Supplementary Information for Eulerian Parameter Inference: A Probabilistic Change of Variables for Model-Based Inference with High-Variability Data Sets

Vincent Wagner[1†], Benjamin Castellaz[1†], Lars Kaiser[1], Sebastian Höpfl[1], Nicole Radde[1*]

[1]Institute for Stochastics and Applications, University of Stuttgart, Pfaffenwaldring 9, Stuttgart, 70569, Germany.

*Corresponding author(s). E-mail(s): nicole.radde@simtech.uni-stuttgart.de;
[†]These authors contributed equally to this work.

Appendix A Supplement to the arithmetic climate model

In Eq. (3), we attempted to model the annual average temperature y of a location on Earth purely based on its latitude coordinate q. We are aware that, in doing so, we neglect many thermodynamic processes crucial to explaining a phenomenon as complex as Earth's climate. The assumption that the average temperature should be approximately proportional to the local intensity of sunlight mainly motivates our model and lies at our model's core. By averaging over all seasons and neglecting diffusion and transport phenomena, this again is proportional to the cosine of the location's latitude. Together with the estimated temperatures of 30°C at the equator and −30°C close to the poles, the named assumptions lead to the defined model

$$s : [0°, 90°] \to [-30, 30], \; q \mapsto y = s(q) := 60 \cdot \cos(q) - 30. \tag{A1}$$

Petty [1] proposes a similar yet different model. He especially considers Earth's black-body radiation and formulates a thermodynamic equilibrium that dictates a

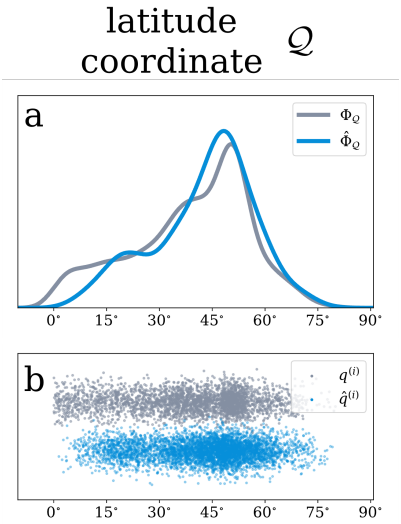

Fig. A1 True and inferred latitude distribution align closely for model (3). a. KDE of the true and inferred latitude parameter sample depicted in **b**. The true sample and KDE approximation are visualised in grey, while the EPI reconstruction is given in blue.

temperature roughly proportional to the fourth root of the irradiation intensity. Similar to our approach, this intensity is approximately proportional to the cosine of the location's latitude. Consequently, their model reads

$$s : [0°, 90°] \to [-30, 30], \; q \mapsto y = s(q) := 60 \cdot \sqrt[4]{\cos(q)} - 30. \tag{A2}$$

Both model formulations yield inferred EPI latitude distributions. Given the availability of true latitude coordinates for all average temperature data points in this example, we can choose the model whose inferred latitude distribution $\hat{\Phi}_{\mathcal{Q}}$ best aligns with the true latitude samples $q^{(i)}$. This comparison clearly favoured the model defined by Eq. (3), for which true and inferred latitudes are depicted in Figure A1. This accurate fit confirms the proper functioning of our inference scheme and the ability of simulation model (3) to explain the data effectively.

A.1 Data acquisition and setup of the SIP

For the annual average temperature data, we used the `meteostat Python` library [2] (Source: Meteostat) to extract the monthly average temperatures of all available weather stations for the years $2020 - 2022$. We only considered stations with 36 reported monthly temperatures and agglomerated the corresponding data to calculate a 3-year average temperature for each station. The data used in this paper were downloaded on 02.07.2023. This led to 3 168 data points for which we obtained the 3-year average temperature and the true latitudinal coordinate.
The `eulerpi` parameter estimation used MCMC sampling with an ensemble of 50 walkers that performed 1 250 steps each. From each resulting sampling trajectory, we

discarded the initial 250 samples as a burn-in. We only used every 10th sample to obtain a result of 5 000 relatively uncorrelated samples from the underlying parameter distribution $\hat{\Phi}_{\mathcal{Q}}$. All sampling parameters are summarised in Table A1.

Parameter	**Value**
Number of sampling walkers	50
Number of sampling steps per walker	1 250
Number of burn-in sampling steps	250
Thinning factor	10

Table A1 Sampling parameters for results related to the annual average temperature and COVID-19 infection model.

Appendix B Derivation of the change of variables formula

The following derivation is adapted from Rezende et al. in the appendix of [3]. We provide a version of their proof adapted to the EPI setting of the following assertion: Let $\Omega_{\mathcal{Q}} = \mathbb{R}^N$, and $\Omega_{\mathcal{Y}} = s(\Omega_{\mathcal{Q}}) \subset \mathbb{R}^M$, and let $s : \Omega_{\mathcal{Q}} \to \Omega_{\mathcal{Y}}$ be a diffeomorphism. Assume that s can be extended to a smooth map between open neighbourhoods in $\mathbb{R}^N$ and $\mathbb{R}^M$, and choose this extension. Then, it holds that

$$\Phi_{\mathcal{Q}}(q) = \Phi_{\mathcal{Y}}(s(q)) \cdot \sqrt{\det\left(\frac{ds}{dq}(q)^\intercal \frac{ds}{dq}(q)\right)} \tag{B3}$$

where $\frac{ds}{dq}(q)$ is the Jacobian of said extension of s evaluated in q.

Proof. First, note that for a set of vectors $y_1, \ldots, y_N \in \mathbb{R}^M$, the volume spanned by those vectors is given by $\sqrt{\det(\mathbf{Y}^\intercal \mathbf{Y})}$, where $\mathbf{Y}$ denotes the matrix with column vectors $y_1, \ldots, y_N$. Now, let $e_1, \ldots, e_N$ denote the canonical basis vectors of $\mathbb{R}^N = \Omega_{\mathcal{Q}}$. For given parameters q, the pushforward maps each basis vector e_i to $\frac{ds}{dq}(q)e_i$. Collecting all transformed vectors as columns of a matrix yields

$$\left(\frac{ds}{dq}(q)e_i\right)_{i=1,\ldots,N} = \frac{ds}{dq}(q)\mathbb{I}_N = \frac{ds}{dq}(q) \tag{B4}$$

where $\mathbb{I}_N$ is the N-dimensional identity matrix. The volume spanned by the transformed vector is thus given by $\sqrt{\det\left(\frac{ds}{dq}(q)^\intercal \frac{ds}{dq}(q)\right)}$. Hence, the transformed density is given by

$$\Phi_{\mathcal{Y}}(s(q)) = \frac{\Phi_{\mathcal{Q}}(q)}{\sqrt{\det\left(\frac{ds}{dq}(q)^\intercal \frac{ds}{dq}(q)\right)}} \tag{B5}$$

Multiplying Eq. (B5) by $\sqrt{\det\left(\frac{ds}{dq}(q)^\intercal \frac{ds}{dq}(q)\right)}$ yields the claim. □

Note that in case $N = M$, Eq. (7) can be simplified to the well-known version of the change of variables formula

$$\Phi_{\mathcal{Q}}(q) = \Phi_{\mathcal{Y}}(s(q)) \cdot \det\left(\frac{ds}{dq}(q)\right). \tag{B6}$$

Appendix C Theoretical considerations when evaluating the data density estimate in general dimensions

Since s has generally more output dimensions than input dimensions and under the assumptions imposed by EPI, the image $\Omega_{\mathcal{Y}}$ of s is a manifold in the data space. For Eq. (6) to be valid, $\hat{\Phi}_{\mathcal{Y}}$ would now have to be a density estimator on this manifold – in general, a challenging task. We propose to calculate a KDE in the higher-dimensional embedding space $\mathbb{R}^M$ and then use the restriction on $\Omega_{\mathcal{Y}}$ as $\hat{\Phi}_{\mathcal{Y}}$ instead. The resulting estimator looses the normalisation property of probability densities needed – hence, $\hat{\Phi}_{\mathcal{Q}}$ must be normalised. This can be done either by using MCMC sampling or by numerical integration. The described data density evaluation is visualised in Figure C2.

We have not yet been able to derive any guarantees for consistency and convergence of this type of density estimation – in particular, it is easy to construct pathological cases in which this method will certainly fail. However, if the curvature of $\Omega_{\mathcal{Y}}$ is small compared to the bandwidth of the density estimator and there are no points where the manifold almost intersects itself, this choice of estimator intuitively makes a lot of sense: points close to the image of s get assigned high probability density values, while points further away from $\Omega_{\mathcal{Y}}$ are associated with low values. This is likely the reason why we have always obtained good results with this estimator in practice. For an illustration of a potentially problematic scenario, consider Figure C2, where $\Omega_{\mathcal{Y}}$ indeed nearly intersects with itself. The influence of data points close to the critical point will influence the density estimation of both close parts of $\Omega_{\mathcal{Y}}$.

Appendix D Supplement to the COVID-19 SEIR-model

COVID-19 infection data were taken from the publicly available file of the RKI on 02.07.2023 [4]. We filtered the 7-day incidence data for the dates 09.03.2020, 16.03.2020, 06.04.2020, 11.05.2020, and 15.06.2020. All of the dates are Mondays and approximate the time points corresponding to 1, 2, 5, 10, and 15 weeks after the beginning of the first COVID-19 wave in Germany. This led to 5 incidence values for 411 German districts (or *Landkreise*), each. During parameter inference, we used MCMC sampling in `eulerpi` with the same hyperparameters already used for the climate model. They are summarised in Table A1. The data and comparable

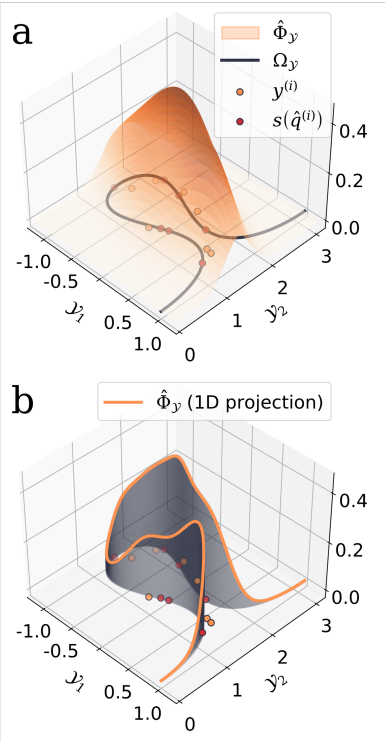

Fig. C2 Visualisation of the data density evaluation on a manifold. a. The two-dimensional data points $y^{(i)}$ are depicted as orange dots with black outline and lie close to the image $\Omega_{\mathcal{Y}}$ of s (grey curve). Exemplary model evaluations $s(\hat{q}^{(i)})$ are given as red dots and lie on $\Omega_{\mathcal{Y}}$ by definition. The KDE $\hat{\Phi}_{\mathcal{Y}}$ is presented as an orange surface and can readily be evaluated at any point in $\mathbb{R}^2$. **b.** By evaluating the data KDE $\hat{\Phi}_{\mathcal{Y}}$ only in simulation results from the image of s, we effectively project $\hat{\Phi}_{\mathcal{Y}}$ to the 1D manifold $\Omega_{\mathcal{Y}}$. Orange and red dots are identical in both graphs.

simulation results are reported in Figure 3c. Please note that the axis of the graph is non-linearly scaled so that the time points 1, 2, 5, 10, and 15 are equidistantly distributed over the width of the figure. The scaled time t is obtained by transforming the linear time t_{ref} using the polynomial

$$\begin{aligned} t : [0,1] &\to [0.703125, 16.3281], \\ t_{\text{ref}} &\mapsto t(t_{\text{ref}}) = \frac{45}{64} + \frac{25}{6} t_{\text{ref}} - \frac{475}{24} t_{\text{ref}}^2 + \frac{500}{6} t_{\text{ref}}^3 - \frac{625}{12} t_{\text{ref}}^4. \end{aligned} \tag{D7}$$

This transformation is used to plot $s(\hat{q}^{(i)})$ over a continuous time t and was derived as the analytical, polynomial interceptor for the points

$$(t_{\text{ref}}, t) = (0.1, 1),\ (0.3, 2),\ (0.5, 5),\ (0.7, 10),\ (0.9, 15). \tag{D8}$$

Appendix E Supplement to the anisotropic heat model

Formally, we consider the following initial boundary value problem (IBVP): Let $U \coloneqq [0,1] \times [0,1]$ and let $T \in \mathbb{R}^+$. Find $u : U \times [0,T] \to \mathbb{R}$, $u : (x,t) \mapsto u(x,t)$ with $u \in C^2\left((U \setminus \partial U) \times]0,T[\to \mathbb{R}\right) \cap C\left((U \times [0,T] \to \mathbb{R})\right)$ such that

$$\begin{aligned} \partial_t u(x,t) &= \operatorname{div}(K \cdot \nabla u(x,t)) \text{ on } U \times [0,T] \\ u(x,0) &= f(x) \text{ on } U \times \{0\} \\ u(x,t) &= g(x) \text{ on } \partial U \times]0,T] \end{aligned} \tag{E9}$$

for f, g continuously differentiable outside a nullset.

E.1 Numerical solution

To keep the numerical solution method as simple as possible, solutions to Eq. (E9) are approximated using the Method of Lines. This approach is adapted from Schiesser [5]; refer to him for a detailed analysis of the method. Spatial discretisation of Eq. (E9) is performed using second-order central differences. The spatial domain is discretised on a regular grid, i.e.

$$x_{i,j} = \begin{pmatrix} i \cdot h_1 \\ j \cdot h_2 \end{pmatrix} \text{ for } i = 0, \ldots, n_1 \text{ and } j = 0, \ldots, n_2 \tag{E10}$$

where $h_1 = \frac{1}{n_1}$ and $h_2 = \frac{1}{n_2}$ are the grid widths in x_1 and x_2 direction, respectively. Then, $u_{i,j}(t)$ is an approximation to the solution at time t:

$$u_{i,j}(t) \approx u(x_{i,j}, t) \tag{E11}$$

The right-hand side of Eq. (E9) can be reorganised as

$$\operatorname{div}\left(K \cdot \nabla u(x,t)\right) = \kappa_{11} \frac{\partial u^2}{\partial x_1^2}(x,t) + 2\kappa_{12} \frac{\partial u^2}{\partial x_1 \partial x_2}(x,t) + \kappa_{22} \frac{\partial u^2}{\partial x_2^2}(x,t) \tag{E12}$$

using $u \in C^2$ for the third identity. The terms on the right-hand side of Eq. (E12) can be approximated using standard second-order central differences:

$$\begin{aligned} \frac{\partial u^2}{\partial x_1^2}(x_{i,j},t) &\approx \frac{u_{i+1,j}(t) - 2u_{i,j}(t) + u_{i-1,j}(t)}{h_1^2} \\ \frac{\partial u^2}{\partial x_1 \partial x_2}(x_{i,j},t) &\approx \frac{u_{i-1,j-1}(t) - u_{i+1,j-1}(t) - u_{i-1,j+1}(t) + u_{i+1,j+1}(t)}{4h_1 h_2} \\ \frac{\partial u^2}{\partial x_2^2}(x_{i,j},t) &\approx \frac{u_{i,j+1}(t) - 2u_{i,j}(t) + u_{i,j-1}(t)}{h_2^2} \end{aligned} \tag{E13}$$

for $i = 1, \ldots, n_i - 1$ and $j = 1, \ldots, n_j - 1$. By substituting Eq. (E13) into Eq. (E12) and the result into Eq. (E9) yields

$$\begin{aligned}\partial_t u_{i,j}(t) \approx \kappa_{11} &\frac{u_{i+1,j}(t) - 2u_{i,j}(t) + u_{i-1,j}(t)}{h_1^2} \\ &+ 2\kappa_{12} \frac{u_{i-1,j-1}(t) - u_{i+1,j-1}(t) - u_{i-1,j+1}(t) + u_{i+1,j+1}(t)}{4h_1 h_2} \\ &+ \kappa_{22} \frac{u_{i,j+1}(t) - 2u_{i,j}(t) + u_{i,j-1}(t)}{h_2^2}.\end{aligned} \tag{E14}$$

This system of ODEs can now be solved by an ODE solver.
For time integration, Heun's method is used. Heun's method is a two-stage Runge-Kutta method of second order. The following description of the method is adapted from [6]. For this, the time axis is discretised too: let the time step width $h_t = \frac{T}{n_t}$ for some number of time steps $n_t \in \mathbb{N}$. Then, the discrete time points are given as

$$t_k \coloneqq k \cdot h_t \text{ for } 0 = 1, \ldots, n_t \tag{E15}$$

Then, $u^k(\cdot)$ is an approximation of the solution at time t_k:

$$u^k(\cdot) \approx u(\cdot, t_k) \tag{E16}$$

Heun's method computes approximate solutions to ODEs with the form

$$\partial_t u(t) = f(u, t) \tag{E17}$$

and is given as

$$u^{k+1} = u^k + \frac{1}{2} h_t \left(f(u^k, t_k) + f(u_k + h_t f(u_k, t_k), t_k + h_t) \right) \tag{E18}$$

First, the following notation for the full discretisation is useful:

$$u_{i,j}^k \approx u(x_{i,j}, t_k) \tag{E19}$$

For the initial values, set

$$u_{i,j}^0 \coloneqq f(x_{i,j}) \tag{E20}$$

For the Dirichlet boundaries, analogously set

$$u_{i,j}^k \coloneqq g(x_{i,j}) \text{ for } (i,j) \in \{0, n_1\} \times \{0, n_2\} \text{ and } k = 1, \ldots, n_t \tag{E21}$$

Substituting the right-hand side of Eq. (E14) for f in Eq. (E18) yields the overall numerical scheme: first, the preliminary time step $\tilde{u}_{i,j}^k$ for $i = 1, \ldots, n_1 - 1$, $j =$

$1, \ldots, n_2 - 1$ is computed:

$$\begin{aligned}\tilde{u}_{i,j}^{k+1} = u_{i,j}^k + h_t \Bigg(& \kappa_{11} \frac{u_{i+1,j}^k - 2u_{i,j}^k + u_{i-1,j}^k}{h_1^2} \\ & + 2\kappa_{12} \frac{u_{i-1,j-1}^k - u_{i+1,j-1}^k - u_{i-1,j+1}^k + u_{i+1,j+1}^k}{4h_1 h_2} \\ & + \kappa_{22} \frac{u_{i,j+1}^k - 2u_{i,j}^k + u_{i,j-1}^k}{h_2^2} \Bigg)\end{aligned} \tag{E22}$$

Then, the next time step is given as:

$$\begin{aligned}u_{i,j}^{k+1} = & \, u_{i,j}^k \\ & + \frac{1}{2} h_t \Bigg(\frac{\kappa_{11}}{h_1^2} \left(u_{i+1,j}^k - 2u_{i,j}^k + u_{i-1,j}^k + \tilde{u}_{i+1,j}^{k+1} - 2\tilde{u}_{i,j}^{k+1} + \tilde{u}_{i-1,j}^{k+1} \right) \\ & + \frac{2\kappa_{12}}{4h_1 h_2} \Big(u_{i-1,j-1}^k - u_{i+1,j-1}^k - u_{i-1,j+1}^k + u_{i+1,j+1}^k \\ & + \tilde{u}_{i-1,j-1}^{k+1} - \tilde{u}_{i+1,j-1}^{k+1} - \tilde{u}_{i-1,j+1}^{k+1} + \tilde{u}_{i+1,j+1}^{k+1} \Big) \\ & + \frac{\kappa_{22}}{h_2^2} \left(u_{i,j+1}^k - 2u_{i,j}^k + u_{i,j-1}^k + \tilde{u}_{i,j+1}^{k+1} - 2\tilde{u}_{i,j}^{k+1} + \tilde{u}_{i,j-1}^{k+1} \right) \Bigg)\end{aligned} \tag{E23}$$

for $i = 1, \ldots, n_1 - 1,\ j = 1, \ldots, n_2 - 1$ and $k = 1 \ldots, n_t$. The Method of Lines imposes a time step condition for the method to be numerically stable - see Schiesser [5] for details. The constraint is given by

$$h_t < \frac{\left(\min(h_1, h_2)\right)^2}{4\kappa_{\max}} \tag{E24}$$

where $\kappa_{\max}$ is the largest eigenvalue of K. An overview of the actual values used in the following investigation can be found in Table E2.

Parameter	**Description**	**Value**
$T = t_{\text{end}}$	Simulation end time	0.1
$n_1 + 1$	Number of spatial points in x_1 direction	21
$n_2 + 1$	Number of spatial points in x_2 direction	21
h_1	Grid width in x_1 direction	0.05
h_2	Grid width in x_2 direction	0.05
h_t	Time step width	$< 0.5 \cdot 10^{-3} \kappa_{\max}^{-1}$

Table E2 Parameters for the numerical solution of the two-dimensional anisotropic IBVP arising from Eq. (8).

Parameter	Description	Value
$\mathcal{Q}_1 = \kappa_{11}$	Random thermal conductivity in x_1 direction	$1.2 + 0.6X_1$
$\mathcal{Q}_2 = \kappa_{22}$	Random thermal conductivity in x_2 direction	$1.2 + 0.6X_2$
$\mathcal{Q}_3 = \kappa_{12}$	Random thermal conductivity diagonal to x_1 and x_2	$0.2 + 0.6X_3$
X_1, X_2, X_3	Random parameter variables	$\overset{i.i.d.}{\sim}$ Beta(α, β)
(α, β)	Shape parameters of beta distribution	$(2, 5)$

Table E3 Setup of the SIP for the head diffusion model.

E.2 Setup of the SIP

To control the underlying assumptions and to be able to make statements about the quality of the inference in the parameter space, an entirely artificial setting was chosen for the third model problem. As parameters to be inferred, the entries of the conductivity matrix K were chosen, i.e.

$$\mathcal{Q}_1 \coloneqq \kappa_{11}, \;\; \mathcal{Q}_2 \coloneqq \kappa_{22}, \;\; \mathcal{Q}_3 \coloneqq \kappa_{12} \tag{E25}$$

To apply marginal EPI, the parameters must be stochastically independent. At the same time, the criterion (11) must always be fulfilled. This is achieved by choosing

$$\begin{aligned} \mathcal{Q}_1 &= 1.2 + 0.6\mathcal{B}_1 \\ \mathcal{Q}_2 &= 1.2 + 0.6\mathcal{B}_2 \\ \mathcal{Q}_3 &= 0.2 + 0.6\mathcal{B}_3 \\ \text{with } \mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3 &\overset{i.i.d.}{\sim} \text{Beta}(\alpha, \beta) \end{aligned} \tag{E26}$$

An overview of the choices discussed above can be found in Table E3. Figure E3 shows a sample from the chosen parameter distribution.

As discussed in the article, the time of the evaluations is restricted to the endpoint of the simulation, i.e. all evaluations are done at $t_{\text{end}} = T = 0.1$. Spatially, we use five evaluation points arranged as on a die: $X = (\binom{0.25}{0.25}, \binom{0.75}{0.25}, \binom{0.5}{0.5}, \binom{0.75}{0.75}, \binom{0.25}{0.75})$. This yields the following model s used for inference:

$$\begin{aligned} s : \mathbb{R}^3 \supseteq \Omega_{\mathcal{Q}} &\to \Omega_{\mathcal{Y}} \subseteq \mathbb{R}^5, \\ (\kappa_{11}, \kappa_{22}, \kappa_{12})^\intercal =: q &\mapsto y = s(q) \coloneqq (u(X_1, t_{\text{end}}), \ldots, u(X_5, t_{\text{end}}))^\intercal \end{aligned} \tag{E27}$$

We evaluated the three marginal parameter densities $\hat{\phi}_{\mathcal{Q}_j}, j = 1, 2, 3$ over an equidistant grid of 100 points, each. Please refer to Supplementary Section F for

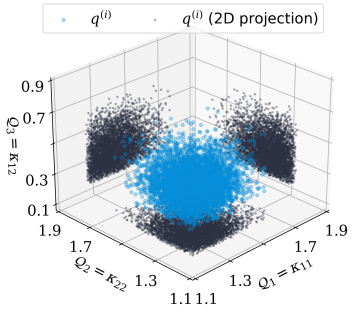

Fig. E3 A sample of size $n = 5\,000$ from the parameter distribution $\Phi_{\mathcal{Q}}$ that we use as true underlying parameters for the heat diffusion model in this work. Steps 1 and 2 in Figure 2 visualise this sample's creation.

details. Specifically, we chose the central parameter point $c = (1.5, 1.5, 0.5)^\intercal$ and the array of grid points defined by

$$q_i = \begin{cases} \frac{i}{100} + 1 \text{ for } j = 1, 2 \\ \frac{i}{100} \text{ for } j = 3 \end{cases} \tag{E28}$$

for $i = 0, \ldots, 100$. For MCMC sampling, the pushforward $s(\hat{q}^{(i)})$ of the inferred parameter sample $\hat{q}^{(i)}$ is generated on the fly, as each parameter sample is simulated anyway and each simulation result represents one sample of the random variable $s(\mathcal{Q})$. However, the grid-based evaluation in parameter point q_i only produces the associated densities $\hat{\phi}_{\mathcal{Q}_j}(q^{(i)})$ and no sample from $s(\mathcal{Q})$. From these independent marginal parameter densities, we therefore subsequently sampled 2 500 parameter samples $\hat{q}^{(i)}$, from which the pushforward could then be computed. All distribution approximations depicted in Figure 4 are obtained through KDEs. This indirectly ensures the normalisation of all $\hat{\phi}_{\mathcal{Q}_i}$.

Appendix F Direct evaluation of marginal distributions

Here, we prove an even stronger version of Eq. (12). Formally, let $\mathbb{R}^N \ni \mathcal{Q} := (\mathcal{Q}_1, \ldots, \mathcal{Q}_L)$ be a random parameter vector with $\mathcal{Q}_i \in \mathbb{R}^{n_i} \forall i = 1, \ldots, L$, such that $\sum_{i=1}^{L} n_i = N$. Furthermore, let $\Phi_{\mathcal{Q}}(q) = \Phi_{\mathcal{Q}}(q_1, \ldots, q_L)$ denote the probability density function of $\mathcal{Q}$ and $\phi_{\mathcal{Q}_i}(q_i)$ denote the corresponding marginal distributions. For $j \in \{1, \ldots, L\}$, choose fixed $c_i \in \mathbb{R}^{n_i}$ for $i = 1, \ldots, L$ with $i \neq j$ in the support of each $\phi_{\mathcal{Q}_i}$, that is, $\phi_{\mathcal{Q}_i}(c_i) \neq 0$. If all $\mathcal{Q}_i$ are independent, i.e. $\Phi_{\mathcal{Q}}(q_1, \ldots, q_L) = \prod_{i=1}^{L} \phi_{\mathcal{Q}_i}(q_i)$, it holds that

$$\phi_{\mathcal{Q}_j}(q_j) = C_j \cdot \Phi_{\mathcal{Y}}\big(s(\tilde{q})\big) \sqrt{\det\left(\frac{ds}{dq}(\tilde{q})^\intercal \frac{ds}{dq}(\tilde{q})\right)} \tag{F29}$$

with $\mathbb{R}^N \ni \tilde{q} \coloneqq (c_1, \dots, c_{j-1}, q_j, c_{j+1}, \dots, c_L)$, where $C_j \in \mathbb{R}$ only depends on $c_1, \dots, c_{j-1}, c_{j+1}, \dots, c_L$.

Proof. Using the independence of $\mathcal{Q}_1, \dots, \mathcal{Q}_L$ and $\phi_{\mathcal{Q}_i}(c_i) \neq 0$, it holds that

$$\begin{aligned} \frac{\Phi_{\mathcal{Q}}(q_1, \dots, q_L)}{\Phi_{\mathcal{Q}}(\tilde{q})} &= \frac{\Phi_{\mathcal{Q}}(q_1, \dots, q_L)}{\Phi_{\mathcal{Q}}(c_1, \dots, c_{j-1}, q_j, c_{j+1}, c_L)} \\ &= \frac{\phi_{\mathcal{Q}_j}(q_j)}{\phi_{\mathcal{Q}_j}(q_j)} \prod_{i=1, i\neq j}^{L} \frac{\phi_{\mathcal{Q}_i}(q_i)}{\phi_{\mathcal{Q}_i}(c_i)} \\ \Rightarrow \Phi_{\mathcal{Q}}(q_1, \dots, q_L) &= \prod_{i=1, i\neq j}^{L} \frac{\phi_{\mathcal{Q}_i}(q_i)}{\phi_{\mathcal{Q}_i}(c_i)} \Phi_{\mathcal{Q}}(\tilde{q}) \end{aligned} \tag{F30}$$

Note the second identity also holds for $\phi_{\mathcal{Q}_j}(q_j) = 0$: If $\phi_{\mathcal{Q}_j}(q_j) = 0$, then $\Phi_{\mathcal{Q}}(q_1, \dots, q_L) = \Phi_{\mathcal{Q}}(c_1, \dots, c_{j-1}, q_j, c_{j+1}, c_L) = 0$ holds almost everywhere and, since the prerequisites for EPI imply $\Phi_{\mathcal{Q}} \in C^1$, this is true even point-wise. Integrating out the marginal distribution of $\mathcal{Q}_j$ yields

$$\begin{aligned} \phi_{\mathcal{Q}_j}(q_j) &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \Phi_{\mathcal{Q}}(q_1, \dots, q_L) dq_1 \cdots dq_{j-1} dq_{j+1} \cdots dq_L \\ &\overset{(F30)}{=} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{i=1, i\neq j}^{L} \frac{\phi_{\mathcal{Q}_i}(q_i)}{\phi_{\mathcal{Q}_i}(c_i)} \Phi_{\mathcal{Q}}(\tilde{q}) dq_1 \cdots dq_{j-1} dq_{j+1} \cdots dq_L \\ &= \left(\prod_{i=1, i\neq j}^{L} \frac{1}{\phi_{\mathcal{Q}_i}(c_i)} \right) \Phi_{\mathcal{Q}}(\tilde{q}) \\ &\cdot \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{i=1, i\neq j}^{L} \phi_{\mathcal{Q}_i}(q_i) dq_1 \cdots dq_{j-1} dq_{j+1} \cdots dq_L \\ &= \left(\prod_{i=1, i\neq j}^{L} \frac{1}{\phi_{\mathcal{Q}_i}(c_i)} \right) \Phi_{\mathcal{Q}}(\tilde{q}) \prod_{i=1, i\neq j}^{L} \underbrace{\int_{\mathbb{R}} \phi_{\mathcal{Q}_i}(q_i) dq_i}_{=1} \\ &= \left(\prod_{i=1, i\neq j}^{L} \frac{1}{\phi_{\mathcal{Q}_i}(c_i)} \right) \Phi_{\mathcal{Q}}(\tilde{q}) \end{aligned} \tag{F31}$$

where the linearity of the integral was used to obtain the third and fourth identities. Setting

$$C_j \coloneqq \prod_{i=1, i\neq j}^{L} \frac{1}{\phi_{\mathcal{Q}_i}(c_i)} \tag{F32}$$

one obtains

$$\begin{aligned}\phi_{\mathcal{Q}_j}(q_j) &\overset{(F31)}{=} C_j \cdot \Phi_{\mathcal{Q}}(\tilde{q}) \\ &\overset{(7)}{=} C_j \cdot \Phi_{\mathcal{Y}}\big(s(\tilde{q})\big)\sqrt{\det\left(\frac{ds}{dq}(\tilde{q})^{\intercal}\frac{ds}{dq}(\tilde{q})\right)}\end{aligned} \tag{F33}$$

where the transformation fundamental to EPI was used for the last identity. □

The corresponding marginal EPI is given by replacing $\Phi_{\mathcal{Y}}(s(\tilde{q}))$ by $\hat{\Phi}_{\mathcal{Y}}(s(\tilde{q}))$ in Eq. (12):

$$\hat{\phi}_{\mathcal{Q}_j}(q_i) = C_j \cdot \hat{\Phi}_{\mathcal{Y}}\big(s(\tilde{q})\big)\sqrt{\det\left(\frac{ds}{dq}(\tilde{q})^{\intercal}\frac{ds}{dq}(\tilde{q})\right)} \tag{F34}$$

Note that this result also allows for the assumption of independent parameter vectors. This can be used, for example, to test pairwise correlations by evaluating the marginal parameter distribution for all pairs $(\mathcal{Q}_i, \mathcal{Q}_j)$ for $i = 1, \ldots, N$ and $j = 1, \ldots, i-1$. Although Eq. (12) is exact, direct evaluation of the marginal parameter distributions may be less accurate in practice than evaluation of the entire distribution with subsequent marginalisation. This is because marginalising the estimate increases its robustness to inaccuracies in the multidimensional KDE. Hence, for finite data and when evaluating a single slice, the estimator's variance is higher than evaluating the whole space followed by marginalisation. Therefore, which of the two options is preferable depends on the available computation time and data.

References

[1] Petty GWGW. A first course in atmospheric radiation. 2nd ed. Madison, Wis: Sundog Publishing; 2006.

[2] Lamprecht CS.: Meteostat Python.

[3] Rezende DJ, Papamakarios G, Racaniere S, Albergo M, Kanwar G, Shanahan P, et al.: Normalizing flows on tori and spheres.

[4] Robert Koch-Institut.: 7-Tage-Inzidenz der COVID-19-Fälle in Deutschland. Available from: https://robert-koch-institut.github.io/COVID-19_7-Tage-Inzidenz_in_Deutschland/.

[5] Schiesser WE. The numerical method of lines: integration of partial differential equations. Elsevier; 2012.

[6] Süli E, Mayers DF. An introduction to numerical analysis. Cambridge university press; 2003.