists). The “predict” step stores the wavelet coefficients representing the regression errors resulting from local regression (in our case, naive regression). Finally, the x values are updated with the scaling coefficients obtained by averaging values within each group. Split, predict and update steps are repeated until only a single scaling coefficient exists which corresponds to the grid average, and the wavelet coefficient grid is fully populated up to six cells (in the toy example’s case two iterations). Within each lifting iteration the grid’s spatial dimension is reduced by approximately a half.

multivariate Gaussian distribution. Further analysis demonstrating the performance of $f_{r,m}$ alongside validation of Gaussian and stationarity assumptions for its residuals are provided in Figures S1–S4 in Supporting Information S1. We illustrate Equation 1, in the left map shown in panel b in Figure 1 that displays fitted $f_{r,m}$ for $m = 1$ (January) and two AR 6 regions with relative humidity in blue and temperature in red. The right map of panel b in Figure 1 shows the residuals $\epsilon_{r,m}$.

2.2. Grid-Cell Level Reconstruction per Region

The fitted model (Equation 1) provides estimates of $V_{r,m}$ for any given yearly GMT trajectory. To further disaggregate to the corresponding monthly, grid-cell level values, we employ a lifting scheme based framework, such that:

$$V_{gc,m,y} = \Lambda_{r,m}(V_{r,m}) \quad (2)$$

where $gc = 1 \dots n$ (n representing the number of grid-cells within region r), $\Lambda_{r,m}$ is a month- and region-specific reverse lifting-scheme operator obtained by inverting the lifting scheme. In the following subsections we start by elaborating on the framework of the lifting scheme, followed by the reverse lifting-scheme operator, $\Lambda_{r,m}$ obtained from it. Finally, we describe the extension of the reverse lifting-scheme operator, $\Lambda_{r,m}^{T-RH}$, to provide the joint temperature ($T_{gc,m,y}$) and relative humidity ($RH_{gc,m,y}$) emulations.

2.2.1. Lifting Scheme Framework

A month- and region-specific lifting scheme is constructed over the training data by iteratively applying the lifting scheme’s split, predict and update steps. These steps had been previously adapted for spatio-temporal data sets by Carreau and Naveau (2023), and are depicted in Figure 2. The specific implementation of these three steps in MERCURY is detailed below.

The split step consists of grouping grid cells—each containing a time series of values—from the original input grid into consecutive x - y pairs across each latitudinal band, unless an odd number of grid-cells exists in which case one group of three is made. The predict step involves estimating y from x using naïve regression—that is, assuming a simple linear relationship of the form $y = x$. The residuals of this local regression, given as,

$$\mathbf{d} = \mathbf{y} - \mathbf{x} \quad (3)$$

are recorded and serve as wavelet coefficients—specifically, second-generation wavelet coefficients, which differ from first-generation wavelets in that they are not constructed from a basis function (see Sweldens (1995) for details). When a group is of size three, the predict step is performed twice (since there are two \mathbf{y} cells), resulting in two wavelet coefficients. Finally, the update step consists of replacing the \mathbf{x} cell with the scaling coefficients \mathbf{c} , corresponding to the average computed within each group:

$$\mathbf{c} = \frac{\mathbf{x} + \mathbf{y}}{2} = \mathbf{x} + \frac{1}{2} \mathbf{d}. \quad (4)$$

The update step must be slightly adapted to account for groups of size three, by averaging over three cells instead of two.

The split, predict and update steps are iterated through until a single scaling coefficient corresponding to the grid average, and a fully populated grid of wavelet coefficients—except for the cell that corresponds to the scaling coefficient - are obtained. Each iteration corresponds to a resolution level and the further the iteration progresses, the lower the resolution, until reaching the regional average.

When reversing the lifting scheme, we iterate backwards from the scaling coefficient of the final iteration. At each iteration, the split step is reversed by inverting the predict and update steps as follows:

$$\mathbf{x} = \mathbf{c} - \frac{1}{2} \cdot \mathbf{d} \quad \text{and} \quad \mathbf{y} = \mathbf{d} + \mathbf{x}. \quad (5)$$

After completing all backward iterations, the exact time series values are recovered for each cell in the original grid and in such, the lifting scheme is a lossless compression.

As a simple interpretative summary, the lifting scheme decomposes a spatio-temporal field into low-frequency scaling coefficients—analogue to spatial averages that preserve regional trends—and high-frequency residuals that is, wavelet coefficients. It can thus be thought of as a simpler alternative to an Empirical Orthogonal Function analysis based on rudimentary arithmetic that is more flexible in the sense that the decomposition does not make any linearity assumptions (apart from the predict and update operations being locally linear). Once, fully iterated over each region and month, the lifting scheme stores a configuration of scaling coefficients, and their corresponding wavelet coefficients (if n is the number of initial grid cells, there are $n - 1$ wavelet coefficients). As the reverse lifting scheme starts from the grid average (which corresponds to the scaling coefficients \mathbf{c} from the final iteration), and these values are unknown when using the emulator to simulate based solely on GMT_y , we may instead use the estimated monthly regional values $V_{r,m}$, see Equation 1, as a substitute \mathbf{c}' . This substitute can then be used to reconstruct grid-cell level values by selecting suitable wavelet coefficients.

2.3. Generating New Emulations

The region- and month-specific reverse lifting-scheme operator $\Lambda_{r,m}$, see Equation 2, is able to go back down to a grid-cell level resolution given a single monthly, regional value. The key step within the reverse lifting-scheme operator $\Lambda_{r,m}$ is identifying the wavelet coefficients that are prototypical for a given monthly, regional value from which to reconstruct the original climate fields with. To do so, we treat the configuration of scaling and wavelet coefficients stored within the lifting scheme as a simple, unordered look-up table that is scenario interoperable and time interchangeable. We start by defining the neighborhood around a given regional, monthly value, $V_{r,m}$ by indexing the 20 scaling values within the lifting scheme's decomposition that are closest to it in value. Wavelet coefficients within the neighborhood are then sampled using Monte Carlo method. The Monte Carlo method builds a multivariate Gaussian distribution using a covariance matrix constructed across the 20 prototypical wavelet coefficients, thus conserving high-frequency spatial structures during sampling.

A detailed algorithm is provided in Algorithm 1. It can be noted that the size of the neighborhood can be thought of as a tunable hyperparameter, however for sake of simplicity we maintain it as constant during this study. As the lifting scheme is a Discrete Wavelet Transform, it is expected to preserve the euclidean distance between time

series within the lower-dimensional, regional space (Chan & Fu, 1999). Hence, reconstruction using the wavelet coefficients should preserve grid-cell level spatio-temporal features within each AR6 region, resulting in smooth timeseries with minimal jumps. Nonetheless, as the lifting scheme is region-specific, seamless blending between AR6 regions during reconstruction needs to be ensured. For this, a buffer zone around each regional boundary is applied before sampling wavelet coefficients, as previously done to overcome boundary issues for image processing exercises (Hee-Seok et al., 2001; Naveau & Oh, 2004). A buffer zone of width one grid-cell is chosen and the lifting scheme is configured using it. Reverse lifting is conducted using the whole region plus the buffer zone, after which buffer zone values are simply discarded.

Algorithm 1: Generating 1,000 emulations using MERCURY

```

m ← month of interest e.g., Jan
r ← region of interest e.g., SAH
c ← Vector containing final scaling coefficients over all samples fitted
on for m and r
d ← Vector containing corresponding grid-cell level wavelet coefficients
Vr,m ← fr,m(GMTy) + εr,m           ▷ Regional value at GMTy level
of interest
N ← argsort(|c - Vr,m|, ascending = True)[:20]           ▷ Define
neighborhood
for emu in range(1,000) do
    demu ← MonteCarlo(d.index(N)).sample
    Vr,m,gcemu ← Λr,m(Vr,m, demu)
end for

```

2.3.1. Reverse Lifting-Scheme Operator for Multivariate Sampling

The reverse lifting-scheme operator, $\Lambda_{r,m}$, is extended to a new operator, denoted $\Lambda_{r,m}^{T-RH}$, which enables the joint sampling of $T_{gc,m,y}$ and $RH_{gc,m,y}$. We first identify the key variable that will be used to define the neighborhood for sampling wavelet coefficients. In our case, we select $T_{r,m,y}$, as its relationship to GMT_y is most established (Herger et al., 2015; Tebaldi & Arblaster, 2014; Tebaldi & Knutti, 2018). Having defined the neighborhood through the key variable, the usual wavelet sampling steps are then carried out. One difference however, is that the wavelet coefficients can now be jointly sampled across each variable's lifting scheme decomposition, given the mutually defined neighborhood. To ensure a strict relationship to the key variable, $T_{r,m,y}$, we additionally impose a conditional sampling of the RH wavelet coefficients on the T wavelet coefficients within the Monte Carlo routine, by calculating their conditional covariance matrix,

$$\Sigma_{RH|T} = \Sigma_{RH,RH} - \Sigma_{RH,T} \cdot \Sigma_{T,T}^{-1} \cdot \Sigma_{T,RH}, \quad (6)$$

where $\Sigma_{RH|T}$ is the conditional covariance matrix constructed from blocks of the covariance matrix constructed across grid-cells and variables,

$$\begin{bmatrix} \Sigma_{T,T} & \Sigma_{T,RH} \\ \Sigma_{RH,T} & \Sigma_{RH,RH} \end{bmatrix} \quad (7)$$

2.4. Evaluation

In evaluating MERCURY, 10 monthly, regional values are produced from $f_{r,m}$, and for each value 100 reconstructions are performed with the reverse lifting-scheme operator $\Lambda_{r,m}^{T-RH}$ resulting in 1,000 spatially resolved, monthly, multivariate emulations. MERCURY is evaluated on the test scenario SSP2-4.5. To evaluate the final multivariate T and RH emulations, we first consolidate them into a single representative compound index, the indoor Wet Bulb Globe Temperature ($WBGT$). $WBGT$ is calculated using Stull's method (Stull, 2011) by first calculating Wet Bulb Temperature (WBT):

$$\begin{aligned}
 WBT_{gc,m,y} = & T_{gc,m,y} \cdot \left(c_1 \cdot \sqrt{RH_{gc,m,y} + c_2} \right) + \tan^{-1}(T_{gc,m,y} + RH_{gc,m,y}) - \tan^{-1}(RH_{gc,m,y} - c_3) \\
 & + c_4 \cdot RH_{gc,m,y}^{\frac{3}{2}} \cdot \tan^{-1}(c_5 \cdot RH_{gc,m,y}) - c_6,
 \end{aligned} \tag{8}$$

where c_1, c_2, c_3, c_4, c_5 and c_6 constants with values 0.16, 8.31, 1.68, 0.0039, 0.023, and 4.69 respectively. $WBGT$ is then obtained from WBT as follows,

$$WBGT_{gc,m,y} = \frac{2}{3} \cdot WBT_{gc,m,y} + \frac{1}{3} \cdot T. \tag{9}$$

Final $WBGT$ emulations are inspected for their representation of spatial structures by means of Spearman correlations. For both the ESM and emulator outputs, a Spearman correlation matrix is constructed across all grid cells. The difference is then taken by simply subtracting the emulator's Spearman correlation matrix from the ESM's Spearman correlation matrix. This provides useful diagnosis into how well dominant spatial structures are approximated, where ideally the difference between the two matrices is zero.

We furthermore evaluate MERCURY's representation of the $WBGT$ distributions on a grid-cell and regionally aggregated level. On a grid-cell level, we investigate MERCURY's ability to approximate the median (50%) and the extreme upper (95% and 97.5%) quantiles of $WBGT$ distributions as calculated from T and RH values outputted by the ESM at each month. We do so by calculating monthly, grid-cell level quantile deviations as used in previous emulator evaluations (Beusch et al., 2020; Nath et al., 2022; Quilcaille et al., 2022). Quantile deviations are calculated by first extracting the quantile time series of an emulated ensemble for quantile q . The proportion of time steps that the ESM values appear below the emulated quantile value is then calculated (q^{ESM}). The quantile deviation is obtained as $q^{ESM} - q$, such that a positive value means that the emulated quantile is larger in value than that of the actual ESM and vice versa.

We evaluate the emulator's representation of $WBGT$ distributions on multiple regionally aggregated levels by means of probability rank distributions. Levels going from AR6 regions to continental to global are investigated, and are obtained by taking latitudinally-weighted averages from emulated, grid-cell level fields. For each level of aggregation, the probability rank distribution is obtained by calculating the probability rank of the actual, ESM value with respect to the emulated ensemble over all time steps. If the emulated ensemble perfectly captures the actual ESM distribution, we would expect the median probability rank value to correspond to 50% and so on, such that their final distribution is uniformly distributed.

3. Results

In the following subsections we first show evaluation results on the test scenario SSP 2–4.5 for the representation of spatial correlations within the multivariate emulations (Section 3.1) and then the representation of the overall distribution (Section 3.2).

3.1. Spatial Evaluation

Differences in the Spearman correlation matrix of the ESM to that of MERCURY are shown in Figure 3. Across ESMs and months, we note a positive difference, indicating an underestimation of spatial correlations within the emulator. This is particularly apparent in July and concentrated toward correlations with grid cells in equatorial latitudinal bands (i.e., -1.25°N). Since the emulator relies on a statistical approximation of spatial covariances, such underestimation of spatial structures is not entirely unexpected. Moreover, equatorial latitudes have stronger temporal correlations (see Durban-Watson test for serial correlation Figure S4 in Supporting Information S1), and this indicates some shortcoming in MERCURY's design choice of not accounting for serial correlations.

Closer inspection of spatial correlations at specific grid-cells is provided in Figure S6 in Supporting Information S1, and it becomes apparent that while localized intra-regional correlations are well represented, the cross-regional correlations are more likely to be underestimated. Moreover, regional boundaries can still be identified on closer inspection, indicating that the buffer zone is not able to eliminate boundary effects altogether. Across all ESMs and months, an overestimation of spatial correlations at latitudes higher than 61.25°N can also be noted. Grid cells above 61.25°N mainly correspond to Greenland, and this could be a product of Greenland being treated

Difference in spatial spearman correlation matrix: ESM-Emulator

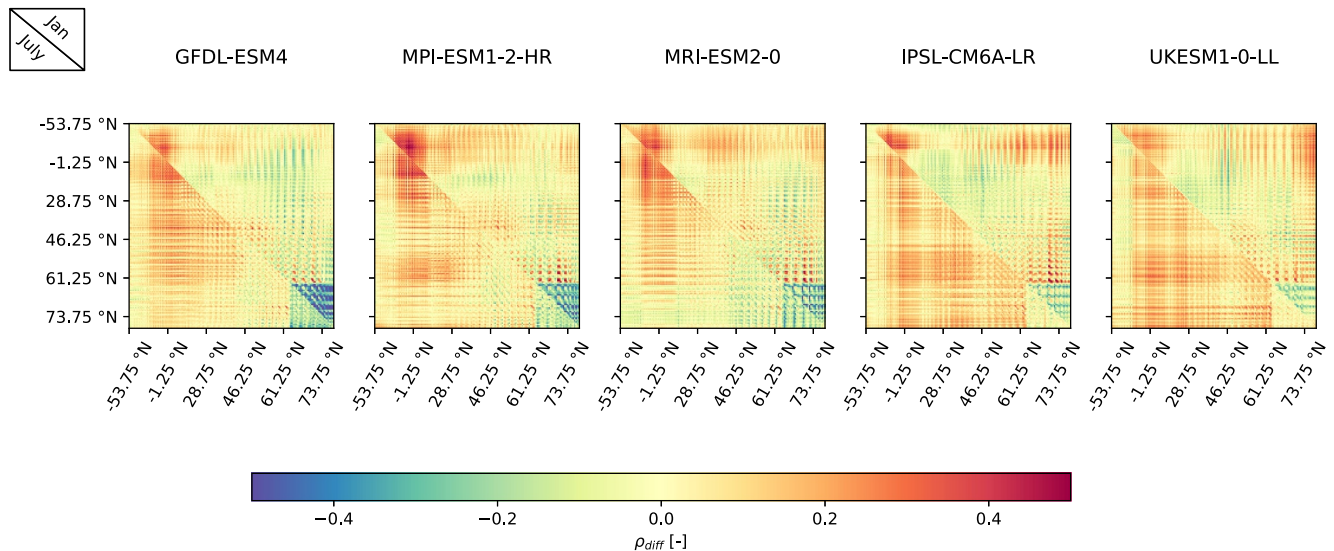


Figure 3. Difference between the Earth System Model and emulator Spearman correlation matrix obtained by subtracting the emulator's Spearman correlation matrix from that of the ESM's for test scenario SSP 2–4.5. The Spearman correlation matrix is calculated over all land grid cells for January (upper triangle) and July (lower triangle). Latitudinal labels refer to the average latitudinal value across the given row/column.

as its own AR6 region, leading to an overestimation of its intra-regional correlations and an underestimation of its cross-regional correlations.

Comparing the results from Figure 3 to inter-comparison results against the existing monthly temperature emulator MESMER-M (Figure S5 in Supporting Information S1) using SSP 5–8.5—which was used in the training set of both emulators—we note that ESMs which had overall poorer Continuous Ranked Probability Skill Scores against the benchmark emulator MESMER-M—and therefore no notable improvement in skill—also show larger underestimation of spatial correlations. For further analysis, we focus on 2 ESMs, MRI-ESM2-0 and UKESM1-0-LL, representative of where MERCURY brings meaningful improvements with respect to MESMER-M and where not so much, respectively. We furthermore focus on July as this is the month where the most underestimation of spatial correlations occurs.

3.2. Representation of the Overall Distribution

Figure 4 shows 50%, 95% and 97.5% quantile deviation maps (panel a) for the two representative ESMs, MRI-ESM2-0 and UKESM1-0-LL, in July. Positive quantile deviations indicate overestimation of the quantile value by MERCURY and vice versa. Whereas 95% and 97.5% quantile deviations are quite low (between –5% and 5%), 50% quantile deviations show large overestimations with some grid-cells for example, in South Asia, having values of up to 20%. This indicates that MERCURY may be overestimating the overall mean response of WBGT to GMT on a grid-cell level. The quantile deviations across all models for both January and July are provided in Figure S7 in Supporting Information S1, where negative quantile deviations in January, particularly for the 50% quantile can be observed. This indicates that while MERCURY captures a good spread, it has a seasonal bias within the WBGT median estimates. This could be driven by other confounding factors such as irrigation and its immediate effects on relative humidity.

Probability rank distributions (panel b) look reasonably uniformly distributed for both ESMs across all aggregation levels, albeit globally aggregated showing lower median values (i.e., again an overestimation by MERCURY). In addition to the quantile deviation maps, this indicates that selection of regions could lead to biases in representation of grid-cell level distributions (e.g., perhaps a grid-cell's responses is not so well correlated to the regional response). Consequently, an overall bias in representation of global distributions—as seen in the globally aggregated probability rank distribution—results. Nevertheless, in this study we seek to represent the most impact relevant regions and note that this is a consequence of that design choice.

Comparison between emulated and actual distribution: WBGT, July

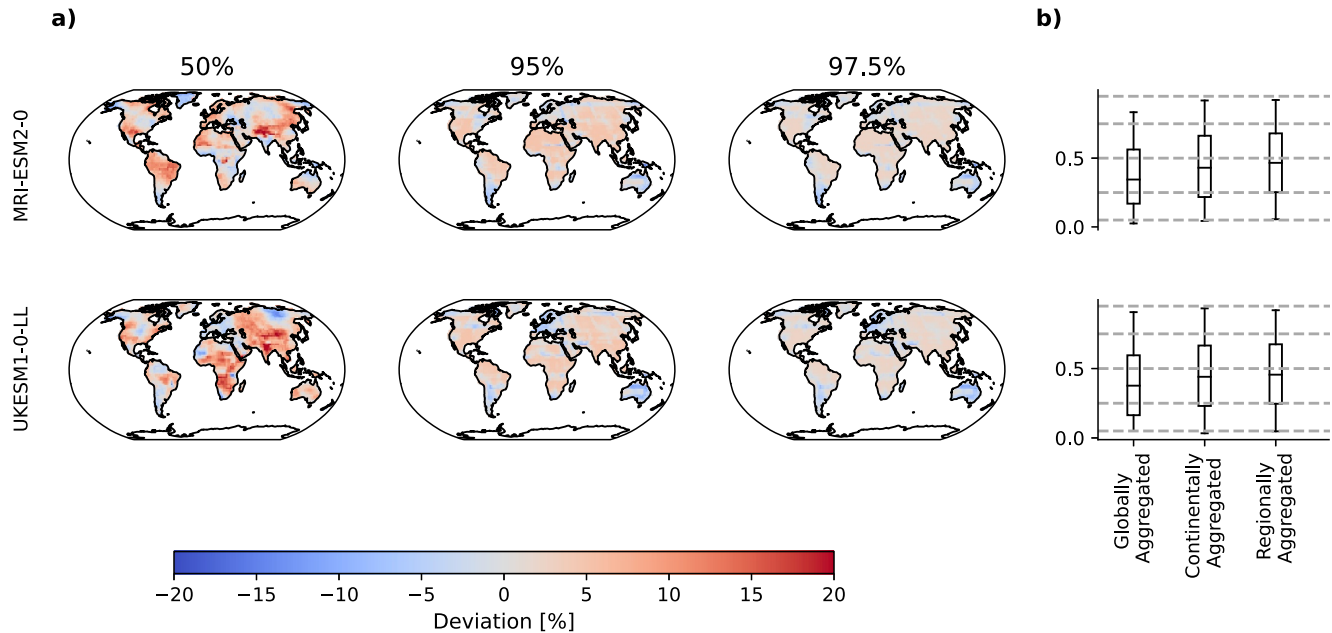


Figure 4. Comparison between the emulated and actual distribution for test scenario SSP 2–4.5. (a) Quantile deviation maps for the 50%, 95% and 97.5% quantiles, where red means that the emulated quantile is warmer than the actual Earth System Model quantile and vice versa for blue. (b) Probability rank distributions of the actual data with respect to the emulated ensemble. Data is aggregated to AR6 regional, continental and global levels before calculating the probability ranks. Whiskers indicate the 5th and 95th percentiles. If the distribution of actual data is captured perfectly, then the median should correspond to 0.5, edges to 0.75 and 0.25, and whiskers to 0.05 and 0.95.

4. Example WBGT Superensemble Time Series

Figure 5 provides 2-D histogram time series of a superensemble pooling together 1,000 WBGT emulations for each ESM, so a total of 5,000 emulations. WBGT values are aggregated to AR6 regions, Sahel and South Asia, and globally. ESM values are also provided for reference. We again show results for the SSP 2–4.5 scenario. A notable spread and divergence in ESM values across regions and globally is apparent. For example, Sahel displays two modes of ESM WBGT values, starting at approximately 22 and 24°C in the year 1850 and increasing at different rates till they converge around 24–25°C by 2100. MERCURY is able to capture both the spread within each ESM initial-condition ensemble as well as the inter-ESM spread in magnitude and rates of change. This provides useful perspective into the potential of multivariate, lightweight emulators to inform impact assessments. To this extent, they not just provide useful approximations of the ESM initial-condition ensemble spread, but also of the inter-ESM spread, which in some cases may be larger and thus have a greater degree of uncertainty.

5. Discussion

We present MERCURY, a fast and versatile emulator framework that allows approximation of spatially resolved risks from multivariate compound hazards such as WBGT. After training on ESM outputs, MERCURY starts from GMT to deterministically approximate monthly, regional temperature and relative humidity separately by means of a month- and region-specific regression (AM, see Equation 1). Region-to-region correlations are then added by sampling from a variable- and month-specific multivariate Gaussian distribution. Grid-cell level temperatures and relative humidity are jointly reconstructed from their regional values using an operator that reverses the “lifting scheme” adapted for spatio-temporal multivariate sampling. The lifting scheme is a discrete wavelet transform that performs an efficient compression of regional fields based on a local regression, to iteratively split and compress irregularly shaped domains into their low-frequency average features and high-frequency residuals. The high-frequency residuals at each iteration are stored as wavelet coefficients, thus enabling reconstruction of the original regional field from a single, regional average value. It should be emphasized that MERCURY does not contain any temporal correlation element. This is mainly as serial

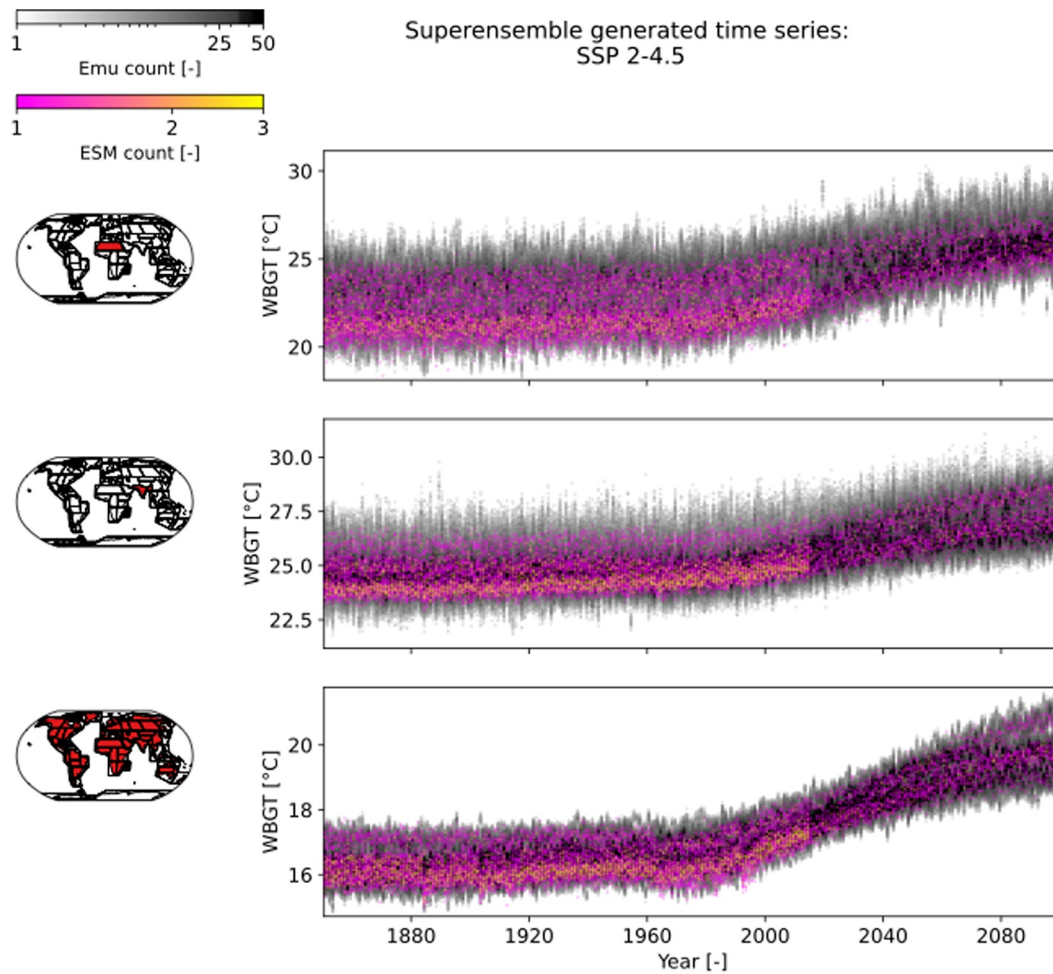


Figure 5. 2-D histogram time series for the emulated superensemble (grays), with the actual Earth System Model (ESM) superensemble overlaid for July, SSP 2–4.5. 1,000 emulations were produced for each ESM initial-condition ensemble member. Note, that since historical runs have more ensemble members, the emulation count is also higher up to 2015.

correlations were not identified as significant (see Durban-Watson test, Figure S4 in Supporting Information S1), and we thus opted for a simpler design. In its current state, MERCURY can only be used in investigating spatially-distributed, multivariate sensitivities to GMT, but not features that accumulate/recur over time, or that are caused by larger modes of climate variability.

The generation of grid-cell level emulations is performed by sampling wavelet coefficients within the “neighborhood” of a given regional, monthly value from the lifting scheme’s decomposition by means of the Monte Carlo method. This allows flexible extension to more variables by sampling their wavelet coefficients from the same neighborhood. In this study, we select a key variable, temperature, through which to define the neighborhood which is then used in sampling wavelet coefficients for any additional variables, that is, relative humidity. In such, we impose a strict, hierarchical dependency of relative humidity on temperature by defining the neighborhood using temperature only and also by sampling relative humidity using its conditional covariance matrix to temperature. In future, a mutually defined neighborhood could be considered instead which would reduce the computationally complexity of our approach. In line with this, future extension to more climate variables may also encounter non-Gaussian variables such as precipitation. Future extensions of this study may thus benefit from more sophisticated sampling approaches such as use of Diffusion-based Neural Networks, and the framework of MERCURY is modular enough to allow flexible extension for such alternative sampling approaches.

A key advantage of MERCURY lies in its ability to treat irregularly shaped domains following multi-resolutional analysis. This means that during emulation generation, MERCURY does not require derivation of high-resolution information for all regions included, but can selectively choose regions of interest to then “zoom” into by means of the reverse lifting-scheme operator. For impact assessments, this means real-time availability of select impact-relevant regional information without the need to generate and sift through global fields which amount to Petabytes of data. To the best of our knowledge, this is the only emulator approach existing so far that tackles the data management problem effectively. In terms of memory space required for storage and emulation generation, the lifting operator itself amounts to 1 GB of storage, however for a given region during reconstruction the load-in memory requirements are 250 MB. MERCURY itself holds low parametric complexity, mainly confined to representing regional responses to GMT, after which it reconstructs grid-cell level responses based on Monte Carlo sampling. This ensures limited growth of parametric uncertainty going down the emulation chain (as otherwise seen in Nath et al. (2024)), as well as imposes less stringent functional forms on the grid-cell level responses to GMT.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

MESMER and MESMER-M are publicly available under Github (Hauser et al., 2021). Code for the lifting scheme can be found within the GitHub repository, available through zenodo Nath et al. (2025). The CMIP6 data are available from the public CMIP archive at <https://doi.org/10.5194/gmd-14-629-2021> (Petrie et al., 2021).

Acknowledgments

We thank the reviewers for their constructive feedback, which helped improve the quality and clarity of this manuscript. We thank the climate modeling groups listed in Table S1 in Supporting Information S1 for producing and making available the CMIP6 model outputs and Urs Beyerle and Lukas Brunner for downloading the CMIP6 data and pre-processing them. SN would like to acknowledge funding from the European Cooperation in Science and Technology, COST action Grant CA19139. J. Carreau would like to acknowledge funding by the Natural Sciences and Engineering Research Council of Canada (NSERC), by Fonds de Recherche du Québec Nature et Technologies (FRQNT) and IVADO. Part of Naveau’s research work was supported by European H2020 XAIDA (Grant agreement ID: 101003469) and the French national programs: 80 PRIME CNRS-INSU, Agence Nationale de la Recherche (ANR) under reference ANR-20-CE40-0025-01 (T-REX), the ANR EXSTA, and the PEPR TRACCS programme under Grant ANR-22-EXTR-0005. SN, PP and CFS acknowledge support from the European Union’s Horizon 2020 research and innovation programmes under Grant Agreement No. 101003687 (PROVIDE).

References

- Alexeeff, S. E., Nychka, D., Sain, S. R., & Tebaldi, C. (2018). Emulating mean patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments. *Climatic Change*, 146(3–4), 319–333. <https://doi.org/10.1007/s10584-016-1809-8>
- Baldwin, J. W., Benmarhnia, T., Ebi, K. L., Jay, O., Lutsko, N. J., & Vanos, J. K. (2023). Humidity’s role in heat-related health outcomes: A heated debate. *Environmental Health Perspectives*, 131(5), 055001. <https://doi.org/10.1289/EHP11807>
- Bassetti, S., Hutchinson, B., Tebaldi, C., & Kravitz, B. (2024). DiffESM: Conditional emulation of temperature and precipitation in Earth system models with 3D diffusion models. *Journal of Advances in Modeling Earth Systems*, 16(10), e2023MS004194. <https://doi.org/10.1029/2023MS004194>
- Beusch, L., Gudmundsson, L., & Seneviratne, S. I. (2020). Emulating Earth system model temperatures with MESMER: From global mean temperature trajectories to grid-point-level realizations on land. *Earth System Dynamics*, 11(1), 139–159. <https://doi.org/10.5194/esd-11-139-2020>
- Carreau, J., & Naveau, P. (2023). A spatially adaptive multi-resolution generative algorithm: Application to simulating flood wave propagation. *Weather and Climate Extremes*, 41, 100580. <https://doi.org/10.1016/j.wace.2023.100580>
- Chan, K.-P., & Fu, A. W.-C. (1999). Efficient time series matching by wavelets. In *Proceedings 15th International Conference on Data Engineering (Cat. No. 99cb36337)* (pp. 126–133). IEEE. <https://doi.org/10.1109/ICDE.1999.754915>
- Daubechies, I. (1992). *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics.
- Hauser, M., Beusch, L., Nicholls, Z., & Schwaab, J. (2021). MESMER-group/mesmer: Version 0.8.3 [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.5802054>
- Hee-Seok, O. H., Naveau, P., & Lee, G. (2001). Polynomial boundary treatment for wavelet regression. *Biometrika*, 88(1), 291–298. <https://doi.org/10.1093/biomet/88.1.291>
- Herger, N., Sanderson, B. M., & Knutti, R. (2015). Improved pattern scaling approaches for the use in climate impact studies. *Geophysical Research Letters*, 42(9), 3486–3494. <https://doi.org/10.1002/2015GL063569>
- Kitsios, V., O’Kane, T. J., & Newth, D. (2023). A machine learning approach to rapidly project climate responses under a multitude of net-zero emission pathways. *Communications Earth & Environment*, 4(1), 355. <https://doi.org/10.1038/s43247-023-01011-0>
- Lesk, C., Anderson, W., Rigden, A., Coast, O., Jägermeyr, J., McDermid, S., et al. (2022). Compound heat and moisture extreme impacts on global crop yields under climate change. *Nature Reviews Earth & Environment*, 3(12), 872–889. <https://doi.org/10.1038/s43017-022-00368-8>
- Liu, Z., Luo, D., Xu, Y., Jaakkola, T., & Tegmark, M. (2023). GenPhys: From physical processes to generative models. *arXiv preprint arXiv:2304.02637*, 1–21. Retrieved from <http://arxiv.org/abs/2304.02637>
- Nath, S., Carreau, J., Kornhuber, K., Pfliederer, P., Schleussner, C.-F., & Naveau, P. (2025). Lifting scheme for MERCURY v2025.3.03 [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.15423119>
- Nath, S., Hauser, M., Schumacher, D. L., Lejeune, Q., Gudmundsson, L., Quilcaille, Y., et al. (2024). Representing natural climate variability in an event attribution context: Indo-Pakistani heatwave of 2022. *Weather and Climate Extremes*, 44, 100671. <https://doi.org/10.1016/j.wace.2024.100671>
- Nath, S., Lejeune, Q., Beusch, L., Seneviratne, S. I., & Schleussner, C. F. (2022). MESMER-M: An Earth system model emulator for spatially resolved monthly temperature. *Earth System Dynamics*, 13(2), 851–877. <https://doi.org/10.5194/esd-13-851-2022>
- Naveau, P., & Oh, H. S. (2004). Polynomial wavelet regression for images with irregular boundaries. *IEEE Transactions on Image Processing*, 13(6), 773–781. <https://doi.org/10.1109/TIP.2003.821345>
- Park, S., & Oh, H.-S. (2022). Lifting scheme for streamflow data in river networks. *Journal of the Royal Statistical Society - Series C: Applied Statistics*, 71(2), 467–490. <https://doi.org/10.1111/rssc.12542>

- Petrie, R., Denvil, S., Ames, S., Levvasseur, G., Fiore, S., Allen, C., et al. (2021). Coordinating an operational data distribution network for CMIP6 data. *Geoscientific Model Development*, *14*(1), 629–644. <https://doi.org/10.5194/gmd-14-629-2021>
- Quilcaille, Y., Gudmundsson, L., Beusch, L., Hauser, M., & Seneviratne, S. I. (2022). Showcasing MESMER-X: Spatially resolved emulation of annual maximum temperatures of Earth system models. *Geophysical Research Letters*, *49*(17), 1–11. <https://doi.org/10.1029/2022GL099012>
- Raymond, C., Matthews, T., & Horton, R. M. (2020). The emergence of heat and humidity too severe for human tolerance. *Science Advances*, *6*(19), eaaw1838. <https://doi.org/10.1126/sciadv.aaw1838>
- Schöngart, S., Gudmundsson, L., Hauser, M., Pfliederer, P., Lejeune, Q., Nath, S., et al. (2024). Introducing the MESMER-M-TPv0.1.0 module: Spatially explicit Earth system model emulation for monthly precipitation and temperature. *EGU sphere*. <https://doi.org/10.5194/egusphere-2024-278>
- Snyder, A., Link, R., Dorheim, K., Kravitz, B., BondLamberty, B., & Hartin, C. (2019). Joint emulation of Earth System Model temperature-precipitation realizations with internal variability and space-time and crossvariable correlation: Fldgen v2.0 software description. *PLoS One*, *14*(10), 1–13. <https://doi.org/10.1371/journal.pone.0223542>
- Stull, R. (2011). Wet-bulb temperature from relative humidity and air temperature. *Journal of Applied Meteorology and Climatology*, *50*(11), 2267–2269. <https://doi.org/10.1175/JAMC-D-11-0143.1>
- Sweldens, W. (1995). The lifting scheme: A construction of second generation wavelets (Technical Report).
- Sweldens, W. (1996). The lifting scheme: A construction of second generation wavelets. *SIAM Journal on Mathematical Analysis*, *29*(2), 511–546. <https://doi.org/10.1137/s0036141095289051>
- Tebaldi, C., & Arblaster, J. M. (2014). Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Climatic Change*, *122*(3), 459–471. <https://doi.org/10.1007/s10584-013-1032-9>
- Tebaldi, C., & Knutti, R. (2018). Evaluating the accuracy of climate change pattern emulation for low warming targets. *Environmental Research Letters*, *13*(5), 055006. <https://doi.org/10.1088/1748-9326/aabef2>
- Tebaldi, C., Snyder, A., & Dorheim, K. (2022). STITCHES: Creating new scenarios of climate model output by stitching together pieces of existing simulations. *Earth System Dynamics*, *13*(4), 1557–1609. <https://doi.org/10.5194/esd-13-1557-2022>
- Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., et al. (2020). A typology of compound weather and climate events. *Nature Reviews Earth & Environment*, *1*(7), 333–347. <https://doi.org/10.1038/s43017-020-0060-z>