Scale-aware space-time stochastic parameterization of subgrid-scale velocity enhancement of sea surface fluxes

Julie Bessac, Hannah M. Christensen, Kota Endo, Adam H. Monahan, & Nils Weitzel

2021

Original citation:
Bessac, J., Christensen, H. M., Endo, K., Monahan, A. H., & Weitzel, N. (2021). Scale-aware space-time stochastic parameterization of subgrid-scale velocity enhancement of sea surface fluxes. *Journal of Advances in Modeling Earth Systems*, *13*(4). https://doi.org/10.1029/2020ms002367

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2020MS002367

Scale-Aware Space-Time Stochastic Parameterization of Subgrid-Scale Velocity Enhancement of Sea Surface Fluxes

Julie Bessac[1], Hannah M. Christensen[2], Kota Endo[3], Adam H. Monahan[3], and Nils Weitzel[4]

[1]Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL, USA, [2]Department of Physics, University of Oxford, Oxford, UK, [3]School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, Canada, [4]Institut für Umweltphysik, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany

Key Points:

- Subgrid-scale enhancement of air-sea fluxes conditioned on the resolved state shows resolution-dependent stochastic spatiotemporal structure
- We propose a statistical scale-aware model of these subgrid-scale tendencies
- The statistical model provides insight into transitions across resolution, spatial variability, and potential dependence on precipitation

Correspondence to:
J. Bessac,
jbessac@anl.gov

Citation:
Bessac, J., Christensen, H. M., Endo, K., Monahan, A. H., & Weitzel, N. (2021). Scale-aware space-time stochastic parameterization of subgrid-scale velocity enhancement of sea surface fluxes. *Journal of Advances in Modeling Earth Systems*, *13*, e2020MS002367. https://doi.org/10.1029/2020MS002367

Received 14 OCT 2020
Accepted 16 FEB 2021

Abstract Stochastic representation of the influence of the subgrid-scales on the resolved scales in weather and climate models has been shown to improve ensemble spread and resolved variability. We propose a statistical scale-aware space-time model to characterize the contribution of mesoscale wind variability to air-sea exchanges. In an earlier study, we analyzed the difference between "true" fluxes computed from a high resolution simulation and "resolved" fluxes obtained by coarse graining. This discrepancy is modeled in space and time, conditioned on the coarse-grained wind and precipitation fields, to parameterize the enhancement of fluxes by mesoscale velocity variations. Stochastic parameterization models have traditionally been developed for particular model resolutions without the explicit capability to adapt to model resolution. We present an approach to develop stochastic models that adapt to resolution in a scale-aware fashion. The scale-aware parameterization is developed from empirical results for systematically coarse-grained high-resolution numerical model output. The statistical model is fit from numerical model output at three different coarsening resolutions. From this scale-aware parameterization, we derive a stochastic parameterization of flux enhancement by subgrid velocity variations for arbitrary resolutions and characterize the conditional distributions and space-time structures of the flux enhancement across model resolutions.

Plain Language Summary Computer models of weather and climate rely on physics-based equations, and because of their discretization need to approximate the effects of fine scale phenomena. Different techniques have been used to develop such approximations through physically based or data-driven methods. In the following, we consider a statistically based approach and propose a statistical model of the enhancement of air-sea exchanges due to small-scale wind variability. In particular, we demonstrate that this enhancement is stochastic and exhibits mean and space-time structures that change with model resolution. Consequently, we build a statistical model that accounts for the resolution-dependence of the air-sea exchange enhancement. The proposed approach is shown to realistically represent the effect of small-scale winds and the change of this effect across resolutions.

1. Introduction

Numerical physically based models are used extensively to simulate phenomena in the Earth system and its components. In general, many physical phenomena happen at scales below the discretization scale of such models. These phenomena interact with the resolved scales. Quantifying and modeling the influence of subgrid-scale (SGS) processes on the resolved scales are needed in order to better represent the full system. In the absence of a scale separation between resolved and unresolved scales, the upscale influence of SGS processes is not a deterministic function of the resolved state. Stochastic models of SGS tendencies have shown great improvement over deterministic models in both uncertainty quantification and simulation of atmospheric variability (e.g. Berner et al., 2017; Dorrestijn et al., 2016; Gagne et al., 2020; Grooms et al., 2015; Palmer, 2019; Strommen et al., 2019). Such stochastic parameterizations, introduced to represent model uncertainty in ensemble forecasts, have been shown to improve forecast reliability (e.g. Leutbecher et al., 2017) as well as the simulated climate mean state and variability (e.g. Christensen et al., 2017).

SGS parameterizations should in general change with increasing resolution, as processes go from being unresolved, to partially resolved (i.e. "permitted"), to being fully resolved. As such, the problem of representing the upscale effect of subgrid variability illustrates the need for "scale-aware" models that adapt with changes in resolution (e.g. Fox-Kemper et al., 2019; National Academies of Sciences, Engineering, and Medicine, 2016). Recent interest has focused on scale-aware stochastic parameterization. The need for parameterizations to be both scale-aware and stochastic is particularly high in the absence of scale separations, such as the "gray zone" in which atmospheric convection or ocean eddies are only partially resolved.

A range of approaches to develop both deterministic and stochastic scale-aware parameterizations have been presented in previous studies. Some of these approaches are based on developing effective models for the reduced dynamics of multiscale systems. For example, Keane et al. (2014); Sakradzija et al. (2015), and Sakradzija et al. (2016) used approaches adapted from statistical physics to parameterize shallow convection in a way that adapts automatically to grid resolution. Visso and Lucarini (2018) used projection methods to develop reduced "slow" dynamics of the Lorenz '96 dynamical system, in which the predicted functional forms of the SGS tendency terms involve both stochastic and non-Markovian components with parameters determined by the effective model "resolution." While such approaches have the benefit of being physics-based, the approximate nature of the reduced systems can result in biased parameterizations that require further quasi-empirical tuning.

An alternative approach to developing stochastic parameterizations is through systematic coarse-graining of high-resolution simulations. For example, Zanna et al. (2017) represented oceanic mesoscale eddy fluxes in terms of an effective, resolution-dependent non-Newtonian viscosity with stochastic perturbations. Several studies have recently explored and modeled statistically the scale dependence of subgrid variability for various atmospheric phenomena and variables. For example, Huang et al. (2014) explored statistically the scale dependence of the first-order moments describing the cloud liquid-water path distribution. Time averaging and spatial coarsening were used interchangeably to study the subgrid variability of cloud liquid water across different seasons and spatial locations. In a subsequent study (Huang & Liu, 2014), the effects on biases in radiative transfer processes of subgrid variability of cloud liquid-water path and its scale dependence were investigated. In Xie and Zhang (2015), a Gamma distribution was used to represent the subgrid variability of cloud water in large-scale models. The shape parameter of the Gamma distribution representing the cloud water inhomogeneity explicitly depended on the model spatial resolution.

In this study, we specifically consider the scale dependence of the enhancement of air-sea fluxes by unresolved, mesoscale wind variations. In the absence of direct eddy covariance measurements, surface fluxes of mass, momentum, and energy between the atmosphere and ocean are calculated using bulk formulae. A generic form for the flux F_x of quantity X is

$$F_X = \overline{\rho}_a c_X(\overline{s})\overline{s}(\overline{X}_a - \overline{X}_o) \tag{1}$$

where ρ_a is the near-surface air density, s is the wind speed, $c(s)$ is a dimensionless exchange coefficient (often modeled as wind speed dependent) and the subscripts a and o, respectively, denote values of X measured in the atmosphere and ocean. In Equation 1 the bar denotes Reynolds averaging.

When such bulk formulae are applied using spatially or temporally averaged fields (such as in weather and climate models, in which resolved variables are grid-box averaged in space and Reynolds-averaged in time), variability on scales below the averaging scale leads to a biased estimate of the true average flux. While in principle this missing part of the flux involves contributions from SGS variations in all of X, ρ_a, and s (as well as their correlations), the dominant contributions come from SGS velocity variations (Blein et al., 2020). These SGS velocity variations result in a systematic enhancement of air-sea fluxes relative to estimates obtained from bulk formulae using the "resolved" wind (e.g. Bessac et al., 2019; Blein et al., 2020; Godfrey & Beljaars, 1991; Mahrt & Sun, 1995; Redelsperger et al., 2000; Vickers & Esbensen, 1998; Williams, 2001; Zeng et al., 2002). A standard approach accounts for the difference between the space-time grid-box averaged wind speed and the norm of the average horizontal wind vector due to SGS velocity variations through a SGS velocity flux enhancement term, s_{SGS}:

$$\overline{\langle s \rangle}^2 = \mid \overline{\langle \mathbf{u} \rangle} \mid^2 + s_{SGS}^2. \tag{2}$$

In Equation 2, $\boldsymbol{u} = (u, v)$ denotes the horizontal wind vector, with $s = \sqrt{u^2 + v^2}$, and angle brackets denote spatial averaging over a grid box. For standard resolutions of global models, SGS motions include contributions from both boundary layer turbulence and mesoscale variability. Studying the dependence on averaging scale of flux enhancements by SGS velocity variations, Mahrt and Sun (1995), Vickers and Esbensen (1998), and Zeng et al. (2002) found that the corrections become larger for coarser resolutions and proposed deterministic power-law expressions for the dependencies.

Focusing on the contribution of mesoscale SGS velocity variations, Bessac et al. (2019) modeled the enhancement of air-sea fluxes statistically by regressing "true" fluxes from a convection-permitting (4×4 km) simulation on coarse-grained "resolved" fluxes and precipitation and representing the residual as an explicitly stochastic space-time field. As boundary layer turbulence is parameterized in the simulations they considered, the subgrid flux enhancement is entirely due to spatial averaging of mesoscale motions. Bessac et al. (2019) demonstrated that the flux enhancement is non-deterministic and provided evidence of continuous changes across scale. Only the probability distributions of SGS flux enhancements at individual averaging scales were considered in this study. We shall refer to such distributions as "scale marginals," in contrast to joint distributions across multiple scales. The focus of the present study is the development of a data-driven scale-aware parameterization of air-sea fluxes appropriate for use in weather and climate models in terms of such joint distributions across scale.

In particular, this study builds on Bessac et al. (2019) by proposing a data-driven, scale-aware, and locally stationary space-time Gaussian process (GP) for air-sea fluxes driven by surface wind speed over the Indian and Western Pacific Oceans. The study relies on coarse-graining of convection-permitting model output, as is commonly used to assess subgrid variability when observational data are not available. The novelty of the current work consists in considering the regression and GPs simultaneously across a range of resolutions relevant to weather and climate models with a parametric representation of the scale dependence. Guided by multiple coarse grained resolutions, we propose a scale-aware parameterization of the regression and space-time GP correlation structure that enables expression of SGS flux enhancement at unobserved resolutions and provides further understanding of this dependence structure. To the best of our knowledge, this is the first data-driven, scale-aware parametrization that accounts for the space-time covariance structure of the random fluctuations. As in Bessac et al. (2019), we adopt an idealized representation of air-sea fluxes to focus on flux enhancement by SGS velocity variations. We do not consider other SGS contributions to the averaged flux.

The study is organized as follows. Section 2 presents the high-resolution numerical model outputs and the coarse-graining setup used to generate realizations at various resolutions as well as to define a ground "truth." In Section 3, we present the regression used in Bessac et al. (2019) and its extension to a scale-aware regression. We also propose a model for the regression residuals that embeds scale information in the spatiotemporal structure of the residuals. In Section 4, we describe the estimation procedure and discuss our interpretation. In Section 5, we discuss the realism of outputs from the proposed scale-aware space-time GP, showing in particular that the model can be fit on a reduced set of resolutions and that the difference between true and resolved fluxes seamlessly extrapolates or interpolates to unseen resolutions. A discussion and conclusions are presented in Section 6.

2. Data and Coarse-Graining Framework

Ideally, SGS wind variability statistics would be measured from observational data sets as in J. Sun et al. (1996). However, such data sets generally come from field campaigns of limited extent and duration, representing a small number of "snapshots" of the flow. Our analysis requires data over a large domain and of a sufficiently high spatial resolution for which suitable observational data sets are not available. Instead, we use as our "truth" an existing high-resolution, convection-permitting model simulation produced as part of the UK Natural Environment Research Council (NERC) Cascade project (Holloway et al., 2012; Love et al., 2011; Pearson et al., 2010). High-resolution model output is averaged over grids at various resolutions to represent relevant resolutions of state-of-the-art weather and climate numerical models (~10 to ~200 km) and serves as a proxy of varying model resolutions.

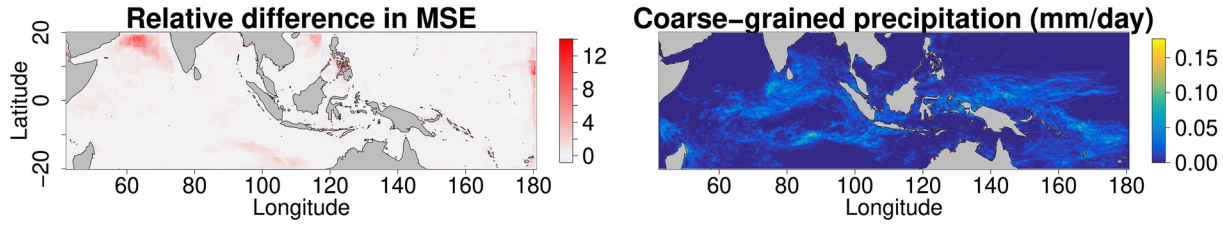

Figure 1. Left: Map of relative difference in mean squared error (MSE) $\left(\frac{MSE_{local} - MSE_{regional}}{MSE_{regional}}\right)$ in log scale of regression residuals at $N = 0.4°$. Right: Time-mean precipitation averaged to $N = 0.4°$.

We use the 4 km resolution Cascade simulation over the tropical Indo-Pacific Warm Pool. This Cascade simulation has proven useful for assessing stochastic parameterization schemes in other coarse-graining studies (e.g. Bessac et al., 2019; Christensen, 2020). A detailed description of the simulation is presented in Holloway et al. (2012). In summary, the simulation was produced by using the limited-area MetUM version 7.1 (Davies et al., 2005), covering the domain 20°S–20°N, 42°–177°E (Figure 1). The 4 km simulation formed one of a hierarchy of simulations. First, a 12 km parameterized convection simulation was produced over a domain 1° larger in each direction, with lateral boundary conditions relaxed to the ECMWF operational analysis. The lateral boundary conditions in the 4 km simulation were specified from the 12 km simulation, through a nudged rim of eight model grid points. The 4 km simulation is convection permitting in that the model resolution falls within the range of scales of mesoscale convective systems. The simulation begins on April 6, 2009 and spans 10 days, chosen as a case study of an active MJO event. The data are stored at full resolution in space and once an hour in time.

We discard the first day of simulation because Holloway et al. (2012) demonstrated a strong spin-up of the simulation over this period. Thorough assessment of the Cascade simulation has been reported by Holloway et al. (2012); Holloway et al. (2013, 2015). The simulation produces a realistic MJO, including realistic convective organization, MJO strength, and propagation speed (Holloway et al., 2013). The model also has a realistic representation of vertical and zonal wind speeds compared with ECMWF operational analysis, although regions of large-scale ascent are less confined than in observations (Holloway et al., 2013). We refer the reader to Bessac et al. (2019) for descriptive statistics of the data at the original and a range of coarse-grained resolutions.

It is important to note that with a resolution of 4 km, the CASCADE simulation does not resolve all SGS motions that would in principle contribute to flux enhancement. The dependence of the SGS flux enhancement on precipitation rate demonstrated in Bessac et al. (2019) indicates that the winds responsible for SGS flux enhancement in the coarse-grained simulation are largely associated with deep convective systems. Not all motions associated with such systems are resolved at 4 km resolution; this resolution is within the convection-permitting gray zone. Neither are motions associated with boundary layer turbulence resolved. As a result, our estimate of the magnitude of the SGS flux enhancement will be biased low. Smaller-scale features of the space-time statistical structure of the enhancement will also not be captured. The resolution of the simulation we consider is similar to that of other convection-permitting models used to assess SGS flux contributions (e.g. the 2.5 km resolution simulation in Blein et al., 2020). This limitation of our analysis is unavoidable because simulations capturing all relevant boundary layer and mesoscale variations cannot be conducted on a domain large enough to capture the organization of these motions on synoptic and planetary scales. Nevertheless, this limitation must be acknowledged.

3. Scale-Aware Space-Time Model for Air-Sea Fluxes Difference

3.1. Initial Single-Scale Model

In the following, we consider a simplified generic power-law representation of the wind speed dependence of air-sea fluxes (as is often found in bulk parameterizations, e.g. Fairall et al., 2003), expressed in nondimensional terms:

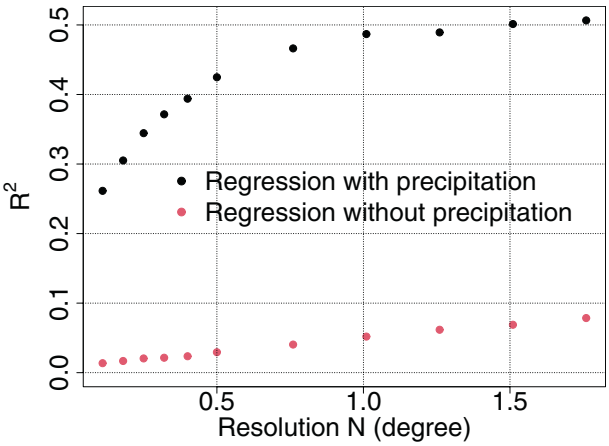

Figure 2. Coefficient $R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}}$ where $SS_{\text{reg}} = \sum_i (\hat{y}_i - \overline{y})^2$ and $SS_{\text{tot}} = \sum_i (y_i - \overline{y})^2$ for flux exponent $n = 2$. The R^2-coefficient represents the percentage of variance explained by the regression prediction $\hat{y}$ of the observed output variable y. The overall trend shows that at coarser resolutions the mean process explains more of the total variability of the true flux than at finer resolutions. R^2-coefficients have been additionally computed for the scale-aware regression proposed in Equation 7. It is not shown here because it is indistinguishable from the scale-specific regression with precipitation.

$$F_n(s) = \left(\frac{s}{s_0}\right)^n. \tag{3}$$

This idealized representation neglects contributions to air-sea fluxes other than those associated with wind speed variations; previous studies have found these neglected contributions to be relatively small (e.g. Blein et al., 2020). In Equation 3, $s = \sqrt{u^2 + v^2}$ is the surface wind speed; u and v are the zonal and meridional components, respectively; $s_0 = 1\ \text{ms}^{-1}$ is a speed scale to make the flux nondimensional. Neglecting the wind speed dependence of the exchange coefficient $c_X(s)$, the exponent n distinguishes between fluxes of different quantities: for example, $n = 1$ for energy and water vapor, and $n = 2$ for momentum. In fact, Blein et al. (2020) find SGS variations of the exchange coefficient to be the second-largest contribution to the difference between true and resolved heat, freshwater, and momentum fluxes, smaller only than those coming from the direct contribution of s in Equation 1. To some extent, variations in c_X can be accommodated in the value of n (e.g. in parameterizations of gas fluxes with $n = 2$, as in Wanninkhof, 2014). Because the wind-speed dependence of the exchange coefficient is different for different quantities, and a range of empirical forms of this dependence exist for any single quantity, we do not explicitly include this SGS contribution and use the idealized flux formulation Equation 3. As shorthand, we will refer to the idealized formulation with the given value of n as the "flux." In the main part study, we focus on the value $n = 2$. Results for $n = 1$ are presented in Appendix B.

Bessac et al. (2019) fit a regression model to $\varepsilon_{N,n}$, defined as the base-10 logarithm of the difference between the true flux averaged over a grid-box and the resolved flux over a grid-box

$$\varepsilon_{N,n} = \log_{10}\left(F^{(T)}_{N,n} - F^{(R)}_{N,n}\right) \tag{4}$$

where the true flux is $F^{(T)}_{N,n} = \left\langle \left(\frac{s}{s_0}\right)^n \right\rangle_N$, representing the grid-box average of pointwise fluxes and the resolved flux $F^{(R)}_{N,n} = \left(\frac{\sqrt{\langle u_N^2\rangle + \langle v_N^2\rangle}}{s_0}\right)^n$ is the flux computed from grid-box averaged (resolved) wind components. The subscript N denotes the grid-box size in degrees ($N \times N$) and $< . >$ denotes the spatial averaging. The log-10 error process $\varepsilon_{N,n}$ is real valued because $F^{(T)}_{N,n} \geq F^{(R)}_{N,n}$ by Jensen's inequality (resolved fluxes always underestimate true fluxes when only SGS velocity variations are accounted for). The previously proposed linear regression by which the SGS quantity $\varepsilon_{N,n}$ is predicted in terms of resolved quantities is given by:

$$\varepsilon_{N,n} = (A_{N,n})_0 + \sum_{k=1}^{3}(A_{N,n})_k \left[\log\left(F^{(R)}_{N,n}\right)\right]^k + \sum_{l=1}^{4}(B_{N,n})_l P^{l/4} + \psi_{N,n}, \tag{5}$$

where P is the coarse-grained precipitation rate. Bessac et al. (2019) showed that the resolved precipitation plays an important role in characterizing the space-time structure of the error $\varepsilon_{N,n}$. The substantial improvement of the R^2-coefficient for $n = 2$ is evident when using the resolved precipitation as a covariate in the regression fit separately at a range of resolutions (Figure 2). Such an improvement is also found for $n = 1$, although it is smaller. In contrast the resolved flux explains more variance for $n = 1$ compared to $n = 2$ (cf. Appendix B and Bessac et al., 2019). For a simpler generalization of our statistical model to different exponents n, we use the resolved flux and precipitation as predictors in this study since both of them explain substantial amounts of the error process variance for some exponents (Bessac et al., 2019).

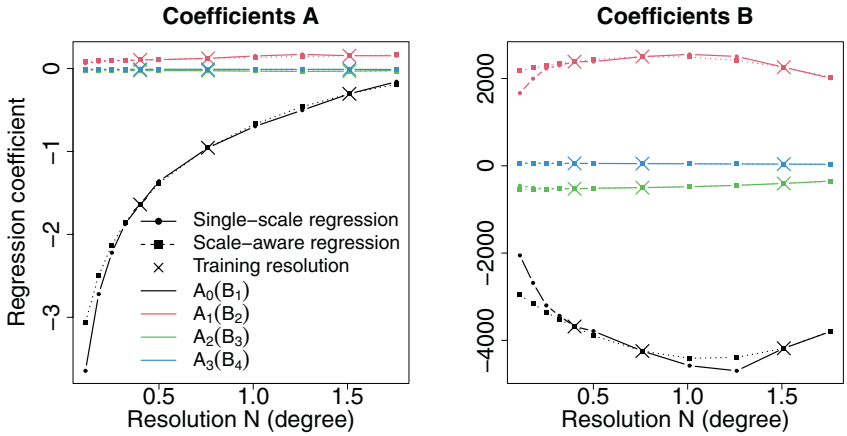

Figure 3. Regression parameters of model Equation 5 for flux exponent $n = 2$ when fit independently across various resolutions and when fit with the scale-aware parameterization from Equation 7 on the reduced set of resolutions $N =$ (0.40°, 0.76°, 1.51°) (marked with "×") and expressed for the whole range of studied resolutions through Equation 7.

In Bessac et al. (2019), the regression model (5) was fit independently for each coarse-grained resolution and the residual $\psi_{N,n}$ was shown to be well approximated by a GP with zero mean and a spatiotemporal correlation structure. This residual represents the part of $\varepsilon_{N,n}$ that can be parameterized stochastically, and was assumed to be statistically independent of the resolved flux and precipitation rate. The following space-time covariance model K was fit locally on the residuals $\psi_{N,2}$:

$$K(r,r',t,t') = \sigma \exp(-d(r,r',t,t')^{\gamma}) + \delta \mathbf{1}_{r,t}(r',t') \tag{6}$$

with the space-time metric $d(r,r',t,t') = \sqrt{\left(\frac{x-x'}{\theta_Z}\right)^2 + \left(\frac{y-y'}{\theta_M}\right)^2 + \left(\frac{t-t'}{\theta_T}\right)^2}$. The parameters θ_Z, θ_M, θ_T, σ, and δ are positive real numbers; and $\gamma \in]0,2]$; $r = (x, y)$ is the geographical position with longitude x and latitude y; and t is the time (in hours). The indicator function $\mathbf{1}()$ is such that $\mathbf{1}_A(x) = 1$ if $u \in A$ and $\mathbf{1}_A(x) = 0$ if $x \notin A$. The parameters θ_Z, θ_M, and θ_T, respectively, determine the correlation scales in the zonal, meridional, and temporal directions. The parameter γ controls the roughness of the stochastic fields, while σ determines their magnitude. The "nugget" parameter δ accounts for local fluctuations not captured by the assumed covariance function. All parameters θ_Z, θ_M, θ_T, γ, σ, and δ are estimated locally by using maximum likelihood. In Appendix A, we provide some details and references for Gaussian processes, covariances, maximum likelihood, and other statistical concepts used in this study.

3.2. From Scale-Specific to Scale-Aware Models

In the following, we explore the statistical characteristics of the regression Equation 5 for $n = 2$ and its residuals $\psi_{N,2}$ as functions of the grid-box size N. Specifically, the regression Equation 5 and the covariance model Equation 6 are fit independently on a succession of resolutions ($N =$ (0.11°, 0.18°, 0.25°, 0.32°, 0.40°, 0.50°, 0.76°, 1.01°, 1.26°, 1.51°, 1.76°)) in order to guide us in the construction of the scale-aware parameterization. Throughout the text, we will refer to the set of models that are fit independently at a given resolution as "scale-specific." Figure 2 illustrates the effects of the resolution on the explained variance, demonstrating that the regression explains more variance on coarser resolutions than finer resolutions. Physically, this behavior is a consequence of the fact that as resolution coarsens, the scale separation between resolved scales and parameterized mesoscales increases with a concomitant decrease in relative importance of the explicitly stochastic process $\psi_{N,2}$. Since regression coefficients and covariance parameters estimated at individual resolutions reveal a smooth dependence on the resolution parameter N (Figures 3 and 5), we propose a statistical formulation of the scale dependence that enables fitting at several resolutions simultaneously both a single scale-aware regression model (Section 3.1) and a scale-aware space-time Gaussian structure of

the residuals (Section 3.2). There exists no obvious physically motivated form to parameterize the scale dependence due to the complexity of the involved processes. Therefore, we choose an empirical form based on the single-scale model fits. In order to demonstrate the capability of the model to extrapolate and interpolate across resolutions, and to illustrate for application purposes that not the entire set of resolutions is needed, the scale-aware parameterizations are fit to a reduced set of resolutions ($N = 0.40°, 0.76°, 1.51°$). From this fit, the parameterization can be expressed for any resolution across the broad range studied.

3.3. Local Versus Regional Linear Regression

In Bessac et al. (2019) and the results presented in Figure 2, the parameters in regression Equation 5 were estimated for each resolution N individually by using all space-time points at that resolution at once. That is, the estimated regression model was across the entire model domain; we denote this model as being regional. Because of the relatively short (9-day) duration of the simulation, it is difficult to distinguish true spatial nonstationarity of the statistics from sampling variability with confidence. Nevertheless, before constructing a scale-aware model of $\varepsilon_{N,2}$, it is important to assess the representativeness of the regional regression. Therefore, we compare the regression Equation 5 fit regionally and locally. Local regressions are defined with the same set of predictants and predictors, taken from moving boxes of ~400 km over the area with overlaps of 40 km to ensure continuity. The local estimation is based on the moving window setup used in Bessac et al. (2019) to estimate a space-time GP on the residuals. Figure 1 shows the relative difference in mean squared error (MSE) of the regression evaluated for a coarse-grained resolution of 0.4° in a local and regional fashion. The overall magnitude and spatial structure of the MSE difference indicate that the local fit has few advantages over the regional fit. Most discrepancies are observed in areas with either low mean precipitation (Figure 1 right panel) over the 9-day duration of the simulation (e.g. in the Arabian Sea and Southwest Indian Ocean) or in coastal locations such as the Philippines archipelago. SGS flux enhancement in coastal areas involves contributions from processes such as offshore advection of land boundary layer features, land-sea breezes, and remote effects of orography that are not present in the open ocean (e.g. Blein et al., 2020; Dörenkämper et al., 2015). In view of the application of this analysis to parameterize fluxes in global weather and climate models, our focus is on open ocean fluxes. Accurate modeling of SGS flux enhancement in coastal areas requires a different approach, which we do not consider in this study.

In the following, since the difference between local and regional regressions is not substantial, we fit the regression in a regional fashion in order to reduce the estimation uncertainty associated with the small (9-day) sample size. This approach is also consistent with the fact that SGS parameterizations in weather and climate models are normally not explicitly dependent on spatial position.

3.4. Scale-Aware Regression Model

The independent fitting of regression Equation 5 at successive resolutions N reveals a smooth dependence of the parameters A and B on the resolution N (Figure 3). The main evolution across resolutions comes from the intercept A_0, which shows a decreasing absolute magnitude as resolution coarsens. To embed the scale-dependence exhibited by the parameters A and B and to provide a scale-aware regression model, we propose the following regression model to be fit regionally using least-squares estimation:

$$\begin{cases} \varepsilon_2(N) = (A_2(N))_0 + \sum_{k=1}^{3} (A_2(N))_k \left[\log_{10}\left(F_2^{(R)}(N)\right)\right]^k + \sum_{l=1}^{4} (B_2(N))_l P^{l/4} + \psi_2(N), \\ (A_2(N))_0 = \alpha_{0,0} + \alpha_{0,1}\ln(N) + \alpha_{0,2}N^2, \\ (A_2(N))_k = \alpha_{k,0} + \alpha_{k,1}N + \alpha_{k,2}N^2 \quad \text{for } k = 1,..3, \\ (B_2(N))_l = \beta_{l,0} + \beta_{l,1}N + \beta_{l,2}N^2 \quad \text{for } l = 1,..4. \end{cases} \tag{7}$$

In Equation 7, the notation for the log-10 error process $\varepsilon_2(N)$, the regression coefficients, and the stochastic residual $\psi_2(N)$ have been changed from Equation 5 to emphasize the continuous dependence on the resolution N (although parameter estimates are made over a discrete set of scales). In order to investigate the predictive capabilities of the scale-aware model, the parameterized forms of the regression coefficient dependence on resolution N are fit on a reduced set of resolutions ($N = 0.40°, 0.76°, 1.51°$). These parametric

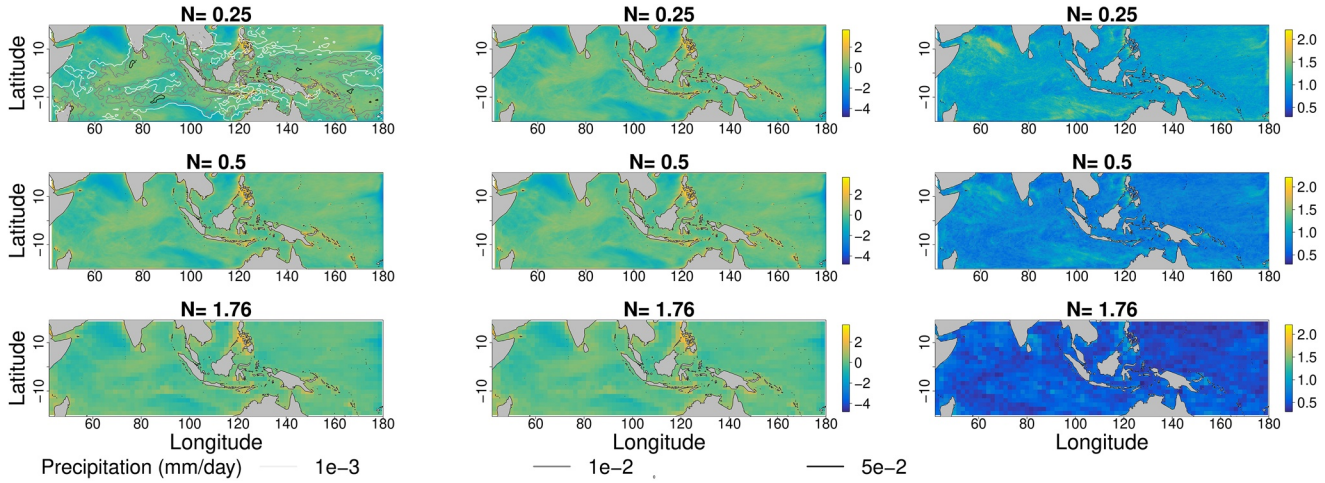

Figure 4. Mean (left and central columns) and standard deviation (right column) of residuals $\psi_2(N)$ at resolution $N = 0.25°$ (top row), $N = 0.5°$ (central row), $N = 1.76°$ (bottom row) when the scale-specific regression Equation 5 is fit at each individual resolution (left column) and when the scale-aware regression Equation 7 is fit (central and right columns). Contours of the mean precipitation (mm per day) over the study period are added in gray scale (top left corner).

fits capture well the scale dependence of the scale-specific coefficients (Figure 3). The R^2-coefficients of the scale-aware regression are not shown but are indistinguishable from the scale-specific values in Figure 2. A closer examination of the residuals is shown in Figure 4, where the mean and standard deviation of residuals from scale-specific regressions Equation 5 (left column) and scale-aware regressions Equation 7 (central column) are compared for different resolutions N, which are not used in the estimation. The intensity and spatial structure of the scale-aware regression residual mean are largely indistinguishable from the scale-specific regression residuals indicating that the scale-aware regression preserves the main structure of residuals at individual resolutions. The standard deviation of the scale-aware regression residuals (right column) is also essentially indistinguishable from that of the scale-specific regression residuals (not shown), indicating the ability of the parameterizations in Equation 7 to capture the dependence of the SGS flux enhancement on precipitation and resolved flux. As discussed earlier in the context of Figure 2, the residuals' variability decreases at coarser resolutions.

In the following, a space-time scale-aware GP structure is proposed for the residuals $\psi_2(N)$ of the scale-aware regression Equation 7.

3.5. Locally Stationary Scale-Aware Gaussian Process

Residuals $\psi_2(N)$ of regression Equation 7 have scale-dependent space-time structures as already observed in Bessac et al. (2019), where the covariance model Equation 6 was fit locally. The evolution across resolution of the spatiotemporal covariance parameters from Equation 6, estimated separately at individual resolutions N as in Bessac et al. (2019), is illustrated by the red boxplots in Figure 5 and shows a smooth evolution of the parameters with the resolution N. In particular, coarser resolutions that smooth out small-scale features display longer ranges in spatiotemporal dependence and more pronounced covariance anisotropy ($\theta_Z > \theta_M$). The variance decreases with coarsening resolutions as the "scale separation" between the mesoscale and the resolved scale increases, and the exponent parameter controlling the smoothness of the field increases slightly, which implies somewhat smoother fields.

We now provide a statistical parameterization of the space-time structure of the residuals as a scale-aware space-time GP. We assume that $\underline{\psi} = (\psi_2(N_0),\ldots,\psi_2(N_1))$ is a space-time process available at different spatial resolutions $N = N_0, \ldots, N_1$, such that each component $\psi_2(N)$ is the space-time residual process of the regression Equation 7 at resolution N. In this case, $N_0 = 0.11°$ and $N_1 = 1.76°$. Each $\psi_2(N)$ is observed at T time-points and K_N spatial points. The number of spatial points varies with resolution because the spatial domain is fixed. To account for a scale-aware parameterization of the space-time model, we jointly model

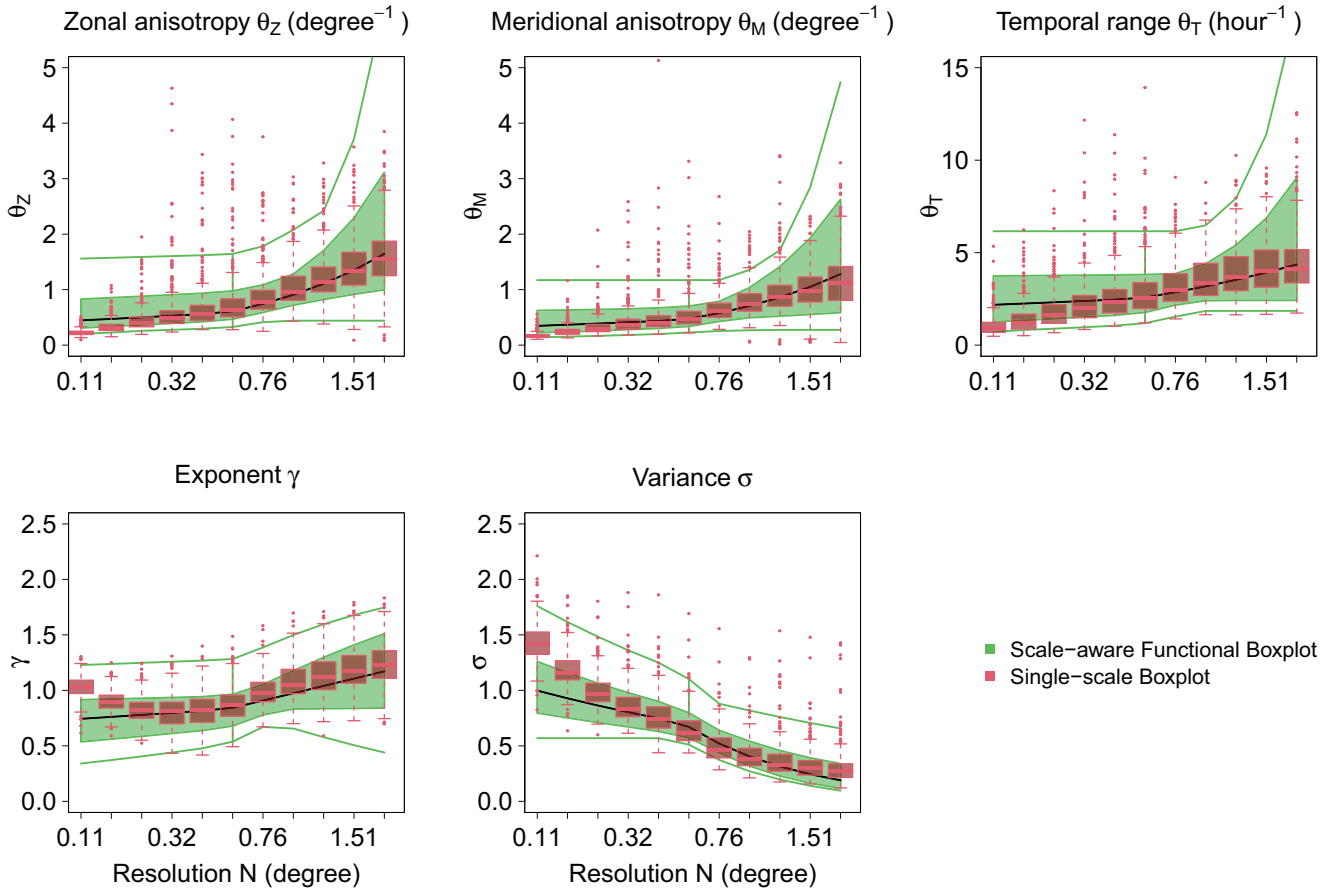

Figure 5. Functional (green) and classical (red) boxplots of respectively scale-aware parameterization of the GP covariance from Equation 9 and scale-specific estimated covariance parameters from Equation 6. The variability in the boxplots arises from parameters being estimated for moving windows over the Indian and Western Pacific Oceans. In red, the boxes are delimited by the 25th and 75th quantiles, the median is shown inside the boxes, and whiskers represent the interquartile range inflated by the factor 1.5. In green, the shaded area is delimited by the 25th and 75th curve quantile, the solid black line is the median curve and green lines show the interquartile envelope inflated by the factor 1.5. Note that for display purposes, the *x*-axis is not scaled linearly, creating visual distortions of some curves. The exponent and variance are nondimensional quantities, where the nondimensionality of the variance is due to the formulation of the flux defined in Equation 3 as nondimensional. GP, Gaussian process.

the process $\underline{\psi} = \left(\psi_2(N_0),\ldots,\psi_2(N_1)\right)$ assuming that $\underline{\psi}$ is a space-time zero-mean GP. Since weather and climate models require stochastic enhancements at a given resolution (that is, they make use only of the "scale marginals"), we assume the cross-dependence between resolutions to be null. While dependence across resolutions exists and will affect the estimation procedure, it does not need to be accounted for in the application. For simplicity and model parsimony, we neglect this dependence. The covariance of $\underline{\psi}$ is modeled by a block-diagonal covariance, where each block is a space-time covariance associated with a resolution N and a parameterization depending on N:

$$Cov(\underline{\psi}) = \begin{bmatrix} Cov(\psi_2(N_0)) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Cov(\psi_2(N_1)) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}(\theta, N_0) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{\Sigma}(\theta, N_1) \end{bmatrix}$$

with the covariance $\boldsymbol{\Sigma}(\theta, N)$ following a three-dimensional powered exponential similar to that used in Equation 6:

$$\begin{aligned}\boldsymbol{\Sigma}(\theta,N) &= \sigma(N)\exp\left(-d_N(r,r',t,t')^{\gamma(N)}\right)\\ d_N(r,r',t,t') &= \sqrt{\frac{(x-x')^2}{\theta_Z(N)}+\frac{(y-y')^2}{\theta_M(N)}+\frac{(t-t')^2}{\theta_T(N)}}\end{aligned} \tag{8}$$

where $r = (x, y)$ is the geographical position with longitude x and latitude y; t is time (hour); and θ is the set of covariance parameters defined below. The nugget parameter δ used by Bessac et al. (2019) is omitted for simplicity since its value was found to be very small in fits to individual resolutions. Although we neglect the cross-correlation among resolution, we model the scale dependence between the scale marginals by parameterizing the parameters θ_Z, θ_M, θ_T, σ, and γ as functions of the resolution N using functional forms suggested by the single-scale fits (Figure 5):

$$\begin{cases}\theta_Z(N) = \theta_{Z,1}\exp(\theta_{Z,2}N) \quad \text{with} \quad \theta_{Z,1} > 0\\ \theta_M(N) = \theta_{M,1}\exp(\theta_{M,2}N) \quad \text{with} \quad \theta_{M,1} > 0\\ \theta_T(N) = \theta_{T,1}\exp(\theta_{T,2}N) \quad \text{with} \quad \theta_{T,1} > 0\\ \sigma(N) = \sigma_1\exp(\sigma_2 N) \quad \text{with} \quad \sigma_1 > 0\\ \gamma(N) = 1+\tanh(\gamma_1+\gamma_2 N) \in \left]0,2\right].\end{cases} \tag{9}$$

The hyperbolic tangent parameterization of $\gamma(N)$ ensures that this parameter falls in the required range]0, 2]. The parameter θ represents the collection of coefficients $\theta = (\theta_{Z,1}, \theta_{Z,2}, \theta_{M,1}, \theta_{M,2}, \theta_{T,1}, \theta_{T,2}, \sigma_1, \sigma_2, \gamma_1, \gamma_2)$. The scale-aware space-time GP model is fit locally by Maximum Likelihood on a subset of residuals selected at three scales representative of the entire span of scales by estimating the optimal parameter θ. The estimation procedure is described in the following section.

4. Covariance Parameter Estimation and Interpretation

In this section, we describe the estimation, based on local estimations and using Maximum Likelihood (Section 4.1). The uncertainty of these estimates is discussed in Section 4.2. Section 4.3 illustrates the spatial distribution of the locally estimated parameters.

Parameters of the scale-aware GP introduced in Section 3.2 are estimated through optimization of the following log-likelihood:

$$\log L(\theta;\psi_2(N_0),\ldots,\psi_2(N_1)) = \sum_{N=N_0}^{N_1}\log\left(\frac{\exp\left(-\frac{1}{2}\psi_2(N)^\top\boldsymbol{\Sigma}(\theta,N)^{-1}\psi_2(N)\right)}{(2\pi)^{K_N/2}\left|\boldsymbol{\Sigma}(\theta,N)\right|^{1/2}}\right). \tag{10}$$

The summation over the different model resolutions results from the block-diagonal covariance matrix enabling to write the joint likelihood as a product of the terms specific to each resolution. Since no analytical solution is available for the maximizer of Equation 10, numerical optimization is used to find the parameters that maximize the logarithm of the likelihood function. The numerical optimization of the log-likelihood Equation 10 is performed in *R* (R Core Team, 2020) according to the following procedure. Details of the likelihood expression can be found in Appendix A.

4.1. Local Maximum Likelihood Estimations

To investigate potential spatial heterogeneity of the space-time structure of $\psi_2(N)$ across the domain, and because of the large amount of data, we make use of local fitting in a moving-window as in Bessac et al. (2019). More specifically, the whole domain is subdivided into smaller regions of size 400 × 400 km. Within each window, stationarity is assumed, and the proposed GP model from Section 3.2 is fit independently to the set of representative resolutions ($\psi_2(0.40)$, $\psi_2(0.76)$, $\psi_2(1.51)$) used in the multi-scale regression model

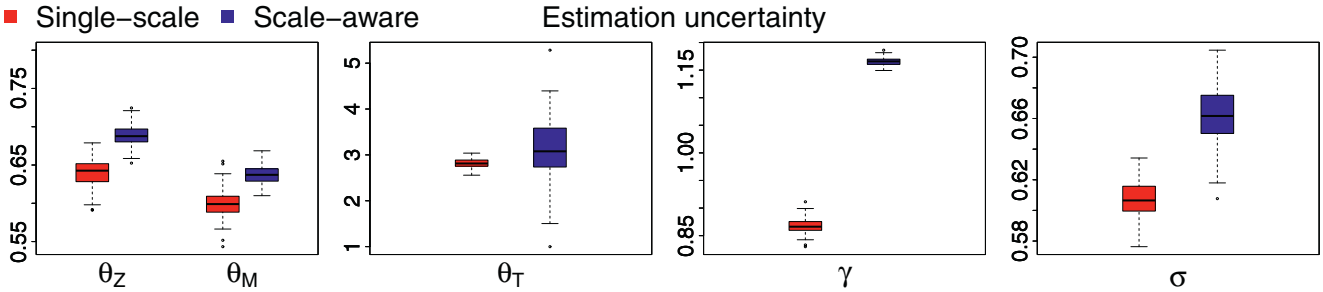

Figure 6. Estimation uncertainty of the maximum-likelihood estimates of the parameters expressed at resolution $N = 0.76°$ for $\psi_2(n)$. Results of random samples from the asymptotic distribution $\mathcal{N}(\hat{\theta}, \mathcal{I}^{-1}(\hat{\theta}))$ of the maximum-likelihood estimates $\hat{\theta}$ are shown for a spatial window realizing the median MSE within the domain. MSE, mean squared error.

(Section 3.1). In order to ensure continuity, the windows overlap by 40 km. Maximum Likelihood estimation (MLE) is used within each window to maximize Equation 10.

Figure 5 displays in green the resulting fits of the functional parameterizations from Equation 9 via functional boxplots (Hyndman & Shang, 2010; Sun & Genton, 2011), where curves are ordered according to the band-depth measure of centrality for functional data. This measure enables an ordering of curves that corresponds to the visual centrality (also called depth in this context) of a curve with respect to a group of curves and provides an extension of point ordering to curve ordering. The ranking of the curves provides a concept analogous to that of quantiles in point-wise statistics. The R-package fda is used to create the functional boxplots (Ramsay et al., 2020). The spread of the estimates in Figure 5 results from estimates in different moving windows of the domain. The estimated scale-aware covariance parameters (green) show good agreement with scale-specific estimates (red) of the covariance parameters: the medians match well, although the 50%-envelope (green shaded area) is wider for the scale-aware parameters than for scale-specific parameters, which indicates more variability across the domain, especially for the spatiotemporal range parameters. We note that functional boxplots based on band-depth ordering, as performed here, are dependent on the curves' shapes, leading to potentially wider envelopes compared to classical point-wise boxplots (Y. Sun & Genton, 2011).

4.2. Estimation Uncertainty

The MLE procedure provides estimates and their estimation variance for the coefficients $\theta_{Z,1}$, $\theta_{Z,2}$, ..., σ_1, σ_2. The estimation variance $\mathcal{I}^{-1}(\hat{\theta})$ associated with the Maximum Likelihood is extracted from the inverse Hessian of the log-likelihood value at the optimum point. The multivariate Delta-method is used to approximate the variance of the transformed parameters (θ_Z, θ_M, θ_M, γ, σ) in Equation 9 from the estimation variance $\mathcal{I}^{-1}(\hat{\theta})$ of the coefficients $\theta_{Z,1}$, $\theta_{Z,2}$, ..., σ_1, σ_2. We provide details of these computations in Appendix A. In order to illustrate the estimation uncertainty, random samples from the normal distribution (which under regularity conditions is the parameters' asymptotic distribution) $\mathcal{N}(\hat{\theta}, \mathcal{I}^{-1}(\hat{\theta}))$ centered on the estimated parameters $\hat{\theta}$ and with estimation variance $\mathcal{I}^{-1}(\hat{\theta})$ as a dispersion parameter are displayed in Figure 6. Results are shown for the particular spatial window that contains the median MSE. Estimates and their variances are shown for the scale-aware model and the model fit on a single resolution without the parameterization Equation 9. For the considered spatial window, the temporal range parameter shows a larger estimation variance for the scale-aware model, as has been observed at other spatial windows (not shown), meaning that the confidence in this parameter estimation is lesser than for the other parameters. This might be due to the short temporal record of data. In contrast, the estimation variances of the zonal and meridional range parameters are smaller for the scale-aware model. Additionally, we note that over the entire domain the MLE estimation procedure is less sensitive to initial conditions and less subject to divergence of the optimization method for the scale-aware model than for scale-specific models, which we attribute to the increased amount of data used in the MLE.

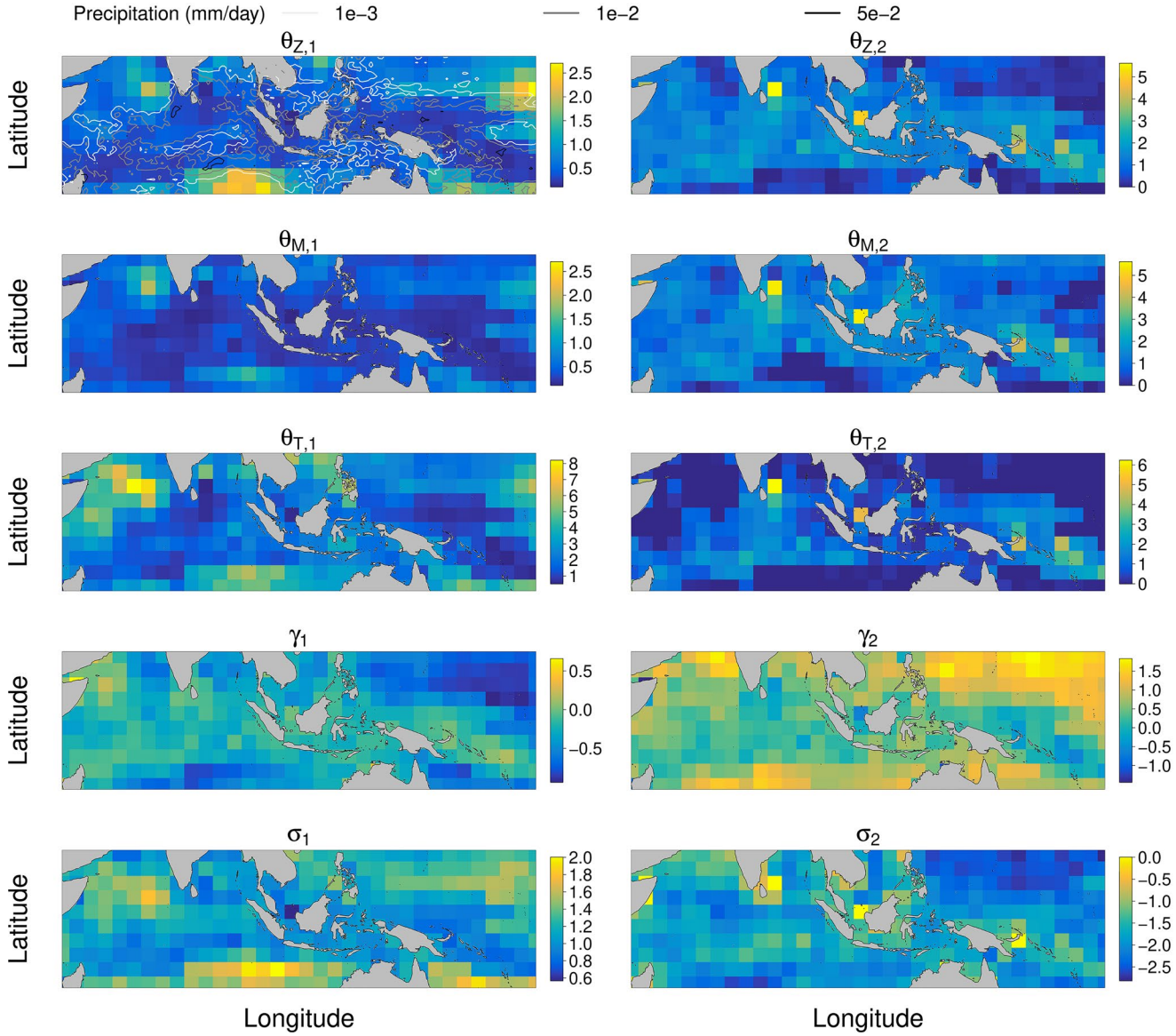

Figure 7. Spatial structure of estimated parameters from Equation 9. Each pixel represents an estimate from a single spatial moving window. Contours of the mean precipitation (mm per day) over the study period are added in gray scale (top left corner).

4.3. Spatial Structure of Estimated Parameters

The spatial structure of the parameters from Equation 9, estimated in a moving window fashion, is displayed in Figure 7. The spatial variability of most parameters appears to be closely related to the distribution of mean precipitation in Figure 1. The $\theta_{.,2}$ values for the zonal, meridional, and temporal ranges are small where precipitation is almost absent, suggesting a weak scale dependence in the absence of precipitation in the area. The γ_2 parameter distribution reveals that the $\psi_2(N)$ fields become smoother with larger N where precipitation is low. In contrast, the field is rougher with weaker scale dependence in areas with more abundant precipitation. Moreover, as observed and discussed earlier, the σ_2 parameter representing the magnitude of the SGS contributions decays more rapidly with increasing N in the low-precipitation areas (Arabian Sea, Southeastern Indian Ocean, and Western Pacific Ocean).

These results suggest that the flux error may have a dependence on precipitation beyond the first-order modeled by the regression. However, some of the spatial variability may be due to the limited duration of

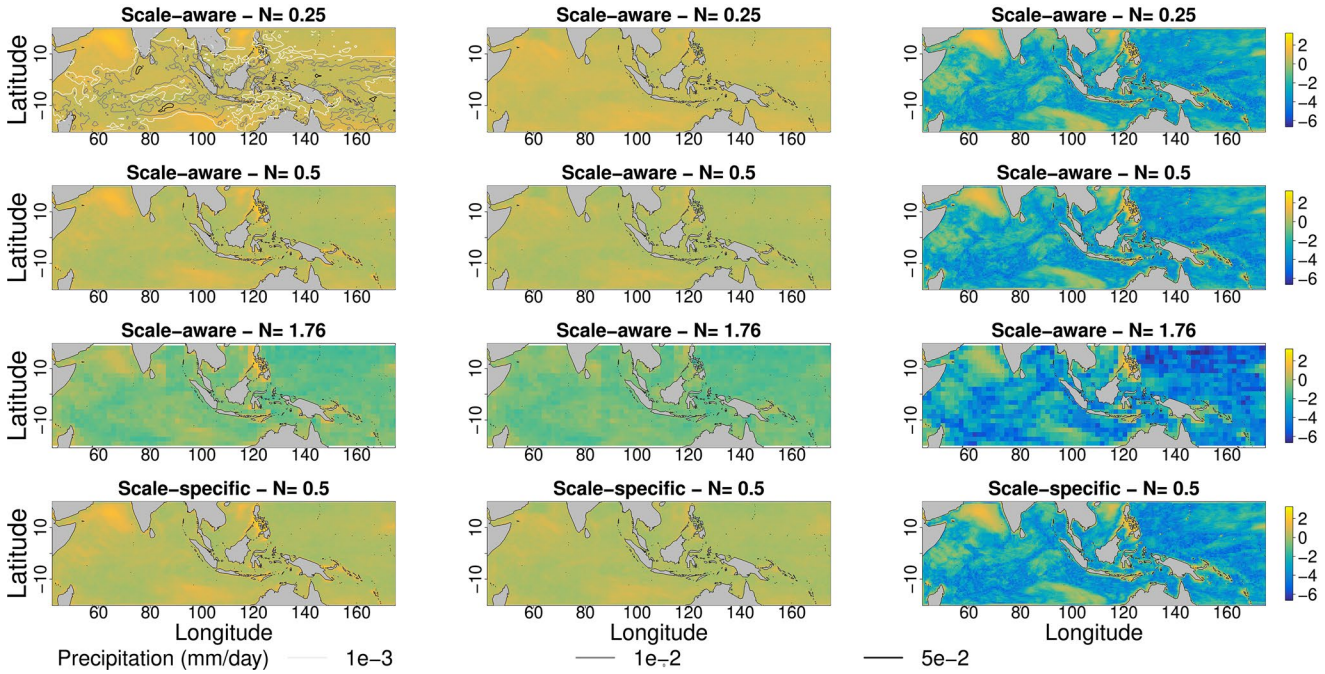

Figure 8. Maps of regional MSE (left column) and its decomposition into centered MSE (central column) and squared bias (right column) for the different extrapolated and interpolated resolutions with the scale-aware model $N = 0.25°$ (top row), $N = 0.5°$ (second row), and $N = 1.76°$ (third row) and for the scale-specific model at $N = 0.5°$ (bottom row). MSE and its components are computed between the data flux error $\varepsilon_2(N)$ from Equations 4 and 30 samples from Equation 7. For ease of readability, results show log(MSE), log(squared bias), and log(centered MSE). Contours of the mean precipitation (mm per day) over the study period are added in gray scale (top left corner). MSE, mean squared error.

the data set: the long-term mean precipitation field in the domain does not contain as much small-scale structure as this 9-day sample. Additionally, some numerical compensation between estimates of the parameter pairs in each of the expressions in Equation 9 might be present and potentially create "spatial artifacts." Determination of whether these apparent structures represent true spatial heterogeneity or result from sampling variability or numerical compensation requires consideration of alternative, longer datasets and is a subject of future research.

5. Evaluation of the Model Performance

The scale-aware model from Section 3 is fit to the following resolutions simultaneously: $N = (0.40°, 0.76°, 1.51°)$ as described in Section 4. In order to assess how well the scale-aware model generalizes to other resolutions, parameters at resolutions not used for the estimation, $N = (0.25°, 0.5°, 1.76°)$, are determined from the model. These parameters are then used to generate realizations of $\varepsilon_2(N)$ by computing the mean from Equation 7 and sampling from the GP $\psi_2(N)$ in Equation 9. Samples from the GP $\psi_2(N)$ in Equation 9 are independent from one another and identically distributed according a multidimensional Gaussian distribution with mean zero and covariance given by Equations 8 and 9 prescribing their space-time covariance structure. These samples of $\varepsilon_2(N)$ can be seen as equally probable realizations of the derived stochastic parameterization. Each sample corresponds to the realization of the flux in an ensemble member of a weather and climate model. Thus, the samples characterize the uncertainty in the flux due to SGS wind speed variability as determined by the GP model. The samples rely on the coarse-grained data for the resolutions, which are used for fitting the scale-aware model, solely through the estimated parameter values. Potential correlations between different scales are not exploited for sample generation to focus on the "scale marginals" as described above. Evaluations of these simulated temporal realizations compared with ones sampled from a scale-specific model are presented across the entire region in Section 5.1 and pointwise in Section 5.2.

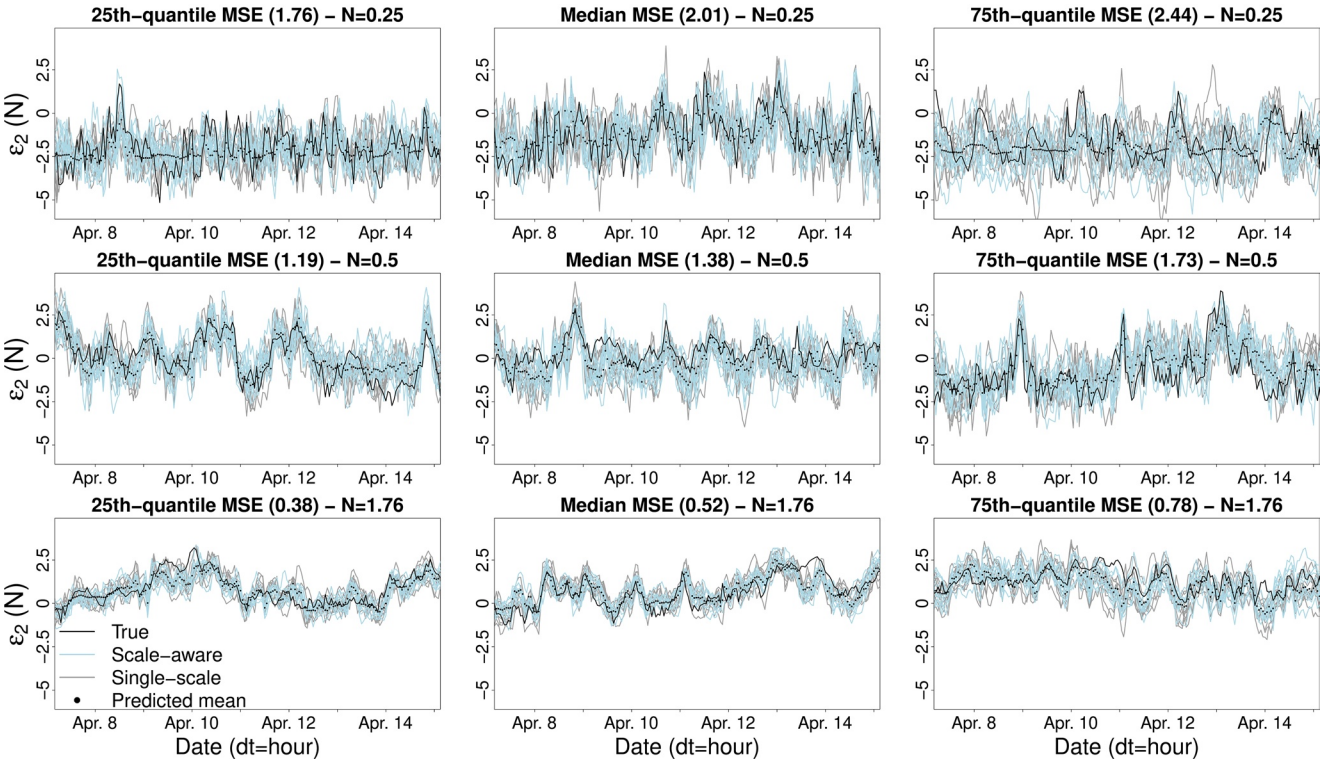

Figure 9. Time series of original and reconstructed flux errors $\varepsilon_2(N)$ for the three resolutions $N = 0.25°$ (top row), $N = 0.5°$ (central row), $N = 1.76°$ (bottom row) at locations corresponding respectively to the first quartile (left column), median quartile (central column), and third quartile (right column) of the MSE over the entire domain. Ten trajectories simulated with the scale-aware model are in blue, 10 trajectories simulated with a scale-specific model are in gray, and the observed flux error is in solid black. MSE, mean squared error.

5.1. Regional Assessment of Scale-Aware Model

In order to quantify the simulation fidelity of the scale-specific and scale-aware models, the MSE of $\varepsilon_2(N)$ (between "true" and sampled fields) and the decomposition of MSE into centered MSE and squared bias are computed from 30 samples of $\varepsilon_2(N)$ (Figure 8). The overall MSE is smaller at coarser resolutions; as discussed above, the need for stochastic corrections to the flux parameterization is less at larger resolutions. The scale-aware model shows results similar to those of the scale-specific model (illustrated for $N = 0.5°$) demonstrating the ability of the scale-aware model to characterize variability on resolutions not used in the estimation. The average relative differences of MSE between the scale-aware model and the scale-specific model are respectively − 4.06%, 4.21%, and − 12.65% for the predicted resolutions $N = 0.25°$, $N = 0.5°$, and $N = 1.76°$, indicating the presence of only modest differences between the two models. The squared bias, which shows the contribution of the regression model, exhibits spatial structures apparently correlated with mean precipitation (Figure 1) such that relatively high biases are observed in low precipitation areas. The centered MSE indicates how the variability of the field is captured by the GP model. It shows weaker spatial structure than the full MSE does, indicating that the model performs well at capturing field variability of $\varepsilon_2(N)$. The similarity of scale-aware and scale-specific centered MSE as well as squared bias shows that both, scale-aware regression and GP model, perform similarly to the respective scale-specific models.

5.2. Pointwise Assessment of Scale-Aware Model

In this section, regional variations of the log-10 flux error process $\varepsilon_2(N)$ are investigated in more detail. Figure 9 provides a visual characterization of the reconstructed $\varepsilon_2(N)$ time series at interpolated and extrapolated resolutions $N = (0.25°, 0.5°, 1.76°)$, for both the original CASCADE simulation and for realizations drawn from the scale-aware and scale-specific models. Time series are shown at different locations representative

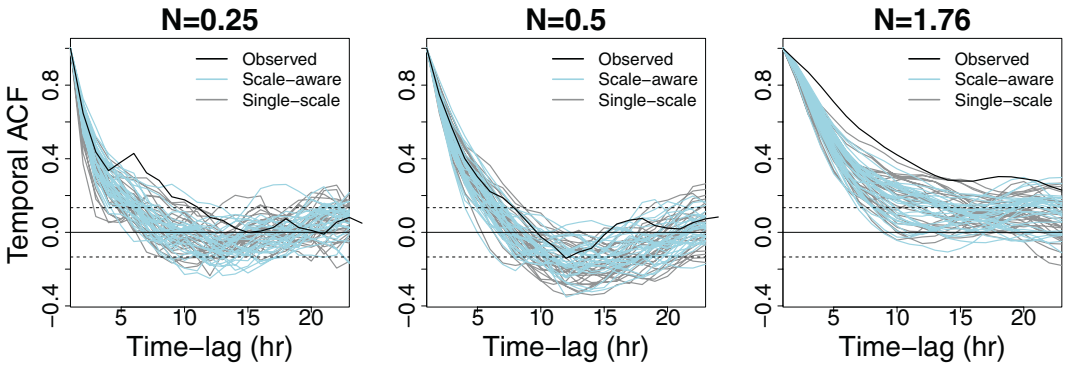

Figure 10. Temporal autocorrelation of observed and simulated flux errors $\varepsilon_2(N)$ for the three resolutions $N = 0.25°$ (left), $N = 0.5°$ (central), $N = 1.76°$ (right) at the location realizing the median MSE over the entire domain. Autocorrelation of 30 trajectories simulated with the scale-aware interpolated model are in blue, trajectories simulated with a scale-specific model are in gray, and the observed flux error is in black. The horizontal dashed lines show the 95% correlation coefficient confidence interval of a white noise and within this interval the autocorrelation can be considered statistically insignificant. MSE, mean squared error.

of the three lower quartiles of the MSE across the whole domain; higher quantiles of MSE are not shown because they are realized in coastal areas that are strongly influenced by the nearby land. Figure 9 shows that the extrapolated and interpolated scale-aware model performs similarly to a scale-specific model in terms of intensity, spread and temporal dynamics, and captures most features of the observed time series. Again, the smaller contribution to the variance of the stochastic residual $\psi_2(N)$ at coarser resolutions is evident. These results are corroborated by the temporal auto-correlation functions of the residual process $\psi_2(N)$ in Figure 10, estimated for the three different resolutions at the point realizing the median MSE over the domain. Both statistical models show generally good agreement with the CASCADE temporal autocorrelation, except at the coarsest resolution. Note again that the resolutions considered in Figures 8–11 are not those used to estimate the parameters of the scale-aware model.

To assess the spatial dependence structure of the $\psi_2(N)$ field, we show in Figure 11 contour lines of the correlation of each moving window central point with points across the rest of the domain. For each central point, a single correlation contour is shown (chosen at correlation strength $z = 0.7$). Contours computed on the original and sampled $\varepsilon_2(N)$ from scale-aware and scale-specific models are compared. The original correlation shows areas with zonally elongated contours, particularly in the Arabian Sea and Southern Indian Ocean. Both sets of realizations show zonally elongated contours and some of the spatial heterogeneity; however, they tend to underestimate this anisotropy (particularly in the Arabian Sea). The simulations reproduce well the observed correlation structure in the Western Pacific (both its anisotropy and its spatial variations). Despite the moving window setup of the GP estimation, however, simulations tend to capture less of the spatial heterogeneity in the northwestern part of the domain. Again, it is not clear to what extent the apparent spatial inhomogeneity is a consequence of sampling variability in this relatively short-duration data set.

These statistics reveal that the proposed scale-aware space-time model provides realistic realizations of flux errors in terms of their intensity and space-time structure. The validation analysis further indicates that the spatial heterogeneity of many of the field aspects such as signal intensity, space-time correlation ranges, anisotropy, and smoothness are captured by the proposed model and its local estimation framework. The scale-aware model performs similar to scale-specific models indicating the accuracy and strong predictive capabilities of the proposed scale dependence in the parameterization.

6. Discussion and Conclusions

In this study, we have built a scale-aware representation of a previously proposed parameterization of air-sea flux enhancement by SGS wind variability for use in climate and weather models (Bessac et al., 2019). The existing model expresses the logarithmic error between a "true" flux and a "resolved" flux, conditioned

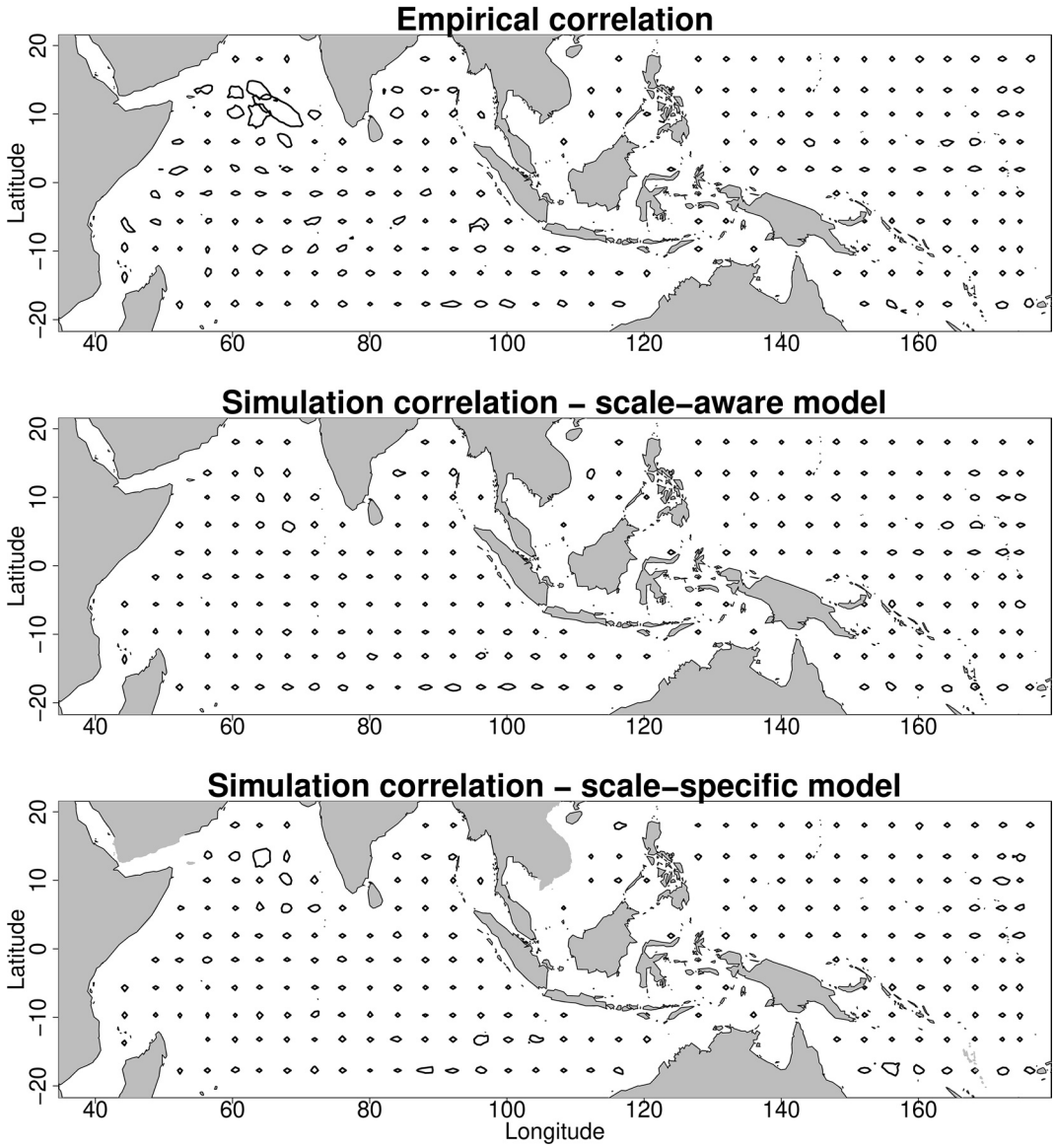

Figure 11. Contours of spatial correlation of the flux error $\varepsilon_2(N)$ at each moving window center with the rest of the field, for the single contour level $z = 0.7$. Correlations are computed on log-10 flux errors $\varepsilon_2(N)$ from Equation 4 (top), simulations of the flux errors from the scale-aware interpolated model (center), and simulations from the scale-specific model (bottom) at resolution $N = 0.5°$.

on the resolved flux and the total precipitation, while accounting for the stochasticity and spatiotemporal structure via a local GP. The "truth" is defined by the output of a high-resolution atmospheric simulation, and "resolved" quantities are coarse-grained versions of the truth, mimicking a model at lower resolution. Fluxes are represented in an idealized fashion, focusing on the contribution of wind speed variations and neglecting other factors. The proposed scale-aware parameterization of both regression and spatiotemporal structures provides a flexible and adaptive framework to model SGS variability in the context of models running at ever-increasing resolution—particularly those operating in the "gray zone" where important physical processes (e.g., convection) are "permitted" but not fully resolved. Scale awareness has been represented by empirical functions which forms are guided by the data-driven combination of CASCADE data and fits of the scale-specific model across a broad range of resolutions. While our focus has been on a wind speed exponent characteristic of momentum and gas fluxes, similar results are obtained for the exponent relevant to heat and freshwater fluxes (Appendix B).

The scale-aware model is fit simultaneously on three resolutions in order to investigate the capabilities of the model to extrapolate and interpolate at unseen resolutions. In assessments performed at extrapolated and interpolated resolutions, the scale-aware model performs similarly to scale-specific models in terms of fidelity of generated outputs, indicating that in this particular setup and application, one can use a reduced set of resolutions and obtain an accurate model at untrained resolutions. Additionally, the MLE estimation procedure is numerically more stable (sensitivity to initial conditions and optimization non-degeneracy) for the scale-aware space-time GPs than for scale-specific GPs, which we impute to the increased amount of data used in fitting the scale-aware model. As discussed in the Introduction, in this case study the dependence across resolutions is smooth and could be represented by simple mathematical functions. However, other systems or phenomena might exhibit complex evolution throughout resolutions such as changing tails or probability distributions, challenging their mathematical representations. The evaluation of whole-domain and pointwise structures reveals that spatiotemporal structures are generally well captured except in areas with very low precipitation during the 9-day study period or with complex features close to the coasts. Further work will focus on assessing the representativeness of the 9-day period by considering outputs from longer atmospheric model simulations in order to improve the characterization of spatial variability, which we expect to be smoother (especially in terms of precipitation). There is also the need to consider the structure of the SGS flux enhancement in spatial domains other than the Warm Pool/Northern Indian Ocean, and to repeat our analysis using other convection-permitting models to assess the robustness of our statistical model. In future work, we will conduct such analysis using a suite of global, convection-permitting simulations generated by the Dynamics of the Atmospheric General Circulation Modeled on Non-Hydrostatic Domains initiative (e.g. Stevens et al., 2019). This work will also investigate the development of a unified parameterization across values of the flux exponent n, so that consistent realizations of SGS enhancements of different fluxes can be generated simultaneously. Such an extension could additionally be applied to non-integer values of n as they could result from including contributions of the exchange coefficient c_X in bulk formulae. Comparing the $n = 1$ results presented in Appendix B with the $n = 2$ results, that we have focused on above, suggests that the statistical features of the SGS velocity flux enhancement vary smoothly with n.

Implementation in operational models of a stochastic parameterization scheme such as we have proposed is neither difficult nor numerically expensive. Because the residuals $\psi_n(N)$ are represented as a Gaussian random process with specified spatial covariance, realizations on a grid can be generated in principle by multiplying a vector of independent unit variance Gaussians by a matricial square-root of the covariance matrix such as the Cholesky decomposition (e.g. Rasmussen & Williams, 2006). In practice such a calculation may not be practical for a large grid, in which case dimensionality reduction such as the use of orthogonal basis functions in space may be necessary (as is done in the stochastic parameterizations used by the European Centre for Medium-Range Weather Forecasts, cf. Palmer et al. (2009)). Temporal dependence can be represented by modeling $\psi_n(N)$ as a multivariate Ornstein-Uhlenbeck process. Since the covariance between two grid boxes is only depending on distance between them and the estimated parameter values, the stochastic parameterization can be implemented for arbitrary grid structures.

As in Bessac et al. (2019), the present study has focused on the modification of model-simulated air-sea fluxes by SGS velocity variations, neglecting the contribution from other variations (and their correlations). In particular, it does not include an explicit treatment of the wind speed dependence of the exchange coefficient in bulk flux formulae. While this choice avoids tying the results of the analysis to any specific semi-empirical flux parameterization, it has the consequence that these results are more a theoretical proof of concept than a final form appropriate for incorporation in operational weather or climate models. Further analyses considering less idealized flux formulations are therefore needed. Finally, the 4×4 km resolution of the atmospheric model used as "truth" implies that our stochastic parameterization accounts only for mesoscale velocity variations; motions on the scale of boundary-layer turbulence are not included.

The statistical model formulation and parameter estimates provide insight into transitions across resolutions that are smooth and exhibit spatial variability in their intensity. As observed in Bessac et al. (2019), zonal anisotropy is stronger at coarser resolutions, and the overall variability decays at coarser resolutions suggesting a smaller need for stochastic enhancements of SGS variability. Precipitation, to which $\varepsilon_2(N)$ displays a strong first-order dependence through the regression model, may also play a role in the intensity

of the scale-dependence and its spatial variability. This result suggests a higher-order dependence of the flux errors on precipitation, although it could also simply reflect sampling variability in both quantities. It is also possible that the regression model could be improved by introducing additional predictors. In future work, it would be interesting to study this higher-order dependence of the flux errors on precipitation and investigate other potential predictors. Such an analysis of higher-order dependence might require the use of nonlinear models such as regime-switching models driven by precipitation states or synoptic conditions. One could as well consider covariates, in particular precipitation, in the description of the scale dependence in Equation 9. Such an approach would be analogous to that of Xie and Zhang (2015), where both model scale and a measure of atmospheric instability are considered in the scale-aware parameterization of cloud water. In general, exploring the processes imprinting on the spatial heterogeneity could serve to improve the accuracy of the statistical model.

Appendix A: Statistical Background

In the following, we provide more details for some of the statistical methods and concepts used above; more details can be found in Cressie and Wikle (2015), Gelfand et al. (2010), Gneiting et al. (2006), Rasmussen and Williams (2006), and Von Storch and Zwiers (2001).

Gaussian Processes

Gaussian distributions are widely used in statistics and other fields due to their mathematical properties, their flexibility in modeling purposes, and their role as the limiting distribution in the Central Limit Theorem. Unidimensional normal distributions can be generalized to a multidimensional setup in the following way. The random vector $Y = (Y_1, ..., Y_d)$ has a d-dimensional normal distribution with mean μ and covariance matrix Σ if its probability density function is given by:

$$f_d(y) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{(y-\mu)^T \Sigma^{-1} (y-\mu)}{2}\right), \text{ for all } y \in \mathbb{R}^d.$$

In this case, the notation $Y \sim \mathcal{N}_d(\mu, \Sigma)$ is standard. Similar to the unidimensional case, a multidimensional normal distribution is entirely characterized by its mean vector and its covariance matrix.

A stochastic process $Y(x)$, where $x \in \mathcal{X} \subset \mathrm{R}^k$, is an infinite-dimensional generalization of a random vector, where the collection of random variables is indexed continuously. In particular, a stochastic process $Y(x)$ is a Gaussian process (GP) if for any set of input "locations" $x_1, ..., x_d$, the joint probability distribution for the vector of responses $Y = (Y(x_1), ..., Y(x_d))$ is a multidimensional normal distribution. Following the properties of the multidimensional Gaussian distribution, a Gaussian process is entirely characterized by a mean function $\mu : \mathcal{X} \to \mathbb{R}$ and a covariance function $K : \mathcal{X} \times \mathcal{X} \to (0, +\infty)$.

In geostatistics, parametric models and covariates are commonly used to express the mean μ whereas the covariance expression K relies mostly on parametric models such as the squared-exponential or Matern covariances. Coefficients describing these parametric models are referred to as parameters in statistics and usually denoted as θ. They are calibrated in order for the parametric model to match the input data as well as possible according to specified criteria (e.g. maximizing the log likelihood). An introductory review of Gaussian Processes is presented in Rasmussen and Williams (2006).

Covariance Kernels

The covariance function K of a Gaussian process can be decomposed as the product of a scalar variance term σ by a correlation function that provides a measure of linear dependence between the stochastic process at two points (in space and/or time). Parametric forms are generally used to represent the correlation such as exponential, squared-exponential or Matern models (e.g. Chapter 4 in Rasmussen &Williams, 2006). A "nugget" term δ is often added to capture additional local variability which is not resolved by the product variance-correlation. For instance, a commonly used full covariance model is the squared-exponential model: $K(d) = \sigma \exp(-(d/\rho)^2) + \delta \mathbf{1}_{d=0}$, where σ is the variance, d is a distance in space or time, ρ the decorrelation range over space or time, δ is the nugget, and $\mathbf{1}$ is the indicator function. In this case, one would estimate the parameters σ, ρ and δ through an estimation procedure (e.g. maximizing the log-likelihood).

One can extend the covariance concept to a spatio-temporal field $Y(t, r)$ where the focus is on the covariance between $Y(t, r)$ and $Y(t + h, r + u)$, where t is time, r a spatial location, h and u are respectively a time lag and a spatial translation vector. This relates to the study of the covariance between spatial locations that are considered at different observation times (Gneiting et al., 2006).

Similarly to spatial covariance, if the covariance $Cov(Y(t, r), Y(t + h, r + u))$ does not depend on the spatio-temporal location (t, r) and depends only on the lags h and u, the covariance is said to be second-order stationary and is denoted by $K(h, u)$. However, by building more complex covariances, it is likely to encounter non-stationarity especially in a space-time context. Mathematical representations of non-stationary space-time covariance is still very challenging. One way among others to deal with non-stationary is by considering local sub-domains in which stationary is assumed, as performed in the above study. A concept associated with space-time covariances is the space-time separability which neglects the interaction between space and time. The assumption is commonly made for computational tractability; however, it is often unrealistic with geophysical data. In the above study, we consider a three-dimensional space-time distance to account for the non-separability of the data set.

Maximum-Likelihood Estimation and its Uncertainty

To estimate model parameters, Maximum Likelihood estimation (MLE) is one of the most popular methods given the optimality of the estimators and their asymptotic properties. The intuition behind the MLE is to estimate a set of parameters that makes the data the most likely to occur under the chosen parametric probabilistic model.

Definition Let $Y_1, ..., Y_n$ be a set of random variables, which realizations are denoted $y_1, ..., y_n$, with joint probability density function $L(\theta; y_1, ..., y_n) = P_\theta(Y_1 = y_1, ..., Y_n = y_n)$ that depends on a set of parameters $\theta \in \mathrm{R}^p$. Then $\hat{\theta}_n$ is the *Maximum-Likelihood estimator* for θ associated to an observed set of realizations $(y_1, ..., y_n)$ if

$$\hat{\theta}_n = \mathrm{argmax}_\theta L(\theta; y_1, \ldots, y_n).$$

In practice, likelihood values for large numbers of data are very small, so the log-likelihood is optimized instead

$$\ell(\theta; y_1, \ldots, y_n) = \log(L(\theta; y_1, \ldots, y_n)).$$

Note that n is the number of realizations (replicates) and has to be distinguished from the dimension d of a random vector. Oftentimes, an analytical form of the maximum is not available due to the complexity of proposed models. Numerical optimization techniques are used instead to compute the optimum iteratively. The choice of optimization method varies depending on the level of non-linearity and the constraints, e.g. parameter positivity, of the likelihood and its parameters. Initial conditions are often hand-picked from a priori empirical analysis or least-square estimation for instance.

Asymptotic Properties and Uncertainty Quantification

Under certain regularity conditions, maximum-likelihood estimators are consistent (they converge in probability to the "true" value θ_0) and asymptotically Gaussian distributed. More precisely, under certain regularity conditions, the following convergence in distribution (denoted by $\xrightarrow{d}$) of $\hat{\theta}$ holds

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \underset{n \to +\infty}{\xrightarrow{d}} \mathcal{N}_p(0, \mathcal{I}^{-1}(\theta_0)),$$

where $\mathcal{I}(\theta_0) = \left(E\left(-\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log L(\theta_0; Y) \right) \right)_{i,j=1..p}$ is the Fisher information matrix. In practice, the Fisher information matrix is extracted from the Hessian (i.e., curvature) of the log-likelihood evaluated at the optimum $\hat{\theta}_n$ and provides the variance-covariance associated with the estimation of the parameters θ. Intuitively, the level of confidence in the estimates of the parameters is inversely proportional to the curvature of the log-likelihood (e.g. poor confidence in the estimates when the curvature is flat).

Delta method: The Delta method enables approximation of an asymptotic distribution of a transformed random variable. It is commonly used to complement the Central Limit Theorem, though it has other applications. For simplicity, let us assume the following convergence in distribution holds for a sequence of random vectors $X_1, ..., X_n$ of $\mathbb{R}^p$: $\sqrt{n}[X_n - \theta_0] \xrightarrow{d} \mathcal{N}_p(0, \Sigma)$. For a function $g : \mathbb{R}^p \to \mathbb{R}^s$ differentiable in θ_0, the following convergence of the transformed random vectors $g(X_1), ..., g(X_n)$ holds:

$$\sqrt{n}[g(X_n) - g(\theta_0)] \xrightarrow{d} \mathcal{N}_s\left(0, \nabla g(\theta_0)^T \cdot \Sigma \cdot \nabla g(\theta_0)\right),$$

where ∇g is the gradient of g. The Delta method is used in Section 4 to compute the estimation variance of θ_Z, θ_M, θ_T, σ, and γ as transformations of the coefficients $\theta_{Z,1}$, $\theta_{Z,2}$, ..., σ_1, σ_2 through the functions (9). Gradients of the functions (9) are computed with respect to each coefficient $\theta_{Z,1}$, $\theta_{Z,2}$, ..., σ_1, σ_2 and combined with the Fisher information matrix of the log-likelihood $\log L(\theta; \psi_2(N_0), \ldots, \psi_2(N_1))$ from Equation 10 with $\theta = (\theta_{Z,1}, \theta_{Z,2}, \theta_{M,1}, \theta_{M,2}, \theta_{T,1}, \theta_{T,2}, \sigma_1, \sigma_2, \gamma_1, \gamma_2)$.

Appendix B: Results for n = 1

In this section, we present results relevant to heat and freshwater fluxes from fitting the scale-aware regression model (7) on the error process $\varepsilon_1(N)$ for $n = 1$ and the local scale-aware GP from Equations 8 and 9 on the regression residuals $\psi_1(N)$. The fit is performed on the same three resolutions considered previously N

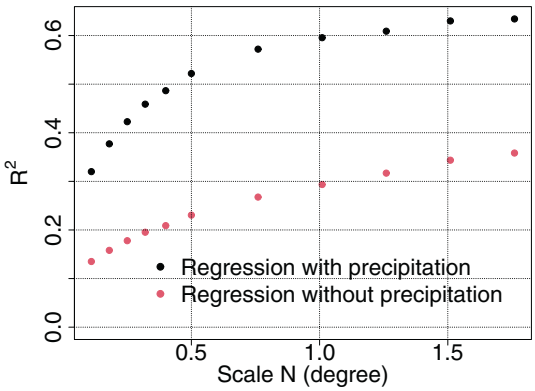

Figure B1. Same as Figure 2, but for flux exponent $n = 1$.

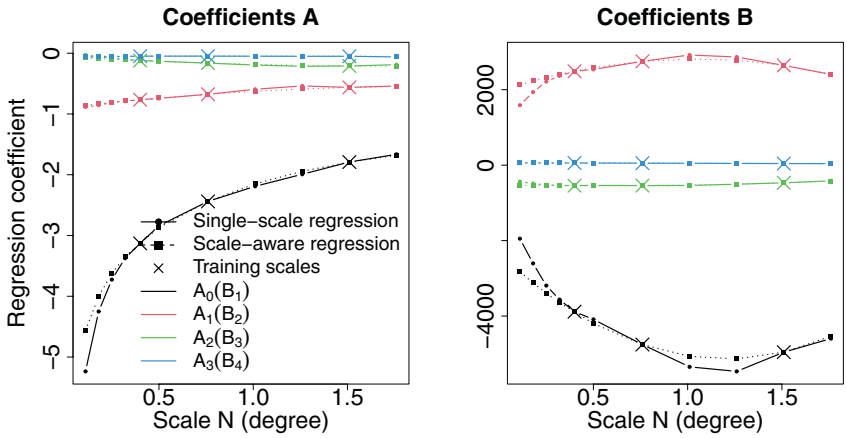

Figure B2. Same as Figure 3, but for flux exponent $n = 1$.

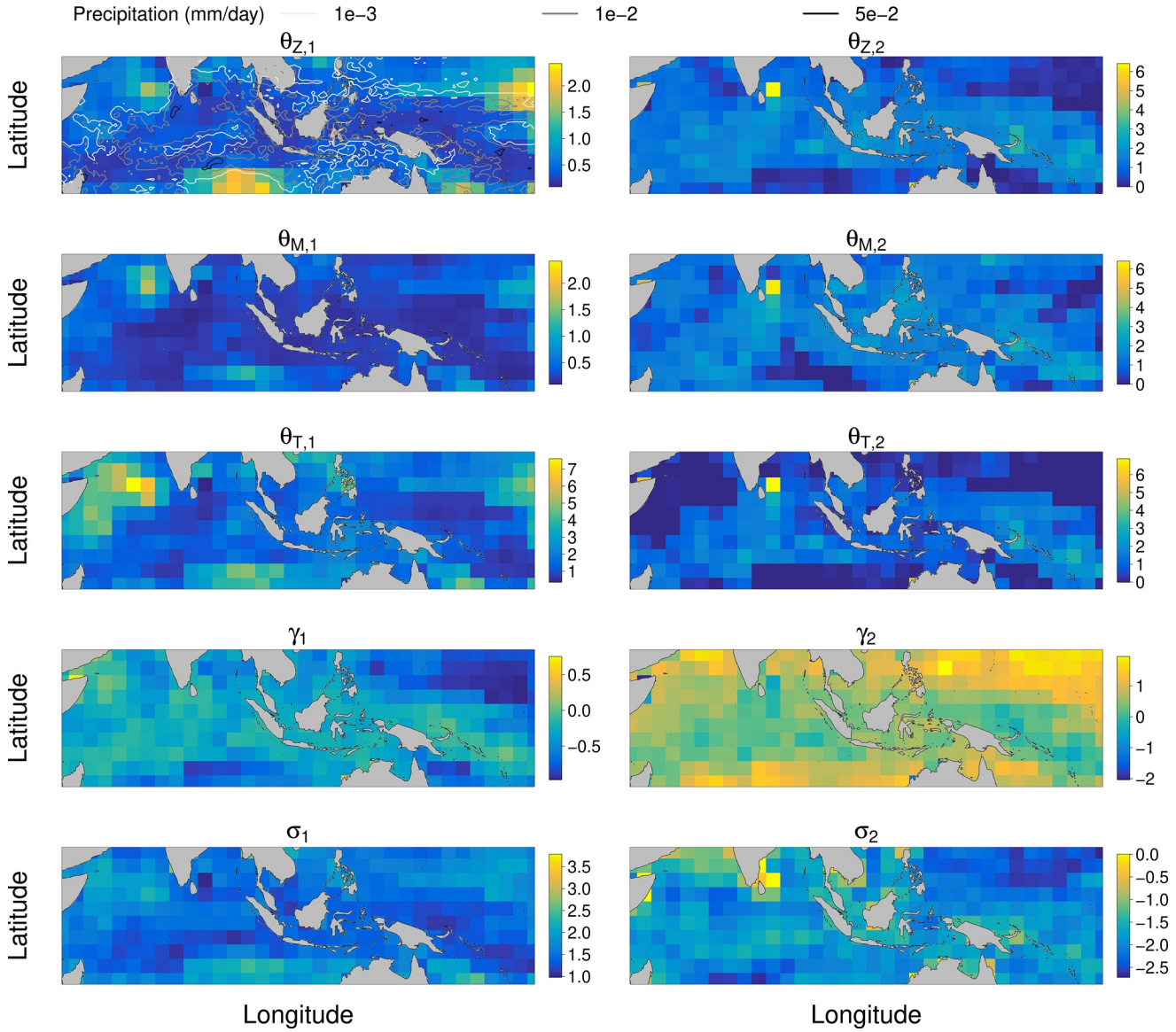

Figure B3. Same as Figure 7, but for flux exponent $n = 1$.

$= (0.40°, 0.76°, 1.51°)$ and in the same fashion as described earlier. Overall results from fitting the full statistical model on $\varepsilon_1(N)$ and $\psi_1(N)$ are similar to those of the $n = 2$ fit and suggest smooth transitions of the quantities of interest across different values of the exponent n. For instance, Figure B1 presents the R^2-coefficient of regression (5) fitted for $n = 1$, indicating a stronger importance of the resolved flux compared to the case $n = 2$. In Figure B2, the fits of the scale-aware regression model (7) are depicted. These show overall similar shapes and quality of fit compared to the case $n = 2$. Figure B3 shows the spatial distribution of the estimated parameters of the local GP (8) and (9) fitted on $\psi_1(N)$ showing similar overall behavior as for the case $n = 2$.

Data Availability Statement

Data from the Cascade project are freely available upon registration with the NERC Centre for Environmental Data Analysis (CEDA) at https://catalogue.ceda.ac.uk/uuid/20981e3052a66ca71c2ba92b94760150 and referenced as (Lister & Woolnough, 2008). The *R* codes used for this study are available for download at https://github.com/jbessac/sgs_fluxes_scaleaware.

Acknowledgments
This research started in a working group supported by the Statistical and Mathematical Sciences Institute (SAMSI). The effort of J. Bessac is based in part on work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contract DE-AC02-06CH11347. The research of HMC was supported by NERC Grant NE/P018238/1. A. H. Monahan and K. Endo acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) (funding reference number RGPIN-2019-04986). N. Weitzel acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project no. 395588486). The authors thank Sébastien Blein and one anonymous reviewer for their constructive comments.

References

Berner, J., Achatz, U., Batté, L., Bengtsson, L., De La Cámara, A., Christensen, H. M., et al. (2017). Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, *98*(3), 565–588.

Bessac, J., Monahan, A. H., Christensen, H. M., & Weitzel, N. (2019). Stochastic parameterization of subgrid-scale velocity enhancement of sea surface fluxes. *Monthly Weather Review*, *147*(5), 1447–1469.

Blein, S., Roehrig, R., Voldoire, A., & Faure, G. (2020). Meso-scale contribution to air-sea turbulent fluxes at GCM scale. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 2466–2495. https://doi.org/10.1002/qj.3804

Christensen, H. M. (2020). Constraining stochastic parametrisation schemes using high-resolution simulations. *Quarterly Journal of the Royal Meteorological Society*, *146*(727), 938–962.

Christensen, H. M., Berner, J., Coleman, D., & Palmer, T. N. (2017). Stochastic parametrisation and the El Niño-Southern oscillation. *Journal of Climatology*, *30*(1), 17–38.

Cressie, N., & Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., & Wood, N. (2005). A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, *131*(608), 1759–1782. https://doi.org/10.1256/qj.04.101

Dörenkämper, M., Optis, M., Monahan, A. H., & Steinfeld, G. (2015). On the offshore advection of boundary-layer structures and the influence on offshore wind conditions. *Boundary-Layer Meteorology*, *155*, 459–482. https://doi.org/10.1007/s10546-015-0008-x

Dorrestijn, J., Crommelin, D. T., Siebesma, A. P., Jonker, H. J. J., & Selten, F. (2016). Stochastic convection parameterization with Markov Chains in an intermediate-complexity GCM. *Journal of the Atmospheric Sciences*, *73*(3), 1367–1382.

Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A., & Edson, J. B. (2003). Bulk parameterization of air-sea fluxes: Updates and verification for the COARE algorithm. *Journal of Climate*, *16*, 571–591.

Fox-Kemper, B., Adcroft, A., Boning, C. W., Chassignet, E. P., Curchitser, E., Danabasoglu, G., et al. (2019). Challenges and prospects in ocean circulation models. *Frontiers in Marine Science*, *6*(65). https://doi.org/10.3389/fmars.2019.00065

Gagne, D. J., Christensen, H. M., Subramanian, A. C., & Monahan, A. H. (2020). Machine learning for stochastic parameterization: Generative adversarial networks in the Lorenz'96 model. *Journal of Advances in Modeling Earth Systems*, *12*(3), e2019MS001896.

Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of spatial statistics*. CRC Press.

Gneiting, T., Genton, M. G., & Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability, and full symmetry. *Monographs on Statistics and Applied Probability*, *107*, 151.

Godfrey, J., & Beljaars, A. (1991). The turbulent fluxes of buoyancy, heat and moisture at the air-sea interface at low wind speeds. *Journal of Geophysical Research*, *96*, 22043–22048.

Grooms, I., Majda, A. J., & Smith, K. S. (2015). Stochastic superparameterization in a quasigeostrophic model of the Antarctic circumpolar current. *Ocean Modelling*, *85*, 1–15.

Holloway, C. E., Woolnough, S. J., & Lister, G. M. S. (2012). Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study. *Quarterly Journal of the Royal Meteorological Society*, *138*(668), 1692–1708.

Holloway, C. E., Woolnough, S. J., & Lister, G. M. S. (2013). The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study, Part I: Characterization of large-scale organization and propagation. *Journal of the Atmospheric Sciences*, *70*(5), 1342–1369.

Holloway, C. E., Woolnough, S. J., & Lister, G. M. S. (2015). The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study, Part II: Processes leading to differences in MJO development. *Journal of the Atmospheric Sciences*, *72*(7), 2719–2743. https://doi.org/10.1175/JAS-D-14-0308.1

Huang, D., Campos, E., & Liu, Y. (2014). Statistical characteristics of cloud variability. Part 1: Retrieved cloud liquid water path at three ARM sites. *Journal of Geophysical Research: Atmosphere*, *119*(18), 10812–0828. https://doi.org/10.1002/2014JD022001

Huang, D., & Liu, Y. (2014). Statistical characteristics of cloud variability. Part 2: Implication for parameterizations of microphysical and radiative transfer processes in climate models. *Journal of Geophysical Research: Atmosphere*, *119*(18), 10829–10843. https://doi.org/10.1002/2014JD022003

Hyndman, R. J., & Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational & Graphical Statistics*, *19*(1), 29–45.

Keane, R. J., Craig, G. C., Keil, C., & Zängl, G. (2014). The Plant-Craig stochastic convection scheme in ICON and its scale adaptivity. *Journal of the Atmospheric Sciences*, *71*(9), 3404–3415. https://doi.org/10.1175/JAS-D-13-0331.1

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S. T. K., Balsamo, G., Bechtold, P., et al. (2017). Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, *143*(707), 2315–2339.

Lister, G., & Woolnough, S. (2008). *CASCADE—Scale interactions in the tropical atmosphere model runs*. Natural Environment Research Council (NERC) Centre for Environmental Data Analysis (CEDA). Retrieved from http://catalogue.ceda.ac.uk/uuid/20981e3052a66ca71c2ba92b94760150

Love, B. S., Matthews, A. J., & Lister, G. M. S. (2011). The diurnal cycle of precipitation over the Maritime continent in a high-resolution atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, *137*(657), 934–947. https://doi.org/10.1002/qj.809

Mahrt, L., & Sun, J. (1995). The subgrid velocity scale in the bulk aerodynamic relationship for spatially averaged scalar fluxes. *Monthly Weather Review*, *123*, 3032–3041.

National Academies of Sciences Engineering, and Medicine. (2016). *Next generation earth system prediction: Strategies for subseasonal to seasonal forecasts*. The National Academies Press. https://doi.org/10.17226/21873

Palmer, T. (2019). Stochastic weather and climate models. *Nature Reviews Physics*, *1*, 463–471. https://doi.org/10.1038/s42254-019-0062-2

Palmer, T., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., et al. (2009). Stochastic parameterization and model uncertainty (Tech. Rep.). European Centre for Medium-Range Weather Forecasts.

Pearson, K. J., Hogan, R. J., Allan, R. P., Lister, G. M. S., & Holloway, C. E. (2010). Evaluation of the model representation of the evolution of convective systems using satellite observations of outgoing longwave radiation. *Journal of Geophysical Research*, *115*(20), 1–11. https://doi.org/10.1029/2010JD014265

R Core Team. (2020). *R: A language and environment for statistical computing [computer software manual]*. Retrieved from https://www.R-project.org/

Ramsay, J. O., Graves, S., & Hooker, G. (2020). *FDA: Functional data analysis [computer software manual]*. Retrieved from https://CRAN.R-project.org/package=fda(Rpackageversion5.1.4)

Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning*. The MIT Press.

Redelsperger, J.-L., Guichard, F., & Mondon, S. (2000). A parameterization of mesoscale enhancement of surface fluxes for large-scale models. *Journal of Climate*, *13*, 402–421.

Sakradzija, M., Seifert, A., & Dipankar, A. (2016). A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems*, *8*(2), 786–812.

Sakradzija, M., Seifert, A., & Heus, T. (2015). Fluctuations in a quasi-stationary shallow cumulus cloud ensemble. *Nonlinear Processes in Geophysics*, *22*, 65–85. https://doi.org/10.5194/npg-22-65-2015

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., et al. (2019). DYAMOND: The DYnamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Prog. Earth and Planetary Sci.*, *6*, 1–17. https://doi.org/10.1186/s40645-019-0304-z

Strommen, K., Christensen, H. M., MacLeod, D., Juricke, S., & Palmer, T. (2019). Progress towards a probabilistic earth system model: Examining the impact of stochasticity in EC-earth v3. 2. *Geoscientific Model Development*, *12*(7), 3099–3118.

Sun, J., Howell, J., Esbensen, S., Mahrt, L., Greb, C., Grossman, R., & LeMone, M. (1996). Scale dependence of air-sea fluxes over the western equatorial pacific. *Journal of the Atmospheric Sciences*, *53*(21), 2997–3012.

Sun, Y., & Genton, M. G. (2011). Functional boxplots. *Journal of Computational & Graphical Statistics*, *20*(2), 316–334.

Vickers, D., & Esbensen, S. K. (1998). Subgrid surface fluxes in fair weather conditions during TOGA COARE: Observational estimates and parameterization. *Monthly Weather Review*, *126*, 620–633.

Visso, G., & Lucarini, V. (2018). A proof of concept for scale-adaptive parameterizations: The case of the Lorenz '96 model. *Quarterly Journal of the Royal Meteorological Society*, *144*, 63–75. https://doi.org/10.1002/qj.3184

Von Storch, H., & Zwiers, F. W. (2001). *Statistical analysis in climate research*. Cambridge University Press.

Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited. *Limnology and Oceanography: Methods*, *12*, 351–362.

Williams, A. G. (2001). A physically based parametrization for surface flux enhancement by gustiness effects in dry and precipitating convection. *Quarterly Journal of the Royal Meteorological Society*, *127*, 469–491.

Xie, X., & Zhang, M. (2015). Scale-aware parameterization of liquid cloud inhomogeneity and its impact on simulated climate in CESM. *Journal of Geophysical Research: Atmosphere*, *120*(16), 8359–8371. https://doi.org/10.1002/2015JD023565

Zanna, L., Porta Mana, P., Anstey, J., David, T., & Bolton, T. (2017). Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction. *Ocean Modelling*, *111*, 66–80.

Zeng, X., Zhang, Q., Johnson, D., & Tao, W.-K. (2002). Parameterization of wind gustiness for the computation of ocean surface fluxes at different spatial scales. *Monthly Weather Review*, *130*, 2125–2133.