CAMBRIDGE
UNIVERSITY PRESS

CORRIGENDUM

Statistical constraints on climate model parameters using a scalable cloud-based inference framework – CORRIGENDUM

James Carzon[1] (ID), Bruno Abreu[4,9], Leighton Regayre[5,6], Kenneth Carslaw[5], Lucia Deaconu[7,8], Philip Stier[7], Hamish Gordon[2,3] and Mikael Kuusela[1]

[1]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[3]Center for Atmospheric Particle Studies, Carnegie Mellon University, Pittsburgh, PA, USA
[4]National Center for Supercomputing Applications, University of Illinois Urbana-Champaign, Urbana-Champaign, IL, USA
[5]Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK
[6]Met Office Hadley Centre, Exeter, UK
[7]Atmospheric, Oceanic and Planetary Physics Department, University of Oxford, Oxford, UK
[8]Faculty of Environmental Science and Engineering, Babes-Bolyai University, Cluj, Romania
[9]Now at: Pittsburgh Supercomputing Center, Pittsburgh, PA, USA
Corresponding author: James Carzon; Email: jcarzon@andrew.cmu.edu

Our original study relied on a workflow codebase in which a bug was identified during follow-up work. This bug interfered with the geospatial matching of the emulated and observed aerosol optical depth measurements and led to some mistakes in the numerical results of our paper. Once we fixed the bug, the distribution of the standardized residuals between the emulated and observed measurements computed according to the model in Eq. (2) of the original manuscript was markedly non-Gaussian after outlier removal. Consequently, the model was found to be misspecified to the extent that all parameter configurations were deemed inconsistent with the observations by the test described in Section 2.4, necessitating an update to the model. Our updated model replaces the Gaussian noise term $\epsilon_{\text{other},x}$ after outlier removal with a more flexible noise model that no longer requires the outlier removal step. The implausibility test statistic remains the same although now we leverage a model-based bootstrap method to obtain its null distribution due to the non-Gaussian noise model.

As we write down the model equation once again,

$$
\begin{aligned}
z(x) &= \zeta(x) + \epsilon_{\text{meas},x} \\
&= \widehat{\eta}_x(u^*) + \epsilon_{\text{emu},x}(u^*) + \epsilon_{\text{meas},x} + \epsilon_{\text{other},x}, \\
\widehat{\eta}_x(u^*) &\coloneqq \mathbb{E}[\tilde{\eta}_x(u^*)|D_{\text{train}}],
\end{aligned}
$$

we relax our previous assumption that the term $\epsilon_{\text{other},x}$ is Gaussian. (It remains the case that $\epsilon_{\text{emu},x}(u^*) \sim N(0, \sigma^2_{\text{emu},x}(u^*))$ and $\epsilon_{\text{meas},x} \sim N(0, \sigma^2_{\text{meas},x})$ are well-founded assumptions by construction and previous literature, respectively, as was held in the original manuscript.) Instead, now we write.

$$
\begin{aligned}
z(x) &= \widehat{\eta}_x(u^*) + \sigma_x(u^*, \delta) d_x, \\
\sigma_x(u^*, \delta) &= \sqrt{\sigma^2_{\text{emu},x}(u^*) + \sigma^2_{\text{meas},x} + \delta^2},
\end{aligned}
$$

where $\delta > 0$ is an unknown parameter and $d_x \overset{\text{i.i.d.}}{\sim} p(0,1)$ for some unknown distribution p with mean 0 and variance 1.

Based on our updated data analysis, it appears that p is symmetric, motivating us to intermediately prescribe a Gaussian ansatz $\tilde{p} = N(0,1)$ toward calibrating a more flexible, yet still tractable implausibility test. With this ansatz, we jointly estimate

$$\widehat{(u,\delta)}_{\text{MLE}} = \underset{u \in \mathcal{U}, \delta > 0}{\operatorname{argmax}} \tilde{\mathcal{L}}\{u, \delta; z(x_1), z(x_2), \ldots, z(x_M)\}. \tag{1}$$

Denote the u- and δ-components of $\widehat{(u,\delta)}_{\text{MLE}}$ by $\widehat{u}_{\text{MLE}}$ and $\widehat{\delta}_{\text{MLE}}$, respectively. Here $\tilde{\mathcal{L}}$ denotes the ansatz-implied form of the likelihood, which is known in closed form and is efficiently optimized with the L-BFGS-B algorithm. Note that $\tilde{\mathcal{L}}$ is exactly the assumed likelihood from the original manuscript. With the maximum likelihood estimates as plug-ins, we use model-based bootstrap sampling to obtain the approximate empirical noise distribution

$$\widehat{p} = \text{Unif}\left\{ \frac{z(x) - \widehat{\eta}_x(\widehat{u}_{\text{MLE}})}{\sigma_x\left(\widehat{(u,\delta)}_{\text{MLE}}\right)} : x \in \mathcal{M} \right\}. \tag{2}$$

Using an intermediate fit to obtain an empirical estimate of the noise distribution is a standard strategy in bootstrapping regression models; see, for example, Davison and Hinkley (1997, Chapter 6). The empirical distribution $\widehat{p}$ enables us to account for heavy tails present in the observations; if we simply used the ansatz $\tilde{p}$, we would get unrealistically tight constraints due to model misspecification.

To test $H_0 : u = u_0$ versus the alternative that $H_1 : u \neq u_0$, we use the same implausibility metric as before as our test statistic,

$$I(u_0; z(x_1), z(x_2), \ldots, z(x_M)) = \sqrt{\sum_{x \in \mathcal{M}} \left(\frac{z(x) - \widehat{\eta}_x(u_0)}{\sigma_x\left(u_0, \widehat{\delta}_{\text{MLE}}\right)} \right)^2}. \tag{3}$$

Note that this test statistic is approximately pivotal for u_0 under the assumed model. Hence it suffices to calibrate the implausibility test for a single choice of u_0, in particular $\widehat{u}_{\text{MLE}}$ for this work. The critical value for the implausibility test is computed by Algorithm 1 leveraging the empirical noise distribution $\widehat{p}$ of Eq. (2). The code to reproduce the results from this corrigendum can be found on GitHub: https://github.com/JamesCarzon/smokeppe-constraints (last version: 21 October 2024).

Figure 1 (which replaces Figure 3 in the original manuscript) shows the resulting significant univariate parametric constraints by our corrected method. The number of test parameters has been increased from 5,000 to 50,000 to ensure a high-fidelity representation of these constraints. We find that the corrected pipeline provides meaningful univariate constraints on four variables: sea spray emission flux, accumulation mode aerosol dry deposition velocity, BVOC SOA, and the median diameter of primary ultrafine anthropogenic sulfate particles. Our data thus constrain the UKESM1 parameters more strongly than we previously reported. We note that our constraints on high values of sea spray emission flux and low values of primary sulfate diameter are consistent with the corresponding constraints displayed in Regayre et al. (2023); see their Figures S12–13. Our constraint on low values of BVOC SOA appears to be new. Our constraint on high values of dry deposition velocity is in contrast with their relative constraint on lower values. However, they do not entirely rule out any values, and thus there is no strict contradiction with our result. On balance, our four conclusive univariate constraints appear to be in greater agreement overall with Regayre et al. (2023) than those reported in our original manuscript.

Combinations of high accumulation mode dry deposition velocity and low BVOC SOA are jointly constrained (Figure 1e). This is in contrast with the original manuscript's constraint on low accumulation mode dry deposition velocity and high BVOC SOA. This updated result can be understood as physically meaning the following: If there is only a little aerosol mass formed by vegetation emissions while the deposition rate is high for relatively large particles, then one will underestimate AOD so much that even
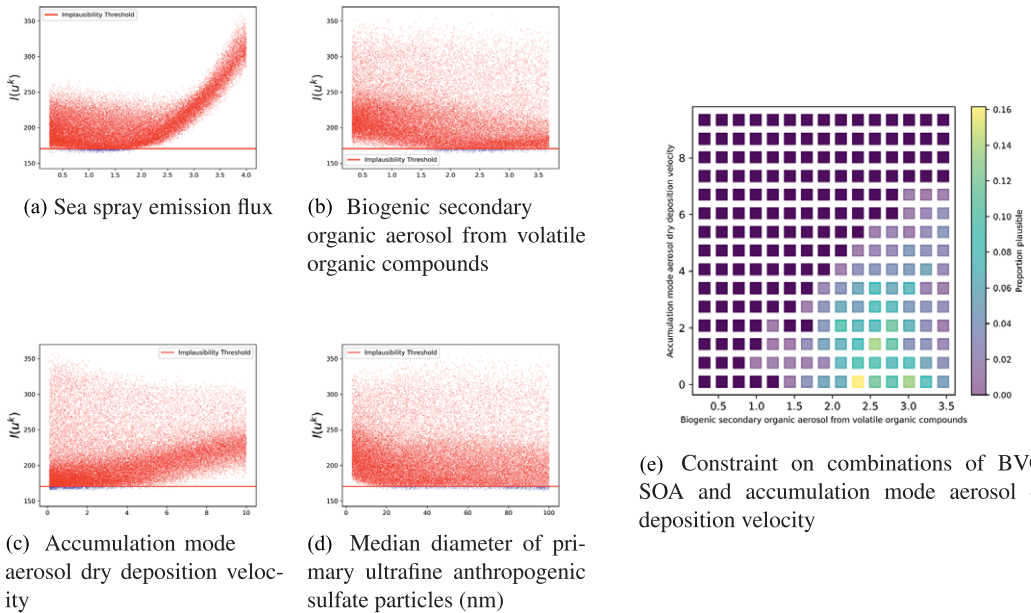
(a) Sea spray emission flux

(b) Biogenic secondary organic aerosol from volatile organic compounds

(c) Accumulation mode aerosol dry deposition velocity

(d) Median diameter of primary ultrafine anthropogenic sulfate particles (nm)

(e) Constraint on combinations of BVOC SOA and accumulation mode aerosol dry deposition velocity

Figure 1. *Parameter constraints at 95% confidence level. (a-d) One-dimensional projections of the* FREQUENTISTCONFSET *using the above revised* IMPLAUSIBILITY TEST. *The* 95*th percentile of the approximate null distribution under* H_0 *is indicated by the horizontal red lines. Each of the four displayed parameters are constrained on their own using only two weeks of data. (e) A two-dimensional projection of the* FREQUENTISTCONFSET *on the BVOC SOA and accumulation mode dry deposition velocity parameters, displayed as described in the manuscript. The dark purple pixels are ruled out as implausible.*

when controlling for the uncertainty of the MODIS retrieval estimates due to instrumental error, the imperfection of our surrogate model, and any other sources of model discrepancy that we estimate for our climate model, this scenario can be ruled out as implausible at the 95% confidence level.

Algorithm 1 `Estimated critical values for implausibility test`
Compute the critical value for the implausibility test at $u = u_0$ at significance level $\alpha \in (0, 1)$.

Input: Parameter value u_0; significance level $\alpha \in (0,1)$; residual distribution p^*, e.g., $\widehat{p}$ in Eq. (2); parameter estimate $\widehat{\delta}$, e.g., Eq. (1); implausibility metric I, e.g., Eq. (3)
Output: Estimated critical value of implausibility test at level α

1: Set $\mathcal{I} \leftarrow \emptyset$
2: **for** b in $\{1, 2, \ldots, B\}$ **do**
3:     Set $\mathcal{Z}_b \leftarrow []$
4:     **for** x in $\mathcal{M}$ **do**
5:         Draw $d^* \sim p^*$
6:         Simulate $z^*(x) \leftarrow \widehat{\eta}_x(u_0) + \sigma_x(u_0, \widehat{\delta}) \cdot d^*$
7:         Set $\mathcal{Z}_b \leftarrow \text{ADDITEM}(\mathcal{Z}_b, z^*(x))$
8:     **end for**
9:     Compute $I_b \leftarrow I(u_0; \mathcal{Z}_b)$
10:     Set $\mathcal{I} \leftarrow \mathcal{I} \bigcup \{I_b\}$
11: **end for**
12: Compute quantile $q \leftarrow Q_{1-\alpha}(\mathcal{I})$
13: **return** q

References

Carzon J, **Abreu B**, **Regayre L**, et al. (2023) Statistical constraints on climate model parameters using a scalable cloud-based inference framework, *Environmental Data Science* 2:e24. doi:10.1017/eds.2023.12.

Davison AC and Hinkley DV (1997) *Bootstrap Methods and Their Application*. Cambridge University Press.

Regayre LA, **Deaconu L**, **Grosvenor DP**, **Sexton DMH**, et al. (2023) Identifying climate model structural inconsistencies allows for tight constraint of aerosol radiative forcing, *Atmospheric Chemistry and Physics 23*(15), 8749–8768.