# MERCURY: A fast and versatile multi-resolution based global emulator of compound climate hazards

S. Nath, J. Carreau, K. Kornhuber, P. Pfleiderer, C.-F. Schleussner, P. Naveau

G–2024–81

December 2024

La collection *Les Cahiers du GERAD* est constituée des travaux de recherche menés par nos membres. La plupart de ces documents de travail a été soumis à des revues avec comité de révision. Lorsqu'un document est accepté et publié, le pdf original est retiré si c'est nécessaire et un lien vers l'article publié est ajouté.

**Citation suggérée :** S. Nath, J. Carreau, K. Kornhuber, P. Pfleiderer, C.-F. Schleussner, P. Naveau (Décembre 2024). MERCURY: A fast and versatile multi-resolution based global emulator of compound climate hazards, Rapport technique, Les Cahiers du GERAD G– 2024–81, GERAD, HEC Montréal, Canada.

Avant de citer ce rapport technique, veuillez visiter notre site Web (https://www.gerad.ca/fr/papers/G-2024-81) afin de mettre à jour vos données de référence, s'il a été publié dans une revue scientifique.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2024 – Bibliothèque et Archives Canada, 2024 The series *Les Cahiers du GERAD* consists of working papers carried out by our members. Most of these pre-prints have been submitted to peer-reviewed journals. When accepted and published, if necessary, the original pdf is removed and a link to the published article is added.

Suggested citation: S. Nath, J. Carreau, K. Kornhuber, P. Pfleiderer, C.-F. Schleussner, P. Naveau (December 2024). MERCURY: A fast and versatile multi-resolution based global emulator of compound climate hazards, Technical report, Les Cahiers du GERAD G-2024-81, GERAD, HEC Montréal, Canada.

Before citing this technical report, please visit our website (https: //www.gerad.ca/en/papers/G-2024-81) to update your reference data, if it has been published in a scientific journal.

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2024 – Library and Archives Canada, 2024

GERAD HEC Montréal 3000, chemin de la Côte-Sainte-Catherine Montréal (Québec) Canada H3T 2A7

Tél.: 514 340-6053 Téléc.: 514 340-5665 info@gerad.ca www.gerad.ca

# MERCURY: A fast and versatile multi-resolution based global emulator of compound climate hazards

Shruti Nath <sup>a</sup> Julie Carreau <sup>b, c</sup> Kai Kornhuber <sup>d, e</sup> Peter Pfleiderer <sup>f</sup> Carl-Friedrich Schleussner <sup>e, g</sup> Philippe Naveau <sup>h</sup>

<sup>a</sup> Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford

- <sup>b</sup> GERAD, Montréal, Canada
- <sup>c</sup> Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal, Canada
- <sup>d</sup> Lamont-Doherty Earth Observatory, Columbia Climate School, New York, USA
- <sup>e</sup> International Institute for Applied Systems Analysis, Vienna, Austria
- <sup>*f*</sup> Institut für Meteorologie, Universität Leipzig, Leipzig, Germany
- <sup>g</sup> IRI THESys and Geography Faculty, Humboldt-Universität zu Berlin, Berlin, Germany
- <sup>h</sup> Laboratory of Climate and Environmental Sciences, Institute Pierre-Simon-Laplace, France

shruti.nath@physics.ox.ac.uk
julie.carreau@polymtl.ca
kaik@ldeo.columbia.edu
peter.pfleiderer@uni-leipzig.de
schleussner@iiasa.ac.at
philippe.naveau@lsce.ipsl.fr

## December 2024 Les Cahiers du GERAD G-2024-81

Copyright © 2024 Nath, Carreau, Kornhuber, Pfleiderer, Schleussner, Naveau

Les textes publiés dans la série des rapports de recherche *Les Cahiers du GERAD* n'engagent que la responsabilité de leurs auteurs. Les auteurs conservent leur droit d'auteur et leurs droits moraux sur leurs publications et les utilisateurs s'engagent à reconnaître et respecter les exigences légales associées à ces droits. Ainsi, les utilisateurs:

- Peuvent télécharger et imprimer une copie de toute publication du portail public aux fins d'étude ou de recherche privée;
- Ne peuvent pas distribuer le matériel ou l'utiliser pour une activité à but lucratif ou pour un gain commercial;
- Peuvent distribuer gratuitement l'URL identifiant la publication.

Si vous pensez que ce document enfreint le droit d'auteur, contacteznous en fournissant des détails. Nous supprimerons immédiatement l'accès au travail et enquêterons sur votre demande. The authors are exclusively responsible for the content of their research papers published in the series *Les Cahiers du GERAD*. Copyright and moral rights for the publications are retained by the authors and the users must commit themselves to recognize and abide the legal requirements associated with these rights. Thus, users:

- May download and print one copy of any publication from the public portal for the purpose of private study or research;
- May not further distribute the material or use it for any profitmaking activity or commercial gain;
- May freely distribute the URL identifying the publication.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

High-impact climate damages are often driven by compounding climate conditions. For Abstract : example, elevated heat stress conditions can arise from a combination of high humidity and temperature. To explore future changes in compounding hazards under a range of climate scenarios and with large ensembles, climate emulators can provide light-weight, data-driven complements to Earth System Models. Yet, only a few existing emulators can jointly emulate multiple climate variables. In this study, we present the Multi-resolution EmulatoR for CompoUnd climate Risk analysis: MERCURY. MERCURY extends multi-resolution analysis to a spatio-temporal framework for versatile emulation of multiple variables. MERCURY leverages data-driven, image compression techniques to generate emulations in a memory-efficient manner. MERCURY consists of a regional component that represents the monthly, regional response of a given variable to yearly Global Mean Temperature (GMT) using a probabilistic regression based additive model, resolving regional cross-correlations. It then adapts a reverse lifting-scheme operator to jointly spatially disaggregate regional, monthly values to grid-cell level. We demonstrate MERCURY's capabilities on representing the humid-heat metric, Wet Bulb Globe Temperature, 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.

**Keywords:** Compounding climate hazards, climate model emulators, multi-resolution analysis, lifting scheme extended to additional climate variables, spatio-temporal emulation

**Acknowledgements:** We thank the climate modelling groups listed in Table 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. JC would like to acknowledge funding by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by Fonds de Recherche du Québec Nature et Technologies (FRQNT). 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 number 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).

# 1 Introduction

High impact climatic events are often driven by a combination of physical variables acting together (Raymond et al., 2020; Lesk et al., 2022). 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 events 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 to information redundancy. This provides an entry point for low-cost statistical emulators focussing on application relevant output indicators. 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 (Snyder et al., 2019; Liu et al., 2023; Schöngart et al., 2024; Bassetti et al., 2024), or jointly sample multivariate fields based on time sampling approaches. Time sampling approaches however, are hindered under scenarios where not many analogue 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-resolution approach thus representing spatio-temporal climate fields as a series of closed sub-spaces, allowing easy extension to multiple climate variables. This approach condenses key information of large-scale responses under climate change, whilst preserving localized small-scale features. It furthermore allows model reduction when representing the overall mean climate responses across multiplt variables – which has previously been emphasised 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 fields (Wim Sweldens, 1996; Park & Oh, 2022) 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 fields into subspaces and compress them into their key features, whilst maintaining their average values. Recently, Carreau & 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 GMT. The second component then uses the lifting scheme to "zoom" into and generate the higher resolution, grid-cell level fields. 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. 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 modelling monthly spatially resolved fields for a given set of variables conditional on the yearly Global Mean Temperature  $(GMT_y)$ . MERCURY's framework is summarised in Figure 1, and is composed of a regional, component generating monthly (m), regional (r) mean responses of a climate variable to GMT (described in Subsection 2.1). Grid-cell level (gc) values are then reconstructed from given regional, monthly values using a month-specific lifting scheme (described in Subsection 2.2). Data used for training and evaluating MERCURY are described in SI section 1.



Figure 1: MERCURY's framework for generating monthly spatially multivariate climate fields. Yearly GMT 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).

## 2.1 Representing mean regional monthly responses

To regress the response variable V on the yearly time series of GMT, a classical Additive Model (AM) is applied for each month and for each AR6 region, i.e.

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

where m = 1, ..., 12, r = 1, ..., 44 and  $V_{r,m}$  represents the climatological response during month mand over region r (either on temperature or relative humidity). As the relationship between  $V_{r,m}$  and  $GMT_y$  may not be necessarily linear,  $f_{r,m}(.)$  denotes a smooth function modelled by a classical cubic spline. The random vector  $\epsilon_{r,m}$  corresponds to a zero-mean Gaussian vector with covariance matrix  $\Sigma_{r,m}$ . This matrix represents monthly regional variability, see Figures S1 and S2 for visual checks of the Gaussian hypothesis and serial correlation diagnostics that indicate only significant positive auto-correlations in tropical latitudes. We illustrate (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  $\epsilon_{r,m}$ .

G-2024-81

## 2.2 Grid-cell level reconstruction per region

After fitting model (1) for each month and each AR6 region, the estimated  $V_{r,m}$  contains the main response signal and can be used to generate monthly time series of regional values for a given yearly GMT trajectory. To then obtain the corresponding monthly, grid-cell level values for a given region, we employ a lifting scheme based framework, such that:

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

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 how the reverse lifting-scheme operator,  $\Lambda_{r,m}$ , is 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 datasets by Carreau & Naveau (2023), and are depicted in Figure 2.

Starting from the original input grid, grid-cells are split into x-y pairs across each latitudinal band, unless an odd number of grid-cells exists in which case one group of three is made. Following this, the wavelet coefficients, d are the errors of a prediction with naive regression as,

$$d = y - x \tag{3}$$

and where a group is of size three, two wavelet coefficients are obtained,

$$d_1 = y_1 - x$$
 and  $d_2 = y_2 - x$ . (4)

Finally, the wavelet coefficients are set aside and the x values in each pair are updated with the scaling coefficients, c, corresponding to the average over each group,

$$c = \frac{x+y}{2} = x + \lambda \cdot d = x + \frac{1}{2} \cdot d \quad \text{or} \quad c = \frac{x+y_1+y_2}{3} = x + \lambda \cdot (d_1 + d_2) = x + \frac{1}{3} \cdot (d_1 + d_2), \quad (5)$$

where  $\lambda$  contains information on the group dimension.

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. It should be noted that the  $\lambda$  coefficient also needs to be tracked at each iteration, to account for differences in group dimensions being merged during the update step. For example, in the second iteration of the toy example depicted in Figure 2, the  $\lambda$  coefficient would be  $\frac{3}{7}$  to account for the differences in the group sizes.



Figure 2: Toy example of the lifting scheme applied on a grid consisting of 7 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 6 cells (in the toy example's case two iterations). Within each lifting iteration the grid's spatial dimension is reduced by approximately a half.

When reversing the lifting scheme, we iterate backwards from the scaling coefficient of the final iteration to the original grid. At each iteration, the x and y values within the corresponding group are obtained as,

$$x = c - \lambda \cdot d$$
 and  $y = d + x = d + (c - \lambda \cdot d) = c + (1 - \lambda) \cdot d.$  (6)

As a simple interpretative summary, the lifting framework decomposes a spatio-temporal field into a single time series i.e., scaling coefficient, that contains the spatial average at each time step and its surrounding spatial patterns i.e., wavelet coefficients, that summarise its spatial structures - as deviations from the average - at a given time step. It can thus be thought of as a simpler alternative to an Empirical Orthogonal Function (EOF) 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). Alternatively, it could be thought of as a naive Convolutional Network composed of 2-D average pooling layers, but suited for irregular grids with adaptive filter sizes and having deliberate attention focused towards remembering the residual spatial patterns.

Once, fully iterated over each region and month, the lifting scheme stores, from the last iteration, a configuration of scaling coefficients (treated as the monthly, regional values), 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, we may use it on estimated monthly regional values  $V_{r,m}$ , see (1), 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 (2), is able to go back down to a grid-cell level resolution given a single monthly, regional value, interchangeable as the fully lifted scaling coefficient. 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 start by defining the neighbourhood around a given regional, monthly value,  $V_{r,m}$  by choosing the twenty scaling values, corresponding to different years, within the lifting scheme's decomposition that are closest to it in value. The associated prototypical wavelet coefficients – referred to as wavelet patterns – are then sampled using Monte Carlo sampling. The Monte Carlo sampling builds a multivariate Gaussian distribution using the 20 wavelet coefficient vectors and samples from that. It assumes normality and stationarity in the wavelet coefficients, and draws from a multivariate Gaussian distribution with covariance matrix constructed across the wavelet coefficients, thus conserving spatial structures within the wavelet coefficients.

Each sampled wavelet pattern represents a possibility of spatial patterns corresponding to the given regional, monthly value and is used to invert the lifting scheme framework and arrive at a final spatially resolved emulation for that region and month. When doing this using regional monthly values provided across all regions as provided by  $f_{r,m}$  (see (1)), we arrive at a single global emulation for a given month. We note that regional cross-correlations are conserved within  $f_{r,m}$ , and given that the lifting scheme conserves the regional average value, discontinuous jumps between regions are expected to be insignificant. Nevertheless, to ensure seamless blending between regions during reconstruction, we apply a buffer zone around each regional boundary before sampling wavelet coefficients, as previously done to overcome boundary issues for image processing exercises (Hee-Seok et al., 2001; Naveau & Oh, 2004). The buffer zone of width one grid-cell is chosen and the lifting scheme is configured using it. Reverse lifting is conducted using for the whole region plus the buffer zone, after which buffer zone values are simply discarded.

#### 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 neighbourhood for sampling wavelet coefficients. In our case, we select  $T_{r,m,y}$ , as its relationship to  $GMT_y$  is most established (Tebaldi & Arblaster, 2014; Tebaldi & Knutti, 2018; Herger et al., 2015). Having defined the neighbourhood 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 neighbourhood. 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,

$$\Sigma_{RH|T} = \Sigma_{RH,RH} - \Sigma_{RH,T} \cdot \Sigma_{T,T}^{-1} \cdot \Sigma_{T,RH},$$
(7)

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

$$\begin{bmatrix} \Sigma_{T,T} & \Sigma_{T,RH} \\ \Sigma_{RH,T} & \Sigma_{RH,RH} \end{bmatrix}$$
(8)

A strict conditionality of the distributional space sampled by the Monte Carlo routine on RH to the key variable T is imposed. To this extent, spatially co-occurring and thus compounding patterns are expected to be conserved within the reconstruction process.

## 2.4 Evaluation

In evaluating MERCURY, ten 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 1000 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):

$$WBT_{gc,m,y} = T_{gc,m,y} \cdot (c_1 \cdot \sqrt{RH_{gc,m,y} + c_2}) + \tan^{-1}(T_{gc,m,y} + RH_{gc,m,y}) - \tan^{-1}(RH_{gc,m,y} - c_3) + c_4 \cdot RH_{gc,m,y}^{\frac{3}{2}} \cdot \tan^{-1}(c_5 \cdot RH_{gc,m,y}) - c_6,$$
(9)

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,

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

$$\tag{10}$$

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 for a quantile,  $q_{gc,m}$ , are calculated by calculating quantile time series from the generated emulations for quantile q, grid-cell gc and month m. The proportion of time steps that the ESM values appear below the emulated quantile value is then calculated  $(q_{gc,m}^{ESM})$ . The quantile deviation is then  $q_{gc,m}^{ESM} - q_{gc,m}$ , 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 a regionally aggregated level by means of probability rank distributions. Multiple layers of aggregation – calculated using a latitudinally weighted average – from AR6 regions to continental to global are investigated. 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 towards correlations with grid cells in equatorial latitudinal bands (i.e. -1.25°N). Since the emulator is a statistical approximation of the ESMs with an approximated spatial covariance matrix, such underestimation of spatial structures is not entirely unexpected. Moreover, equatorial latitudes have stronger temporal correlations (see Durban-Watson test for serial correlation Figure S2, and this indicates some shortcoming in MERCURY's design choice of not accounting for serial correlations. Across all ESMs and

G-2024-81



Difference in spatial spearman correlation matrix: ESM-Emulator

Figure 3: Difference between the ESM 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).

Comparing the results from Figure 3 to inter-comparison results against the existing monthly temperature emulator MESMER-M (Figure S3), we note that ESMs which had overall poorer CRPSS 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. In case readers are interested in an in-depth analysis of spatial correlations for select grid-cells please refer to SI section 4.

### 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 e.g., 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.

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 ESM 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 1000 WBGT emulations for each ESM, so a total of 5000 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, South Asia displays two modes of ESM WBGT values, starting at approximately 22 °C and 24 °C in the year 1850 and increasing at different rates till they converge around 24°C-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 wet bulb globe temperature (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 (1)). Region-to-region correlations are then added by means of a variable- and monthspecific multivariate Gaussian process. 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 performs an efficient compression of regional fields based on a local regression, to iteratively split and compress irregularly shaped fields into their key features. The differences between each iteration are stored as wavelet coefficients, thus enabling reconstruction of the original regional field from a single, regional averaged value.



Figure 5: 2-D histogram time series for the emulated superensemble (greys), with the actual ESM superensemble overlaid for July, SSP 2-4.5. 1000 emulations were produced for each ESM initial-condition ensemble member. Note, that snce historical runs have more ensemble members, the emulation count is also higher up to 2015.

The generation of grid-cell level emulations is performed by sampling wavelet coefficients within the "neighbourhood" 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 neighbourhood. In this study, we select a key variable, temperature, through which to define the neighbourhood which is then used in sampling wavelet coefficients for any additional variables, i.e., relative humidity. In such, we impose a strict, hierarchical dependency of relative humidity on temperature by defining the neighbourhood using temperature only and also by sampling relative humidity using its conditional covariance matrix to temperature. It would be possible to instead have a more egalitarian approach with a mutually defined neighbourhood between variables from which to sample from. In future, a mutually defined neighbourhood 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 approached 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 regions. MERCURY doesn't require to derive high resolution information for all regions included, but one 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 selected impact-relevant regional information without the need to generate and sift through global fields which amount to Peta-bytes of data. To the best of our knowledge, this is the only emulator approach existing so far aiming to tackle the data

management problem effectively. 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.

# **Open research section**

MESMER and MESMER-M are publicly available under Github (Hauser et al., 2021; Nath, 2022). Code for the lifting scheme can be found within the GitHub repository: https://github.com/snathxoc/Lifting. The CMIP6 data are available from the public CMIP archive at https://esgf-node. llnl.gov/projects/esgf-llnl/ (ESGF-LNLL, 2022).

## References

- Alexeeff, Stacey E., Nychka, Doug, Sain, Stephan R., & Tebaldi, Claudia. 2018. Emulating mean patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments. Climatic Change, 146(3–4), 319–333.
- Baldwin, Jane W., Benmarhnia, Tarik, Ebi, Kristie L., Jay, Ollie, Lutsko, Nicholas J., & Vanos, Jennifer K. 2023. Humidity's Role in Heat-Related Health Outcomes: A Heated Debate. Environmental Health Perspectives, 131(5).
- Bassetti, Seth, Hutchinson, Brian, Tebaldi, Claudia, & Kravitz, Ben. 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).
- Beusch, Lea, Gudmundsson, Lukas, & Seneviratne, Sonia 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.
- Carreau, Julie, & Naveau, Philippe. 2023. A spatially adaptive multi-resolution generative algorithm: Application to simulating flood wave propagation. Weather and Climate Extremes, 41(May), 100580.
- Daubechies, Ingrid. 1992. Ten Lectures on Wavelets. Philadelphia, Pennsylvania: Society for industrial and applied mathematics.
- Hauser, Mathias, Beusch, Lea, Nicholls, Zebedee, & Schwaab, Jonas. 2021. MESMER-group/mesmer: version 0.8.3.
- Hee-Seok, O. H., Naveau, Philippe, & Lee, Geunghee. 2001. Polynomial boundary treatment for wavelet regression. Biometrika, 88(1), 291–298.
- Herger, Nadja, Sanderson, Benjamin M., & Knutti, Reto. 2015. Improved pattern scaling approaches for the use in climate impact studies. Geophysical Research Letters, 42(9), 3486–3494.
- Kitsios, Vassili, O'Kane, Terence John, & Newth, David. 2023. A machine learning approach to rapidly project climate responses under a multitude of net-zero emission pathways. Communications Earth & Environment, 4(1), 355.
- Lesk, Corey, Anderson, Weston, Rigden, Angela, Coast, Onoriode, Jägermeyr, Jonas, McDermid, Sonali, Davis, Kyle F., & Konar, Megan. 2022. Compound heat and moisture extreme impacts on global crop yields under climate change. Nature Reviews Earth & Environment, 3(12), 872–889.
- Liu, Ziming, Luo, Di, Xu, Yilun, Jaakkola, Tommi, & Tegmark, Max. 2023. GenPhys: From Physical Processes to Generative Models. 1–21.
- Nath, Shruti. 2022. snath-xoc/Nath\_et\_al\_ESD\_2022\_MESMER-M. 4.
- Nath, Shruti, Lejeune, Quentin, Beusch, Lea, Seneviratne, Sonia I., & Schleussner, Carl Friedrich. 2022. MESMER-M: an Earth system model emulator for spatially resolved monthly temperature. Earth System Dynamics, 13(2), 851–877.
- Nath, Shruti, Hauser, Mathias, Schumacher, Dominik L., Lejeune, Quentin, Gudmundsson, Lukas, Quilcaille, Yann, Candela, Pierre, Saeed, Fahad, Seneviratne, Sonia I., & Schleussner, Carl-Friedrich. 2024. Representing natural climate variability in an event attribution context: Indo-Pakistani heatwave of 2022. Weather and Climate Extremes, 44(6), 100671.

- Naveau, Philippe, & Oh, Hee Seok. 2004. Polynomial wavelet regression for images with irregular boundaries. IEEE Transactions on Image Processing, 13(6), 773–781.
- Park, Seoncheol, & Oh, Hee-Seok. 2022. Lifting Scheme for Streamflow Data in River Networks. Journal of the Royal Statistical Society Series C: Applied Statistics, 71(2), 467–490.
- Quilcaille, Yann, Gudmundsson, Lukas, Beusch, Lea, Hauser, Mathias, & Seneviratne, Sonia I. 2022. Showcasing MESMER-X: Spatially Resolved Emulation of Annual Maximum Temperatures of Earth System Models. Geophysical Research Letters, 49(17), 1–11.
- Raymond, Colin, Matthews, Tom, & Horton, Radley M. 2020. The emergence of heat and humidity too severe for human tolerance. Science Advances, 6(19).
- Schöngart, Sarah, Gudmundsson, Lukas, Hauser, Mathias, Pfleiderer, Peter, Lejeune, Quentin, Nath, Shruti, Seneviratne, Sonia Isabelle, & Schleußner, Carl-Friedrich. 2024 (5). Introducing the MESMER-M-TPv0.1.0 module: Spatially Explicit Earth System Model Emulation for Monthly Precipitation and Temperature.
- Snyder, Abigail, Link, Robert, Dorheim, Kalyn, Kravitz, Ben, BondLamberty, Ben, & Hartin, Corinne. 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.
- Stull, Roland. 2011. Wet-bulb temperature from relative humidity and air temperature. Journal of Applied Meteorology and Climatology, 50(11), 2267–2269.
- Tebaldi, Claudia, & Arblaster, Julie M. 2014. Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. Climatic Change, 122(3), 459–471.
- Tebaldi, Claudia, & Knutti, Reto. 2018. Evaluating the accuracy of climate change pattern emulation for low warming targets. Environmental Research Letters, 13(5).
- Tebaldi, Claudia, Snyder, Abigail, & Dorheim, Kalyn. 2022. STITCHES: creating new scenarios of climate model output by stitching together pieces of existing simulations. Earth System Dynamics, 13(4), 1557–1609.
- Wim Sweldens. 1996. The lifting scheme: A construction of second generation wavelets. SIAM Journal on Mathematical Analysis.
- Zscheischler, Jakob, Martius, Olivia, Westra, Seth, Bevacqua, Emanuele, Raymond, Colin, Horton, Radley M., van den Hurk, Bart, AghaKouchak, Amir, Jézéquel, Aglaé, Mahecha, Miguel D., Maraun, Douglas, Ramos, Alexandre M., Ridder, Nina N., Thiery, Wim, & Vignotto, Edoardo. 2020. A typology of compound weather and climate events. Nature Reviews Earth & Environment, 1(7), 333–347.