Stochastic parameterization of subgrid-scale velocity enhancement of sea surface fluxes

Julie Bessac[1,*], Adam H. Monahan[2], Hannah M. Christensen[3], and Nils Weitzel[4,5]

[1] Mathematics and Computer Science Division
Argonne National Laboratory, Lemont, IL, USA
[*] Corresponding author, email: jbessac@anl.gov

[2] School of Earth and Ocean Sciences
University of Victoria, Victoria, BC, Canada

[3] Department of Physics
University of Oxford, Oxford, UK

[4] Institut für Geowissenschaften und Meteorologie
Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

[5] Institut für Umweltphysik
Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany

January 31, 2019

Abstract

Subgrid-scale (SGS) velocity variations result in grid-scale sea-surface flux enhancements that must be parameterized in weather and climate models. Traditional parameterizations are deterministic in that they assign a unique value of the SGS velocity flux enhancement to any given configuration of the resolved state. In this study, we assess the statistics of SGS velocity flux enhancement over a range of averaging scales (as a proxy for varying model resolution) through systematic coarse-graining of a convection-permitting atmospheric model simulation over the Indian Ocean and West Pacific Warm Pool. Conditioning the statistics of the SGS velocity flux enhancement on (1) the fluxes associated with the resolved winds, and (2) the precipitation rate, we find that the lack of a separation between "resolved" and "unresolved" scales results in a distribution of flux enhancements for each configuration of the resolved state. That is, the SGS velocity flux enhancement should be represented stochastically rather than deterministically. The spatial and temporal statistics of the SGS velocity flux enhancement are investigated by using basic descriptive statistics and through a fit to an anisotropic space-time covariance structure. Potential spatial inhomogeneities of the statistics of the SGS velocity flux enhancement are investigated through regional analysis, although because of the relatively short duration of the simulation (9 days) distinguishing true inhomogeneity from sampling variability is difficult. Perspectives for the implementation of such a stochastic parameterization in weather and climate models are discussed.

1 Introduction

Near-surface winds exert an important control on exchanges of mass, energy, and momentum between the atmosphere and the underlying surface. In weather and climate models, air-sea exchanges are generally expressed as a combination of the concentration difference between the atmosphere and the sea surface and a function of the near-surface wind speed s (conventionally at the anemometer height of 10 m):

$$\text{surface flux of } X = \overline{\rho_a} c_x(\overline{s})\overline{s}\left[\overline{X}_s - \overline{X}_a\right] \quad (1)$$

In Eq. (1), X_s and X_a are respectively the "concentrations" of quantity X (in units of X per unit mass of air) at the surface and at the anemometer height, ρ_a is the surface air density, and $c_x(s)$ is a non-dimensional function of the wind speed (and potentially other variables such as near-surface stratification). The exchange coefficient $c_x(s)$ depends on wind speed through, for instance, changes in surface roughness, or bubble injection/spray production by breaking surface waves (e.g., Drennan, 2006; Edson, 2008; Garbe et al., 2014). The overbars in Eq. (1) denote time averaging (typically over windows of ~ 10 min) separating the turbulent and Reynolds-averaged variations. Although based on theoretical foundations, these parameterizations are generally largely empirical. Furthermore, although they are averaged in time, the expressions relate fluxes at a single point in space to the atmospheric state (and specifically the wind speed) at that location.

Numerical weather and climate models have finite spatial resolution, and require surface fluxes averaged over model gridboxes. Through the dependence of c_s on $\overline{s}$, the bulk flux parameterization is generally a nonlinear function of wind speed. Thus the flux averaged over a region of space (such as a gridbox) does not equal the flux that would be computed from the averaged wind speed. Furthermore, the gridbox-averaged wind speed itself is not available from the weather or climate model. Rather, the models directly simulate the gridbox-averages of the horizontal wind components. Denoting spatial averaging by angle brackets and the wind vector field by $\mathbf{u} = (u, v)$ we have

$$\langle s \rangle = \langle |\mathbf{u}| \rangle \geq |\, \langle \mathbf{u} \rangle \,|, \tag{2}$$

The inequality in Eq. (2), which follows mathematically from Jensen's inequality and the fact that wind speed is a convex function of the wind components, results physically from the existence of subgrid-scale (SGS) velocity variations and is generally most important under conditions of weak mean winds. A similar inequality holds for the time averaging used to separate the turbulent and Reynolds-averaged parts of the flow (e.g., Beljaars, 1994; Mahrt and Sun, 1995).

For many fluxes (e.g. momentum, gases, and particles), the function that sets the dependency of flux on wind speed, $h(s) = c_x(s)s$, is found to be convex ($h'' \geq 0$). It follows that

$$\overbrace{\left|\left\langle \overline{\mathrm{flux}(s)} \right\rangle\right| \geq \mathrlap{\underbrace{\phantom{\left|\mathrm{flux}\left(\overline{\langle s \rangle}\right)\right| \geq \left|\mathrm{flux}\left(|\overline{\langle \mathbf{u} \rangle}|\right)\right|}}_{\text{SGS velocity variations}}}\left|\mathrm{flux}\left(\overline{\langle s \rangle}\right)\right|}^{\text{convexity of } c_x(s)s} \geq \left|\mathrm{flux}\left(|\overline{\langle \mathbf{u} \rangle}|\right)\right|. \tag{3}$$

where the first inequality follows again from Jensen's inequality applied to the function h, while the second follows from inequality (2). The spatially and temporally averaged flux (the left-hand quantity in inequality (3)) is what is desired, while the flux computed from the norm of the space-time mean of the wind vector (the quantity on the right of (3)) is what is directly available from the resolved state in models.

The fact that space-time-averaged fluxes exceed the fluxes computed from the space-time-averaged wind vector has been recognized for many years, and a number of studies have considered ways of parameterizing this difference (e.g., Godfrey and Beljaars, 1991; Mahrt and Sun, 1995; Vickers and Esbensen, 1998; Redelsperger et al., 2000; Williams, 2001; Zeng et al., 2002). A standard approach accounts for the difference between $\overline{\langle s \rangle}$ and

$|\overline{\langle \mathbf{u} \rangle}|$ due to SGS velocity variations through a SGS velocity flux enhancement term:

$$\overline{\langle s \rangle}^2 = |\overline{\langle \mathbf{u} \rangle}|^2 + s_{SGS}^2. \tag{4}$$

Standard parameterizations of s_{SGS} account for surface flux enhancement due to disorganized near-surface SGS flow associated with shallow and deep convection:

$$s_{SGS}^2 = \left(\beta \overline{\langle w^* \rangle}\right)^2 + g\left(\overline{\langle P \rangle}\right), \tag{5}$$

where $\overline{\langle w^* \rangle}$ is the free convective velocity scale determined by the resolved surface buoyancy flux and $\overline{\langle P \rangle}$ is the gridbox-averaged precipitation rate. The coefficient $\beta \sim 1$ and the function $g\left(\overline{\langle P \rangle}\right)$ have typically been determined empirically from field measurements and cloud-resolving model simulations. Mahrt and Sun (1995) replaced $g\left(\overline{\langle P \rangle}\right)$ with a term that represented all mesoscale contributions to the area-mean flux (not just those associated with convective precipitation). The observationally based study of Vickers and Esbensen (1998) parameterized s_{SGS} from observations taken under fair weather conditions. Redelsperger et al. (2000) and Williams (2001) demonstrated the importance of the contribution of deep convection (represented by precipitation) to observed subgrid-scale velocity variations. Parameterization of the contribution of boundary-layer eddies and deep convection to surface flux enhancement was also considered by Zeng et al. (2002), using results from 1 km $\times$ 1 km simulations of a cloud resolving model (CRM) on a 512 km $\times$ 512 km domain in the tropical North Atlantic. Studying the dependence of SGS flux enhancements on averaging scale, Mahrt and Sun (1995),Vickers and Esbensen (1998), and Zeng et al. (2002) all found that the corrections become larger for coarser resolutions and proposed power-law expressions for the dependences. In all these studies, deterministic parameterizations of the subgrid-scale velocity flux enhancement were obtained by empirical fits to data. The physical significance of the scatter of these data around the parameterization curves was not addressed.

Another approach to accounting for SGS velocity variations is to explicitly model these through an assumed parametric probability distribution conditioned on resolved scales (e.g., Cakmur et al., 2004; Capps and Zender, 2008; Ridley et al., 2013; Zhang et al., 2016). While the parameterizations considered in these studies are probabilistic, they are still deterministic. The probability distributions employed are used to compute statistical moments across the gridbox, rather than to generate a sequence of random values.

Deterministic parameterizations of subgrid-scale processes, in which a unique configuration of the resolved variables is associated with a unique value of the parameterized tendency, are only theoretically justified in the presence of a large separation between resolved and unresolved scales. In the absence of such a scale separation, a distribution of parameterized tendencies will be associated with each configuration of the resolved state, and the mathematical form of the parameterization will be stochastic (see the recent review by Berner et al., 2017). The data scatter around the curves corresponding to deterministic parameterizations of SGS velocity flux enhancement demonstrates the existence of such stochastic fluctuations (particularly in the CRM-based study of Zeng et al., 2002, in which the deviations clearly cannot be attributed to measurement error). As is detailed in the review of Berner et al. (2017), the importance of explicitly accounting for stochastic variations around a deterministic parameterization has been demonstrated in a number of studies on weather, seasonal, and climate time scales. In the specific context of air-sea fluxes, Williams (2012) demonstrated that including stochastic flux fluctuations has an effect not just on model variability, but on its mean state (through rectified deepening of the simulated mixed layer). Including stochastic parameterizations into climate models also improves the representation of processes sensitive to air-sea coupling, such as the El Niño-Southern Oscillation (Christensen et al., 2017; Yang et al., 2019), through improving the high-frequency atmospheric response to changes in sea surface temperature.

In this study, we revisit the question of SGS flux enhancement using a nine-day simulation of a convection-permitting (4 km resolution) atmospheric model on a large tropical domain (20°S to 20°N, from East Africa to 180°W). By systematically coarse-graining the high-resolution simulation, we are able to analyze the relationship between "true" gridbox-averaged fluxes and the fluxes computed from the gridbox-mean vector wind (the "resolved flux"). We extend previous analyses not only by estimating the deterministic dependence of the true fluxes on resolved variables, but also by modelling the residuals around this empirical fit as a space-time random field. We emphasize the distinction between such a parameterization and the probabilistic but deterministic ones of Cakmur et al. (2004); Capps and Zender (2008); Ridley et al. (2013), and Zhang et al. (2016). The parameterization we develop samples from a random space-time-field at each time step: it is explicitly stochastic.

Rather than explicitly develop a parameterization of s_{SGS}, we instead consider the difference between the true and resolved fluxes as a random variable conditioned on the resolved flux and the precipitation rate. While this approach is more abstract, it has the benefit of being able to simultaneously account for the differences in resolved and true fluxes due to SGS velocity variations and the nonlinearity of the dependence of the flux on wind speed. Parameterizations constructed in terms of s_{SGS} account only for the first of these two issues.

This study is organized as follows. A description of the high-resolution simulation used in our analysis is presented in Section 2. Section 3 presents the results of the analysis. A discussion and conclusions are presented in Section 4.

2 Model Description

Ideally, subgrid-scale wind variability statistics would be measured from observational data sets. However, our analysis requires data of a sufficiently high spatial resolution over a large domain, for which a suitable observational data set is not available. Instead, we use an existing high-resolution model simulation as our "truth", produced as part of the UK Natural Environment Research Council (NERC) "Cascade" project (Pearson et al., 2010; Love et al., 2011; Holloway et al., 2012). The Cascade project produced convection-permitting, cloud system-resolving simulations with resolutions ranging from 1.5 km to 12 km over several large tropical domains using the UK Met Office's Unified Model (MetUM).

For this paper, we use the Cascade 4 km resolution tropical Indo-Pacific Warm Pool integration. This Cascade simulation has proven useful for assessing stochastic parameterization schemes in other coarse-graining studies (Christensen, 2018). For full details of the simulation, see 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. The model is semi-Lagrangian and non-hydrostatic. The model has 70 terrain-following hybrid vertical levels, with a variable vertical resolution ranging from tens of meters in the boundary layer to 250 m in the free troposphere, and with the model top at 40 km. The time step was 30 s. Initial conditions were specified from the ECMWF operational analysis. The 4 km simulation formed one of a hierarchy of simulations. First, a 12 km parametrized 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 8 model grid points.

The 4 km resolution simulation is "convection permitting". The Gregory and Rowntree (1990) convection scheme is adapted such that at large Convectively Available Potential Energy (CAPE) values the convection scheme is effectively turned off, allowing the model's dynamical equations to represent strong convective events. The convection scheme is active only for weakly unstable situations. The chosen simulation uses Smagorinsky subgrid mixing in the horizontal and vertical dimensions. The simulation begins on 6 April 2009 and spans 10 days, chosen as a case study of an active Madden-Julian oscillation (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 validation of the Cascade simulation has been reported by Holloway et al. (2012, 2013, 2015). The simulation is shown to produce a realistic MJO, including realistic convective organization, MJO strength, and propagation speed (Holloway et al., 2013). This is likely because the model accurately captures fundamental convective processes, including a realistic vertical heating structure (Holloway et al., 2015), realistic generation of eddy available potential energy (Holloway et al., 2013), improved profiles of moist static energy and saturation moist static energy compared to simulations with parameterized convection (Holloway et al., 2012), and a precipitation distribution that is similar to that diagnosed from Tropical Rainfall Measuring Mission (TRMM) observations (Holloway et al., 2012). 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).

Figure 1 presents maps of the mean and standard deviation of the wind speed at the base 4 km $\times$ 4 km resolution. Large-scale structure in the mean wind speed field across the domain is evident, with a particular contrast between high wind speeds in the equatorward flanks of the subtropical highs in the Southern Indian, North Pacific, and South Pacific oceans; and relatively small wind speeds in the equatorial band and Northern Indian Ocean. The wind speed standard deviation field displays more localized regions of relatively large values. Maps of the 50th and 95th percentiles of precipitation rate (Figure 1) also show considerable spatial heterogeneity. In particular, there are large regions of the domain in which the median precipitation rate is 0 mm/day; a precipitation rate of zero is also the 95th percentile in the Arabian Sea. When interpreting these and subsequent figures, one must remember that the simulation is of quite short duration. We expect that sampling variations will contribute to spatio-temporal variations of statistics.

3 Results

In this study, we focus on the effects of spatial averaging on air-sea fluxes computed from bulk formulae (i.e., Eq. 1). As such, all fields we consider are assumed to be Reynolds averaged. This assumption is also consistent with the parameterized nature of the model used to produce the simulations that we analyze. For the rest of the study, we will no longer use an overbar to denote time averages.

Rather than focusing on expressions for specific fluxes (such as water vapor, sensible heat, gases, or aerosols), we consider a generic power-law form for the dependence of flux on wind speed. Furthermore, our focus is on subgrid-scale variations in winds, so we will not consider the explicit dependence of fluxes on other state variables (such as the dependence of the exchange coefficient $c_x(s)$ on near-surface stability through the Obukhov length). We therefore take as a simplified nondimensional representation of air-sea flux:

$$F_n = \left(\frac{s}{s_0}\right)^n, \tag{6}$$

where $s_0 = 1\ \mathrm{ms}^{-1}$ is a speed scale. Scaling the wind speed dependence in this way facilitates comparisons of the nondimensional flux F for different values of the exponent n. Note that for $n = 1$, the flux function is linear in the wind speed and the difference between true and resolved fluxes results only from the difference between $\langle s \rangle$ and $|\langle \mathbf{u} \rangle|$.

The power-law dependence of fluxes on wind speed assumed here is a simplifying approximation. Neglecting the wind speed dependence of the exchange coefficients c_X and the effect of surface currents, atmospheric boundary layer theory predicts values of $n = 1$ for heat and water vapour fluxes and $n = 2$ for momentum fluxes (e.g. Drennan, 2006). When sea-state dependence of exchange coefficients is parameterized in terms of local wind speed, these functional dependencies are changed (and may not be polynomial). A range of empirically-based values of n have been reported for gases (e.g. Fig. 2.10 of Garbe et al., 2014), with variations depending on factors such as how fluxes are influenced by bubble injection beneath breaking waves. Relatively high values of n are often used for aerosol fluxes, which are strongly affected by the production of spray in whitecaps (e.g. the value of 3.41 used for sea salt in Zhang et al., 2016). For illustrative purposes, we consider the values $n = 1, 2$, and 3 in this study.

True fluxes averaged over a "gridbox" domain of area $N \times N$ (with N in degrees) are then defined as

$$F_{N,n}^{(T)} = \left\langle \left(\frac{s}{s_0} \right)^n \right\rangle_N . \tag{7}$$

As the focus of this study is air-sea fluxes, gridboxes at any coarsening scale containing land points are excluded from the analysis. Estimates of the probability density function (pdf) of $\log_{10}\left(F_n^{(T)}\right)$ computed from the raw 4 km resolution model output for $n = 1$, 2, and 3 are presented in Figure 2 (left column). The flux distributions move to larger values as n increases.

The resolved flux (that is, the flux that would be computed from the gridbox-mean vector wind $(\langle u \rangle_N, \langle v \rangle_N)$) is defined as

$$F_{N,n}^{(R)} = \left(\frac{\sqrt{\langle u \rangle_N^2 + \langle v \rangle_N^2}}{s_0} \right)^n . \tag{8}$$

For $n \geq 1$, we know that $F_{N,n}^{(T)} \geq F_{N,n}^{(R)}$ (with equality holding only if $n = 1$ and in the absence of SGS velocity variations). Because the difference between true and resolved fluxes varies over orders of magnitude, our analysis will focus on the log-10 error process:

$$\varepsilon_{N,n} = \log_{10}\left(F_{N,n}^{(T)} - F_{N,n}^{(R)}\right) . \tag{9}$$

The pdfs of $F_n^{(T)}$ are all positively skewed (Figure 2; this fact is somewhat obscured by the logarithmic scaling). The resolved fluxes are also positively skewed (not shown). Positive skewness results in the mean flux exceeding the most likely value, and provides occasional large magnitude perturbations which are physically realistic and can potentially improve ensemble spread in a forecast setting. It is important that the parameterized error process $10^{\varepsilon_{N,n}}$ respects this skewness. In fact, we show in the next section that the distribution of $\varepsilon_{N,n}$ is approximately Gaussian so the difference between true and resolved fluxes is lognormal and therefore positively skewed. The Gaussianity of the log-10 error process is also of practical importance for generating realizations (particularly in the multivariate setting when the error is considered as a space-time random field). It is interesting to note that other stochastic parameterizations have proposed the use of positively skewed univariate distributions for the stochastic perturbations (Craig and Cohen, 2006; Ollinaho et al., 2017).

3.1 Whole domain analysis

We first study the log-10 error process $\varepsilon_{N,n}$ using wind speeds from across the entire domain. Estimates of the pdfs of $\varepsilon_{N,n}$ for $N = 0.125°$, $0.25°$, $0.5°$, $1°$, $2°$, and $4°$ and $n = 1$, 2 and 3 are shown in Figure 2 (center column). For all values of n, the distributions of $\varepsilon_{N,n}$ are unimodal such that the most likely value increases with averaging scale N: larger averaging scales correspond to larger average differences between true and resolved fluxes. In addition, the values of the log-10 error generally increase for larger n. For the smallest coarsening scales considered, the errors $F^{(T)}_{N,n} - F^{(R)}_{N,n}$ are generally orders of magnitude smaller than the true or resolved fluxes. As the coarsening scale increases, the range of typical error values becomes comparable to the range of typical flux values. In contrast, the distributions of $\varepsilon_{N,n}$ become narrower as N increases. Larger averaging areas result in more averaging of SGS fluctuations and a reduction of the standard deviation of $\varepsilon_{N,n}$, denoted by $\mathrm{std}(\varepsilon_{N,n})$. Consistent with the absence of a spectral gap in SGS velocity variations, the most likely error becomes smaller, but the need for stochastic corrections becomes larger as N is reduced from coarser to finer resolution.

As a measure of the practical importance of accounting for the difference between $F^{(T)}_{N,n}$ and $F^{(R)}_{N,n}$, we consider the probability of the relative error $F^{(T)}_{N,n}/F^{(R)}_{N,n} - 1$ exceeding 10% as a function of the quantiles of resolved flux (Figure 2, right column). Across all resolutions and flux exponents, the probability of the relative error exceeding this threshold decreases with increasing $F^{(R)}_{N,n}$: relative errors are generally larger for smaller fluxes. The probability of exceeding the 10% relative error threshold also increases with increases of both N and n. For a resolution of $N = 1^\circ$ typical of a contemporary general circulation model (GCM), the relative errors in the bottom quartile of fluxes exceed 10% at least 10% of the time for $n = 1$, 35% of the time for $n = 2$, and 60% of the time for $n = 3$.

3.1.1 Distribution of $\varepsilon_{N,n}$ conditioned on resolved fluxes

Developing an empirical parameterization of $\varepsilon_{N,n}$ requires conditioning this quantity on resolved variables. We first study the dependence of the log-10 error process on the resolved flux. Probability distributions of $\varepsilon_{N,n}$ conditioned on $F^{(R)}_{N,n}$ for $N = 1^\circ$ are presented in the left column of Figure 3. The spreads of these conditional distributions around the conditional means represent variations in the difference between true and resolved fluxes that cannot be accounted for by the resolved flux alone. While the spreads of the conditional distributions are similar for all three values of n considered, there are evident differences in the deterministic dependence of $\varepsilon_{1^\circ,n}$ on $F^{(R)}_{1^\circ,n}$. For $n = 1$, the median of $\varepsilon_{1^\circ,1}$ decreases with $F^{(R)}_{1^\circ,1}$: the absolute errors are smaller for larger values of the resolved flux. There is little dependence of $\mathrm{median}\left(\varepsilon_{1^\circ,2}\right)$ on resolved flux for $n = 2$, while for $n = 3$, $\mathrm{median}\left(\varepsilon_{1^\circ,3}\right)$ increases with resolved flux (errors are larger for larger fluxes).

These general features of the conditional dependence of $\mathrm{median}\left(\varepsilon_{N,n}\right)$ on $F^{(R)}_{N,n}$ are found for all coarsening scales considered (Figure 3, central column). Consistent with the behavior of the relative error, the median of $\varepsilon_{N,n}$ increases with coarsening scale: coarser grids result in larger differences between resolved and true fluxes for all values of $F^{(R)}_{N,n}$. In contrast, the spread of the distribution of the log-10 error process decreases with coarsening scale (Figure 3, right column): more averaging results in a narrower distribution. While the interquartile range (iqr) of $\varepsilon_{N,n}$ clearly depends on $F^{(R)}_{N,n}$, this dependence is weaker than that of the median log-10 error (for $n = 1$ and $n = 3$).

To account for the deterministic dependence of $\varepsilon_{N,n}$ on $F^{(R)}_{N,n}$, we construct a polynomial regression model:

$$\varepsilon_{N,n} = \sum_{k=0}^{K} (A_{N,n})_k \left[\log_{10}\left(F^{(R)}_{N,n}\right)\right]^k + \zeta_{N,n}. \tag{10}$$

From inspection, we determined that a reasonable fit is obtained for cubic regression, namely $K = 3$. The results are not strongly sensitive to the value selected for K; qualitatively similar results were obtained for $K = 1$ (not shown). The residual process $\zeta_{N,n}$ is that part of the log-10 error process that cannot be accounted for deterministically by the resolved flux and must be represented stochastically or by further conditioning on other state variables. As we will see in the next section, we can account for some of the variability of $\zeta_{N,n}$ by conditioning on precipitation rate. Values of $(A_{N,n})_{k=1,\ldots,K}$ for $N = 0.25°$ and $N = 1°$ are presented in Table 1.

Inspection of the pdfs of $\zeta_{1°,n}$ conditioned on $F^{(R)}_{1°,n}$ (Figure 4, left column) demonstrates that the regression model Eq. (10) has accounted for most of the deterministic dependence of $\varepsilon_{1°,n}$ on the resolved flux. This fact is also true for the other coarsening scales considered (not shown). Quantile-quantile plots of $\zeta_{N,n}$ against a normal distribution (Figure 4, center column) demonstrate that while the distribution of the residual process $\zeta_{N,n}$ is not exactly Gaussian, deviations from Gaussianity are generally modest. In general, $\zeta_{N,n}$ becomes more non-Gaussian with increasing exponent n.

By construction, the iqr of $\zeta_{N,n}$ conditioned on $F^{(R)}_{N,n}$ is the same as that of $\varepsilon_{N,n}$. Returning to Figure 3, we can see that for each value of the exponent n, changes in N affect the overall value of iqr $(\zeta_{N,n})$ more than the shape of the dependence on $F^{(R)}_{N,n}$. Almost linear behavior in log-log plots of the unconditional iqr $(\zeta_{N,n})$ against N (Figure 4, right column, inset) implies that the iqr of $\zeta_{N,n}$ can be well approximated by a power-law dependence on resolution:

$$\mathrm{iqr}\left(\zeta_{N,n}\right) \simeq \gamma_n N^{\alpha_n}. \tag{11}$$

Values of α_n, γ_n estimated from linear regression of $\ln\left(\mathrm{iqr}\left(\zeta_{N,n}\right)\right)$ on $\ln N$ are presented in Table 2. We note that these parameters depend only weakly on n and that this dependence is not systematic. Rescaling $\zeta_{N,n}$ according to Eq. (11),

$$\hat{\zeta}_{N,n} = \frac{\zeta_{N,n}}{\gamma_n N^{\alpha_n}}, \tag{12}$$

results in the curves of the conditional interquartile range iqr $\left(\hat{\zeta}_{N,n}|F_{N,n}^{(R)}\right)$ largely collapsing on single curves for each n (Figure 4, right column). Agreement among the rescaled iqr value is generally poorest for smaller values of the resolved flux (possibly due to sampling variability since relatively few data fall in this range). Overall, these results indicate that the resolution dependence of the scale N of the explicitly stochastic part of $\varepsilon_{N,n}$ can be well approximated by a power law.

The spatial patterns of the temporal mean and standard deviation of the residuals $\zeta_{N,2}$ are shown in Figure 5 (right column) for $N = 0.25°$ and $N = 1°$. Spatial structure is evident in both fields, although spatial variations are smoother and less pronounced at coarsening scale $N = 1°$ (second and fourth rows). The stochastic variability of the field is weaker at coarsening scale $N = 1°$, as discussed earlier. Because of the short (9-day) duration of the simulation, we are unable to determine to what extent these structures represent true spatial nonhomogeneity in the residual field and to what extent they result from sampling variability.

The results demonstrate that by using velocity information alone, the log-10 error $\varepsilon_{N,n}$ can be approximated by a Gaussian random variable with a mean that depends on the resolved flux $F_{N,n}^{(R)}$ and a variance that is independent of the resolved flux but that varies as a power law in "resolution" N.

3.1.2 Conditioning the residual process $\zeta_{N,n}$ on the precipitation rate

In addition to intrinsic indeterminacy due to the lack of a scale separation in the velocity field, variability of $\zeta_{N,n}$ can result from variations in other physically relevant quantities not accounted for in the regression model Eq. (10). Previous studies have shown a relationship between SGS flux enhancement and convective precipitation (e.g., Redelsperger et al., 2000; Williams, 2001; Zhang et al., 2016), resulting from disorganized mesoscale surface flows associated with moist convection. The 4 km resolution of the model simulation we are considering is at the edge of being convection-permitting. As such, modelled precipitation contains contributions from both resolved and parameterized convection. These resolved and parameterized precipitation fields are available separately, and the relative contribution of both to the total precipitation rate can be determined. Above a threshold precipitation rate of about 0.6 mm/day, all of the modelled precipitation is associated with resolved processes (not shown). Since the strongest relationship between $\zeta_{N,n}$ and precipitation rate P is found above this threshold (Figure 6, left column), in the following calculations we will not distinguish between precipitation derived from resolved or parameterized motions.

The pdf of $\zeta_{N,n}$ conditioned on P for $N = 0.25°$ shows a relatively weak dependence for $P \lesssim 0.1$ mm/day and a steady increase with P above this value (Figure 6, left column). Such a transition indicates a systematic contribution to SGS flux enhancements of disorganized velocity fluctuations associated with deep convection. The transition from relatively weak to strong dependence moves to larger values of P and becomes less abrupt for larger coarsening scales. On larger coarsening scales, the sharpness of the transition is smoothed out because the averaging areas will contain regions of larger and smaller precipitation rates. The slope of the dependence of median($\zeta_{N,2}$) on P for large P is about the same for all coarsening scales. The breadth of the conditional distributions systematically decreases with increasing coarsening scale. Again, more averaging results in smaller fluctuations around the mean.

In the same way that we obtained $\zeta_{N,n}$ as the residual of a regression of $\varepsilon_{N,n}$ on $\log_{10}\left(F^{(R)}_{N,n}\right)$, we represent the deterministic dependence of $\zeta_{N,n}$ on P through a regression model on $P^{1/4}$:

$$\zeta_{N,n} = \sum_{l=0}^{L} \left(B_{N,n}\right)_l P^{l/4} + \psi_{N,n}. \tag{13}$$

The fourth-root transformation of P was determined empirically through inspection of the dependence of the median of $\zeta_{N,n}$ conditional on P (not shown). In the calculation of these regression coefficients, values of $P = 0$ mm/day were neglected (since they represent a point probability mass at this particular value). Plots of the regression model with $L = 4$ for $N = 0.25^\circ, 1^\circ$, and 2° are shown in Figure 6. The residual $\psi_{N,n}$ is that part of the error process that depends neither on the resolved flux nor on the precipitation rate and that must be represented as explicitly stochastic in the absence of further conditioning. The sequential conditioning of $\varepsilon_{N,n}$ first on $F^{(R)}_{N,n}$ and then on P is justified by the absence of a strong statistical relationship between resolved flux and precipitation rate (not shown). Values of the coefficients $(B_{N,n})$ for $N = 0.25^\circ$ and $N = 1^\circ$ are presented in Table 1.

Quantile-quantile plots of $\psi_{N,n}$ against a normal distribution (Figure 6, center column) show that except for large values of the coarsening scale and exponent the distribution of $\psi_{N,n}$ does not deviate substantially from Gaussian. The deviations that are present are somewhat larger than we found for $\zeta_{N,n}$, perhaps because of the simple form of the regression Eq. (13) does not capture all of the deterministic dependence of $\zeta_{N,n}$ on P.

As was the case for $\zeta_{N,n}$, the dependence of the unconditional iqr of $\psi_{N,n}$ on coarsening scale N can be approximated as a power law:

$$\mathrm{iqr}\left(\psi_{N,n}\right) \simeq \mu_n N^{\lambda_n} \tag{14}$$

(Figure 6, right column, inset). Estimated values of μ_n and λ_n (obtained from regressing

$\ln(\mathrm{iqr}(\psi_{N,n}))$ on $\ln N$) are given in Table 2. As was the case with γ_n and α_n, the dependence of μ_n and λ_n on n is weak and not systematic. Interquartile ranges of $\psi_{N,n}$ rescaled by this power law,

$$\hat{\psi}_{N,n} = \frac{\psi_{N,n}}{\mu_n N^{\lambda_n}}, \tag{15}$$

and conditioned on P, collapse to a reasonable approximation on a single curve for each n (Figure 6, right column). While $\mathrm{iqr}\left(\hat{\psi}_{N,n}|P\right)$ does show systematic dependence on P, the variations around a value of 1 are sufficiently small that it is a reasonable first approximation to model this quantity as a constant. Furthermore, this figure clearly shows that the iqr of $\hat{\psi}_{N,n}$ conditioned on P depends only weakly on the value of the exponent n. The fact that the values of μ_n are smaller than γ_n is a result of the reduction of the spread in $\psi_{N,n}$ relative to $\zeta_{N,n}$ because of the further conditioning on P.

Maps of the time-mean and standard deviation of the residuals $\psi_{N,2}$ are shown in Figure 5 for $N = 0.25°$ and $N = 1°$. The statistics of $\psi_{N,n}$ show much less spatial structure than of the ones of $\zeta_{N,n}$, particularly for the standard deviation and at coarser graining-scale ($N = 1°$). This fact indicates that much of the spatial structure of $\zeta_{N,n}$ is inherited from the precipitation field. The most pronounced features of mean($\psi_{N,n}$) are the negative values in the Indian Ocean, east of Australia, and on the eastern boundary of the domain, and the positive values around the maritime continent and the Northwestern coast of Australia. Overall, the use of the precipitation field significantly improves the spatial homogeneity of the residuals.

Using data from across the analysis domain, we conclude that the difference between the true and resolved fluxes can be modelled as a lognormal distributed variable, with a median that depends on the value of the resolved flux and the precipitation rate and an iqr that is to a first approximation independent of $F^{(R)}_{N,n}$, P, and n, and that depends on the coarsening scale through a simple power law.

3.1.3 Spatial and temporal correlation structure of $\zeta_{N,n}$ and $\psi_{N,n}$

So far, we have considered only pointwise (marginal) statistics of the error $\varepsilon_{N,n}$ and the residuals $\zeta_{N,n}$ and $\psi_{N,n}$. Since the residuals at nearby times and spatial locations may not be independent, it is appropriate to treat $\zeta_{N,n}$ and $\psi_{N,n}$ as space-time random processes. Basic descriptive characterizations of the temporal and spatial structure of these processes are provided by autocorrelation functions in space and time.

Plots of the temporal autocorrelation functions (acf) of $\zeta_{N,2}$ for lags of up to 48 h, composited across all points in the model domain, show that on average the memory of the residual process increases with coarsening scale (Figure 7, upper left). For example, the value of the acf falls below e^{-1} in about 3 hours for $N = 0.25°$ and 7 hours for $N = 2°$. On top of the overall decay of correlations, the acf shows a clear diurnal periodicity.

Conditioning $\zeta_{N,n}$ on the precipitation rate reduces both the autocorrelation decay timescale and the amplitude of the diurnal cycle in the spatial composite acf of the residuals $\psi_{N,n}$ (Figure 7, upper right). These changes are consistent with having accounted deterministically for the contribution to SGS velocity variations from organized convective motion associated with precipitation.

Temporal autocorrelation functions at individual spatial locations display considerable variation around the composites shown in the upper panels of Figure 7. The lower panels of this figure show the $\zeta_{N,2}$ and $\psi_{N,2}$ acf composites for $N = 0.25°$ and $N = 1°$, as well as the interdecile range across all spatial locations. While the spatial spread of the acf of $\psi_{N,2}$ is slightly smaller than that of $\zeta_{N,2}$, both acfs show substantial spatial variations (although the confidence intervals corresponding to a null hypothesis of zero correlation coefficient are broad because of the relatively few degrees of freedom, particularly if the serial dependence of the time series is accounted for). The acf decay length scales increase slightly with n (not shown).

Compositions of the spatial correlation function of $\zeta_{N,2}$ for $N = 0.25°$, $1°$, and $2°$ (Figure 8, upper row) were obtained by averaging the estimated spatial correlation functions centred at a range of different base locations across the domain. The spatial correlation functions are evidently anisotropic, with decay lengthscales in the zonal that are larger than those in the meridional. This anisotropy and the values of the correlation length scales tend to increase at coarser averaging scales. Similar behavior is seen for the spatial correlation function of $\psi_{N,2}$ (Figure 8, second row), although the correlation length scales of $\psi_{N,2}$ are smaller than those of $\zeta_{N,2}$. As was the case for the temporal dependence structure, we find that removing the deterministic dependence on precipitation results in a residual field that is more local in space. While spatial correlation scales increase slightly with increases in n (not shown), results similar to those shown in Figure 8 are found for $n = 1$ and $n = 3$.

We now consider variations of the spatial correlation function across the domain. For each base point $\mathbf{x}$, the spatial autocorrelation function results in a different map. Since a complete characterization of the spatial correlation structures of $\zeta_{N,n}$ and $\psi_{N,n}$ is therefore not practical, we adopt the following approach. For $N = 1^\circ$, the spatial correlation field across the entire domain is computed at each of a set of base points on a coarse $4^\circ \times 4^\circ$ grid. Around each base point, a contour is drawn corresponding to a squared correlation value of 0.5 for the spatial correlation field with that base point. Within such a contour around any base point, the squared spatial correlation values are larger than 0.5. The resulting maps (Figure 8, third and fourth rows) give some evidence of variations of the spatial correlation functions of $\zeta_{1^\circ,2}$ and $\psi_{1^\circ,2}$ across the domain. In particular, regions of relatively large correlation lengthscales for $\zeta_{1^\circ,2}$ are found in the Arabian Sea and Bay of Bengal, as well as in a band extending from the Horn of Africa to west of Australia. Similar features are seen in the spatial correlation structure of $\psi_{1^\circ,2}$, although the variations across the domain are less pronounced, and no atypical structure is seen in the Bay of Bengal. Broadly similar behavior is found for different coarsening scales N and flux exponents n (not shown).

From the perspective of developing stochastic parameterizations of SGS flux enhancements, in the following section we propose a statistical model that embeds the pointwise and space-time characteristics of $\varepsilon_{N,n}$ presented above. This statistical framework provides a complementary quantification of features described above (spatio-temporal dynamics and marginal distributions) and also allows generation of realistic space-time samples of the SGS flux enhancement. A Gaussian process is used here to model the space-time residual processes $\zeta_{N,n}$ and $\psi_{N,n}$. Gaussian processes are tractable stochastic processes in a multidimensional context (space-time in our case) and the choice of Gaussian marginal distribution is supported by Figures 4 and 6. Since Gaussian processes are characterized by their first and second moments only and the mean of the residuals does not need to be accounted for, we only consider the specification of the space-time covariance structure in the following.

3.1.4 Fitting spatio-temporal covariance structures

In order to quantify the spatio-temporal dynamics and the spatial anisotropy observed in Figures 7 and 8, as well as the dependence on the coarsening scales, parametric anisotropic spatio-temporal covariance structures have been fit locally for each of the two residual processes $\zeta_{N,n}$ and $\psi_{N,n}$ for various coarsening scales N.

Spatio-temporal covariance model The ellipsoidal contour lines present in the observed spatial correlation (Figure 8) suggest the use of an anisotropic correlation model with different dependence in the meridional and zonal directions determined respectively by parameters θ_1 and θ_2. For simplicity, we assume that the semimajor and semiminor axes of the correlation align with the zonal and meridional directions. We also include temporal dependence scale θ_3 in the correlation structure. The commonly used power exponential correlation is considered with a 3D-anisotropic distance for the space-time coordinates, and is fit to the data:

$$K(l, l', t, t') = \sigma \exp(-d(l, l', t, t')^{\gamma}) + \delta \mathbb{I}_{l=l', t=t'} \tag{16}$$

with the space-time 3D-distance $d(l,l',t,t') = \sqrt{\left(\frac{x-x'}{\theta_1}\right)^2 + \left(\frac{y-y'}{\theta_2}\right)^2 + \left(\frac{t-t'}{\theta_3}\right)^2}$. The parameters θ_1, θ_2, θ_3, σ, γ, and α are positive real numbers estimated by a least-squares method described below, and x, y and t respectively represent the latitude, longitude and temporal coordinates. For more flexibility in the correlation decay, the distance exponent $\gamma \in]0,2]$ is estimated as a parameter of the covariance model. A nugget $\delta > 0$ is added to the covariance to capture local variance that is not accounted for in the parametric exponential part of the model Eqn. (16).

Estimation of the local covariance structure A moving-window framework is used to estimate the spatial variations of the covariance structure (as in Haas, 1990; Kuusela and Stein, 2017). More specifically, the whole domain is sub-divided into smaller regions of size 400 km$\times$400 km. Within each window, stationarity is assumed, and the proposed covariance model Eq. (16) is fit independently to the residuals $\zeta_{N,n}$ and $\psi_{N,n}$. In order to ensure continuity, the windows overlap by 40 km.

Figures 9a, 9b, 9c, 10, and 11 respectively show maps of the estimated values of the parameters θ_1, θ_2, θ_3, γ, and δ. Parameters are depicted for both processes $\zeta_{N,n}$ and $\psi_{N,n}$, and for the two coarsening scales $N = 0.25°$ and $N = 1°$. Spatially heterogeneous structure is evident in the maps of the estimates of θ_1, θ_2, and θ_3, as expected given the large size of the domain and the limited temporal duration of the simulation. As observed in the empirical correlation structure (Figure 8), the parameters estimated from $\psi_{N,n}$ exhibit more homogeneity across the domain, shorter spatial length scales, and less anisotropy (similar ranges of values for θ_1 and θ_2) than those of $\zeta_{N,n}$. Again, we see that the precipitation field explains much of the spatio-temporal structure of the error process $\varepsilon_{N,n}$. Zonal correlation elongation (larger values of θ_2 than θ_1) is evident for both coarsening scales considered. This anisotropy tends to be slightly stronger at coarser scales than finer ones. As indicated by the composite spatial and temporal correlation structures (Figures 7 and 8), the spatial and temporal scales θ_1, θ_2 and θ_3 are longer for coarser averaging scales . The larger scales of space and time variations of the error process $\varepsilon_{N,n}$ when N is large results from the averaging out of smaller scales.

Figure 10 shows estimates of the parameter γ that determines the smoothness of the field. This parameter also shows evidence of spatial heterogeneity. The parameter value is larger when the precipitation is not regressed out, which is another indication that the precipitation results in localized spatial structure in the error process $\epsilon_{N,n}$, resulting in a less structured and less smooth residual process. In contrast with the other parameters considered, conditioning on precipitation and varying the coarsening scales have less influence on the intensity of this parameter.

In Figure 11, the ratio of the nugget δ and the variance of the error $\epsilon_{N,n}$ shows that the nugget parameter tends to have slightly less importance in the overall variance when the precipitation is regressed out. In that case, the residual fields present less unexplained information that cannot be captured by the proposed parametric covariance with a single decay scale in each direction.

Some regions of the maps display atypical behaviors, such as the Arabian Sea and the Southeastern part of the Indian Ocean, where the correlation structure is not influenced by the precipitation field. These behaviors are expected because precipitation was almost absent in those regions during the simulation time.

Simulating the error process In order to assess the quality of the statistical models we have developed for $\varepsilon_{N,n}$, we generated samples of $\zeta_{1^\circ,2}$ and $\psi_{1^\circ,2}$ from a Gaussian distribution with zero mean and a covariance specified by the estimated version of (16). The choice of a Gaussian distribution is justified by the quantile-quantile plots shown in Figures 4 and 6. Samples of the error process $\varepsilon_{1^\circ,2}$ were then constructed via Eq. (10) and (13).

Figure 12 shows sample time series of the "true" error process and its simulated samples at an arbitrary location for both models Eq. (10) and (13). While both models capture the range of variations of $\varepsilon_{1^\circ,2}$ reasonably well, structure is evident in the log10-error process that is captured by Eq. (13) but not by Eq. (10). In particular, the large sustained increase in $\varepsilon_{1^\circ,2}$ starting on 11 April is captured by the model including precipitation rate as a regressor, but not by the model based only on the resolved flux. The benefit of conditioning on P is evident from this result. Consistent with the results of Section 3.1.2, the spread of the ensemble of simulations around $\varepsilon_{1^\circ,2}$ is smaller for the model Eq. (13) than for Eq. (10): including precipitation as a regressor improves the resolution (sharpness) of the ensemble forecast. The statistical consistency between the observed error process and its samples is further explored through rank histograms at a single location in Figure 12. When a perfect match exists between the distributions of observations and samples, the rank histogram is expected to be uniform. We observe that the use of precipitation in the regression (lower panel) provides a better statistical calibration than does the regression based on the resolved fluxes only (upper panel). The fact that the rank histogram is not flat for either model reflects that the statistical model does not exactly fit the statistics of either residual process.

The simulated time series also capture the true temporal autocorrelation structure of $\varepsilon_{1^\circ,2}$ (Figure 13). The broader range of acf curves for $\varepsilon_{1^\circ,2}$ constructed from realizations of $\zeta_{1^\circ,2}$ than from realizations of $\psi_{1^\circ,2}$ is consistent with the latter being more constrained by resolved variables (which are the same among all realizations). We note that the correlation values for the shortest time lags tend to be underestimated by the proposed models.

Figure 14 depicts maps of the mean square error (MSE) between the "true" error process $\epsilon_{1^\circ,2}$ and the simulated samples. The overall magnitude of the MSE is smaller for the model Eq. (13) than for Eq. (10). Moreover, the former model shows a weaker spatial structure due to the use of the precipitation information. The total MSE is decomposed into its squared bias and centered MSE components to assess the respective contributions of the mean features and of the fluctuations of the fields (Taylor, 2001). The squared bias contribution is significantly less than the difference in variability, indicating that both proposed models capture the global mean of the error process reasonably well. However, the bias term exhibits more spatial structure than does the centered MSE, indicating that the proposed models capture well the structure of the stochastic variability of the error process. Including the precipitation field as a predictor improves the ability of the statistical model to account for the mean and fluctuations over the domain, particularly accounting for much of the structure in the squared bias term. Including further predictors might be able to reduce the squared bias term further, particularly around the Arabian Sea and the Southeast Indian Ocean.

3.2 Local domain analysis

Because of the relatively short duration of the simulation we are considering, some of the apparent spatial non-stationarity in the temporal and spatial autocorrelation functions may result from sampling variability. For example, an animation of the surface wind field over the simulation period (not shown) shows the migration of a strong cyclone from the Arabian Sea to the Bay of Bengal; such a circulation feature is not observed to occur elsewhere in the domain in this nine-day period. Nevertheless, the potential for spatially non-stationary structure motivates repeating the analysis of the relationships between $\varepsilon_{N,n}$, $F_{N,n}^{(R)}$, and P in different subregions of the model domain. Furthermore, previous empirical studies of SGS flux enhancement have considered either observations or model simulations in spatial domains much smaller than the one we study. We therefore re-examine our analysis in model subdomains.

In order to examine regional variations in $\zeta_{N,n}$ and $\psi_{N,n}$, regression Eqs. (10) and (13) were fit separately on three subdomains (the Western Pacific, Arabian Sea, Southern Indian Ocean) depicted on Figure 1. Similarities are evident among the statistical properties of the residuals $\zeta_{N,n}$ and $\psi_{N,n}$ in the subregions, in terms of marginal distributions (Figure 15), spatial correlation (Figure 16), and temporal structure (not shown). For the most part, we find that coarser averaging scales result in larger departures of the residuals from normality.

Consistent with the global analysis, spatial correlations at coarser resolutions appear stronger than the ones at finer resolutions.

However, we also note differences in statistical features between these different regions. As previously observed, the Arabian Sea has atypical characteristics, especially in terms of spatial and temporal dynamics. The spatial correlation scales of $\zeta_{N,n}$ and $\psi_{N,n}$ are longer than in the other two subdomains. The absence of precipitation in this area during the period of the model simulation (Figure 1) is likely responsible for this variant behavior.

Given the short temporal amount of data, it is difficult to distinguish sampling variability from true spatial heterogeneity in the fields. However, the very low precipitation rates over large parts of the model domain (lower than long-term climatological values) do indicate that the limited temporal duration of the simulation is an important factor for the spatial structure.

4 Discussion and Conclusions

In this study, we have considered the empirical parameterization of the subgrid-scale velocity enhancement of spatially-averaged sea surface fluxes in weather and climate models. Using output from a relatively high-resolution, convection-permitting model simulation, we have shown that the SGS flux enhancement is not a deterministic function of the resolved state. Considering a range of different coarsening scales and flux exponents, and regressing the differences between the true and resolved fluxes on (nonlinearly transformed) resolved flux and precipitation rates, we have obtained residual fields characterizing the essentially stochastic nature of the SGS flux enhancement. The final model that we propose takes the lognormal form

$$F_{N,n}^{(T)} = F_{N,n}^{(R)} + 10^{\varepsilon_{N,n}} \tag{17}$$

with

$$\varepsilon_{N,n} = \sum_{k=0}^{K} (A_{N,n})_k \left[\log_{10}\left(F_{N,n}^{(R)}\right)\right]^k + \sum_{l=0}^{L} (B_{N,n})_l \, P^{l/4} + \psi_{N,n}, \tag{18}$$

where $\psi_{N,n}$ is a Gaussian space-time field with a variance that scales as a power law of the coarsening scale N. The residual field $\psi_{N,n}$ has been shown to be correlated in space and time, such that increases in N result in increases of both the spatial and temporal correlation decay scales. Modelling the statistics of $\psi_{N,n}$ as a function of coarsening scale N is an important step in allowing this parameterization to be scale aware.

Space-time Gaussian process models have been fit through the estimation of parametric covariances. In order to account for potential spatial inhomogeneity, covariances were fit in a set of overlapping moving windows. This estimation provides insights into the space-time characteristics of the residual fields: we were able to better quantify the spatial and temporal correlation ranges across coarsening scales and across the domain, and to assess the spatial anisotropy of the fields. Furthermore, this framework provides a space-time sampling distribution that could be used in future implementations.

In this study we have treated a 4 km simulation as 'truth', since observational data do not exist at a high-enough resolution over such a large spatio-temporal domain. Because of the realism of the simulation (Holloway et al., 2012, 2013, 2015), the results of this study are a good first indication of the statistics of sub-grid scale fluxes. Furthermore, the relatively large precipitation rates which have the strongest deterministic relationship with the error process $\varepsilon_{N,n}$ (Figure 6) are associated with resolved dynamics rather than parameterized convection. Nevertheless, details of the proposed model, such as the precise values of the regression coefficients, could change if a different model simulation were coarse grained. A further limitation of the study is the restricted spatial domain and length of the simulation: the statistics of SGS fluxes could vary depending on region of the globe and meteorological conditions. A follow up study is planned which will apply these techniques to a different dataset that covers a larger space-time domain to assess the generality of the parameterization.

Because the 4 km resolution of the model is still relatively coarse and the model equations are Reynolds averaged, this analysis does not account for those contributions to SGS velocity flux enhancement that are associated with the model's existing gustiness parameterization (Walters et al., 2017). Since the main goal of this analysis is to demonstrate the importance of explicitly accounting for the stochasticity of the parameterization, the fact that not all SGS velocity variations are accounted for is not a critical limitation. We expect that if output from higher-resolution observations or model output were used, the magnitude of stochastic fluctuations around the deterministic parameterization would increase.

To construct an empirical parameterization of SGS flux enhancements, we have used the resolved flux and precipitation rate as deterministic predictors of the error process $\varepsilon_{N,n}$. It is possible that $\varepsilon_{N,n}$ may depend on other modelled quantities and that by including these in the regression model we would further reduce the stochasticity of our parameterization. For example, the dependence of the exchange coefficient $c_x(s)$ on sea surface temperature (through, e.g., changes in near-surface stability) has been neglected. Furthermore, the dependence of the error process on resolved variables may depend on the specific parameterization schemes used in the model. Further investigation of these questions is an interesting direction of future study.

Following standard practice (e.g., Williams, 2001), we have neglected the dependence between variations in air density, wind speed, and air-sea concentration difference that can affect area-averaged fluxes (Eq. 1). Furthermore, our parameterization is based on the general resolved flux rather than specifically the surface heat flux (through the free convective scale) as in standard gustiness parameterizations (e.g., Beljaars, 1994; Mahrt and Sun, 1995; Williams, 2001). While our approach has the advantage of not requiring an iterative calculation of fluxes, it is further removed from the basic boundary-layer physics used in justifying expressions such as Eq. (5). Moreover, many choices regarding the structure of the statistical model (such as the fourth-root transformation of precipitation rate, and the number of terms K and L in the resolved flux and precipitation rate regressions) were determined through experimentation rather than systematic optimization. A more systematic and objective approach to optimizing the values of these quantities should be considered in future research. Similarly, the consideration of alternative formulations of the statistical model Eq. (18), in terms of both the predictor fields chosen and the model architecture, is an interesting direction of future study. The development of physically based parameterizations (such as that of Williams, 2001) rather than empirically based ones is also a potentially important direction of research. Finally, repeating this analysis with longer time series on a larger spatial domain would allow a better determination of spatial and temporal heterogeneities in the statistics of SGS flux enhancements.

The goal of this study has been to demonstrate (via a systematic coarse-graining analysis) the fundamentally stochastic nature of the dependence of area-averaged fluxes on the resolved state and to characterize the structure of the stochastic space-time fields needed to parameterize this dependence. This analysis demonstrated the existence of spatial and temporal dependence in the stochastic parameterization and provided empirical evidence for the inclusion of such correlations in stochastic parameterization schemes (as opposed to treating this structure as a pragmatic solution to improve ensemble spread, see, e.g., Leutbecher et al., 2017). This analysis also highlighted the resolution dependency of such spatio-temporal correlations, which is not currently included in operational stochastic schemes. A future study will report on the result of implementing and testing such a stochastic sea surface flux parameterization in weather and climate models.

Acknowledgments

Data from the Cascade project is available on request from the NERC Centre for Environmental Data Analysis (CEDA). This research started in a working group supported by the

Statistical and Mathematical Sciences Institute (SAMSI). AHM acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC), and thanks SAMSI for hosting him in the autumn of 2017. The effort of Julie Bessac is based in part on work supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357. The research of HMC was supported by NERC grant number NE/P018238/1. We thank Aneesh Subramanian and two anonymous reviewers for their helpful comments.

References

Beljaars, A. C. (1994). The parameterization of surface fluxes in large-scale models under free convection. *Q. J. R. Meteorol. Soc.*, 121:225–270.

Berner, J., Achatz, U., Batté, L., Bengtsson, L., de la Cámara, A., Christensen, H. M., Colangeli, M., Coleman, D. R., Crommelin, D., Dolaptchiev, S., Franzke, C. L., Friedrichs, P., Imkeller, P., Järvinen, H., Juricke, S., Kitsios, V., Lott, F., Lucarini, V., Mahajian, S., Palmer, T. N., Penland, C., Sakradzija, M., von Storch, J.-S., Weisheimer, A., Weniger, M., Williams, P. D., and Yano, J.-I. (2017). Stochastic parameterization: Toward a new fiew of weather and climate models. *Bull. Amer. Meteo. Soc.*, 3:565–588.

Cakmur, R., Miller, R., and Torres, O. (2004). Incorporating the effect of small-scale circulations upon dust emission in an atmospheric general circulation model. *J. Geophys. Res.*, 109:D07201.

Capps, S. B. and Zender, C. S. (2008). Observed and CAM3 GCM sea surface wind speed distributions: Characterization, comparison, and bias reduction. *J. Climate*, 21:6569–6585.

Christensen, H. M. (2018). Constraining stochastic parametrisations using high-resolution model simulations. *Q. J. Roy. Meteor. Soc.*, Submitted.

Christensen, H. M., Berner, J., Coleman, D., and Palmer, T. N. (2017). Stochastic parametrisation and El Niño-Southern Oscillation. *J. Climate*, 30(1):17–38.

Craig, G. C. and Cohen, B. G. (2006). Fluctuations in an equilibrium convective ensemble. part i: Theoretical formulation. 63(8):1996–2004.

Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N. (2005). A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. Roy. Meteor. Soc.*, 131(608):1759–1782.

Drennan, W. (2006). On parameterizations of air-sea fluxes. In Perrie, W., editor, *Atmosphere-Ocean Interactions*, volume 2, pages 1–33. WIT Press.

Edson, J. B. (2008). Review of air-sea transfer processes. In *ECMWF Workshop on Ocean-Atmosphere Interactions*, pages 7–24.

Garbe, C. S., Rutgersson, A., Boutin, J., de Leeuw, G., Delille, B., Fairall, C. W., Gruber, N., Hare, J., Ho, D. T., Johnson, M. T., Nightingale, P. D., Pettersson, H., Piskozub, J., Sahlée, E., tin Tsai, W., Ward, B., Woolf, D. K., and Zappa, C. J. (2014). Transfer across the air-sea interface. In Liss, P. and Johnson, M., editors, *Ocean-Atmosphere Interaction*, pages 55–112, doi 10.1007/978–3–642–25643–1_2.

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

Gregory, D. and Rowntree, P. R. (1990). A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Weath. Rev.*, 118(7):1483–1506.

Haas, T. C. (1990). Lognormal and moving window methods of estimating acid deposition. *Journal of the American Statistical Association*, 85(412):950–963.

Holloway, C. E., Woolnough, S. J., and Lister, G. M. S. (2012). Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study. *Q. J. Roy. Meteor. Soc.*, 138(668):1692–1708.

Holloway, C. E., Woolnough, S. J., and 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., and 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.

Kuusela, M. and Stein, M. L. (2017). Locally stationary spatio-temporal interpolation of Argo profiling float data. *arXiv preprint arXiv:1711.00460*.

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R., Goddard, J., Haiden, T., Hogan, R., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P., Subramanian, A., Vitart, F., Wedi, N., and Weisheimer, A. (2017). Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Q. J. Roy. Meteor. Soc.*, 143(707):2315–2339.

Love, B. S., Matthews, A. J., and Lister, G. M. S. (2011). The diurnal cycle of precipitation over the Maritime Continent in a high-resolution atmospheric model. *Q. J. Roy. Meteor. Soc.*, 137(657):934–947.

Mahrt, L. and Sun, J. (1995). The subgrid velocity scale in the bulk aerodynamic relationship for spatially averaged scalar fluxes. *Mon. Weath. Rev.*, 123:3032–3041.

Ollinaho, P., Lock, S.-J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R. M., Haiden, T., Hogan, R. J., and Sandu, I. (2017). Towards process-level representation of model uncertainties: Stochastically perturbed parametrisations in the ECMWF ensemble. 143(702):408–422.

Pearson, K. J., Hogan, R. J., Allan, R. P., Lister, G. M. S., and 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: Atmospheres*, 115(20):1–11.

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

Ridley, D., Heald, C., Pierce, J., and Evans, M. (2013). Toward resolution-independent dust emissions in global models:Impacts on the seasonal and spatial distribution of dust. *Geophys. Res. Lett.*, 40:2873–2877.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7):7183–7192.

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

Walters, D., Baran, A., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujie, M., Whitall, M., Williams, K., and Zerroukat, M. (2017). The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES global land configurations. *Geosci. Model Dev. Discuss*, pages doi:10.5194/gmd–2017–291.

Williams, A. G. (2001). A physically based parametrization for surface flux enhancement by gustiness effects in dry and precipitating convection. *Q. J. R. Meteorol. Soc.*, 127:469–491.

Williams, P. D. (2012). Climatic impacts of stochastic fluctuations in air-sea fluxes. *Geophys. Res. Lett.*, 39:doi:10.1029/2012GL051813.

Yang, C., Christensen, H. M., Corti, S., von Hardenberg, J., and Davini, P. (2019). The impact of stochastic physics on the El Niño-Southern Oscillation in the EC-Earth coupled model. *Clim. Dynam.* sub. jud.

Zeng, X., Zhang, Q., Johnson, D., and Tao, W.-K. (2002). Parameterization of wind gustiness for the computation of ocean surface fluxes at different spatial scales. *Mon. Weath. Rev.*, 130:2125–2133.

Zhang, K., Zhao, C., Wan, H., Qian, Y., Easter, R. C., Gahn, S. J., Sakaguchi, K., and Hu, X. (2016). Quantifying the impact of sub-grid surface wind variability on sea salt and dust emissions in CAM5. *Geosci. Model Dev.*, 9:607–632.

n	N	$(A_{N,n})_0$	$(A_{N,n})_1$	$(A_{N,n})_2$	$(A_{N,n})_3$	$(B_{N,n})_0$	$(B_{N,n})_1$	$(B_{N,n})_2$	$(B_{N,n})_3$	$(B_{N,n})_4$
1	0.25°	-1.37	-0.93	-0.09	-0.12	-0.14	0.77	-0.08	-0.02	0.004
2	0.25°	-0.75	0.03	-0.015	-0.03	-0.14	0.78	-0.11	-0.01	0.003
3	0.25°	-0.37	0.31	0.12	-0.01	-0.14	0.80	-0.12	-0.008	0.003
1	1.0°	-0.65	-0.72	-0.23	-0.20	-0.28	0.73	-0.22	0.04	-0.003
2	1.0°	-0.06	0.10	-0.01	-0.05	-0.29	0.75	-0.25	0.05	-0.003
3	1.0°	0.37	0.31	0.02	-0.02	-0.29	0.78	-0.26	0.05	-0.003

Table 1: Estimated regression coefficients for the models Eqn. (10) and Eqn. (13) for coarsening scales $N = 0.25°$ and $N = 1°$.

n	γ_n	α_n	μ_n	λ_n
1	0.63	-0.21	0.45	-0.34
2	0.59	-0.19	0.42	-0.31
3	0.60	-0.18	0.43	-0.31

Table 2: Coefficients of the scaling relationships Eq. (11) and (15) relating the interquartile range of $\zeta_{N,n}$ (γ_n, α_n) and $\psi_{N,n}$ (μ_n, λ_n) to coarsening resolution N.

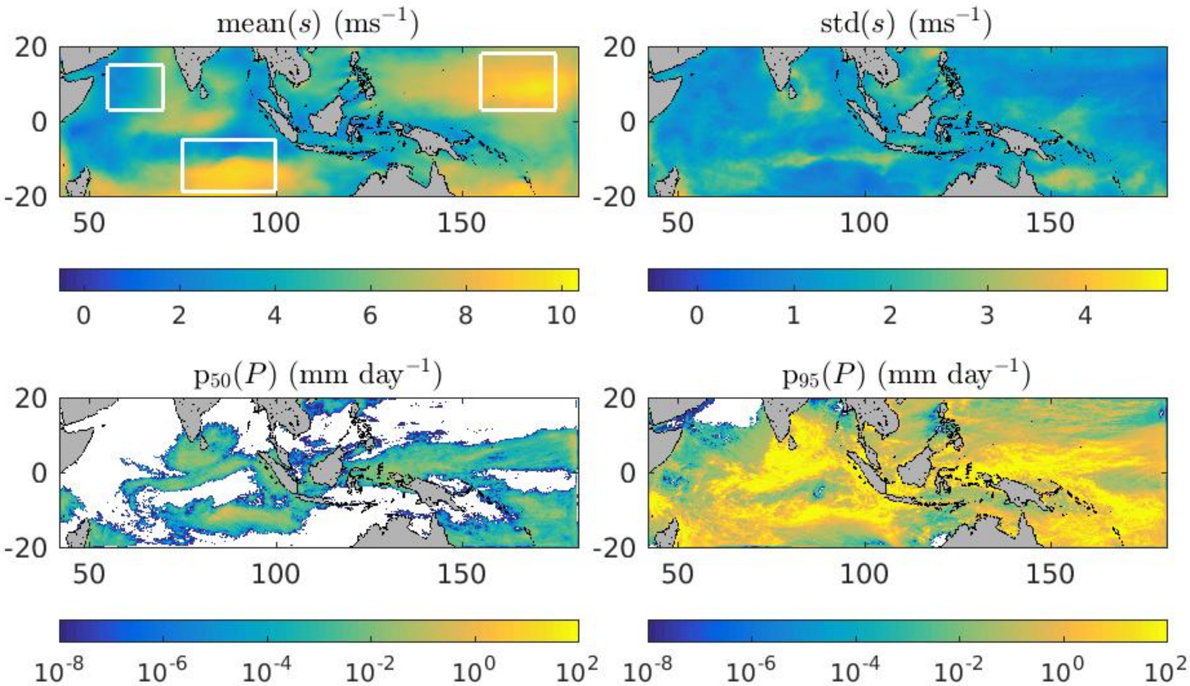


Figure 1: Mean and standard deviation of the simulated wind speed (upper panels) and 50th and 95th percentiles of precipitation rate (lower panels) at the base 4 km $\times$ 4 km resolution of the simulation. White areas in the precipitation plots correspond to zero precipitation rates. The white boxes in the mean wind-speed panel delimit the subregions considered in Section 3.2.

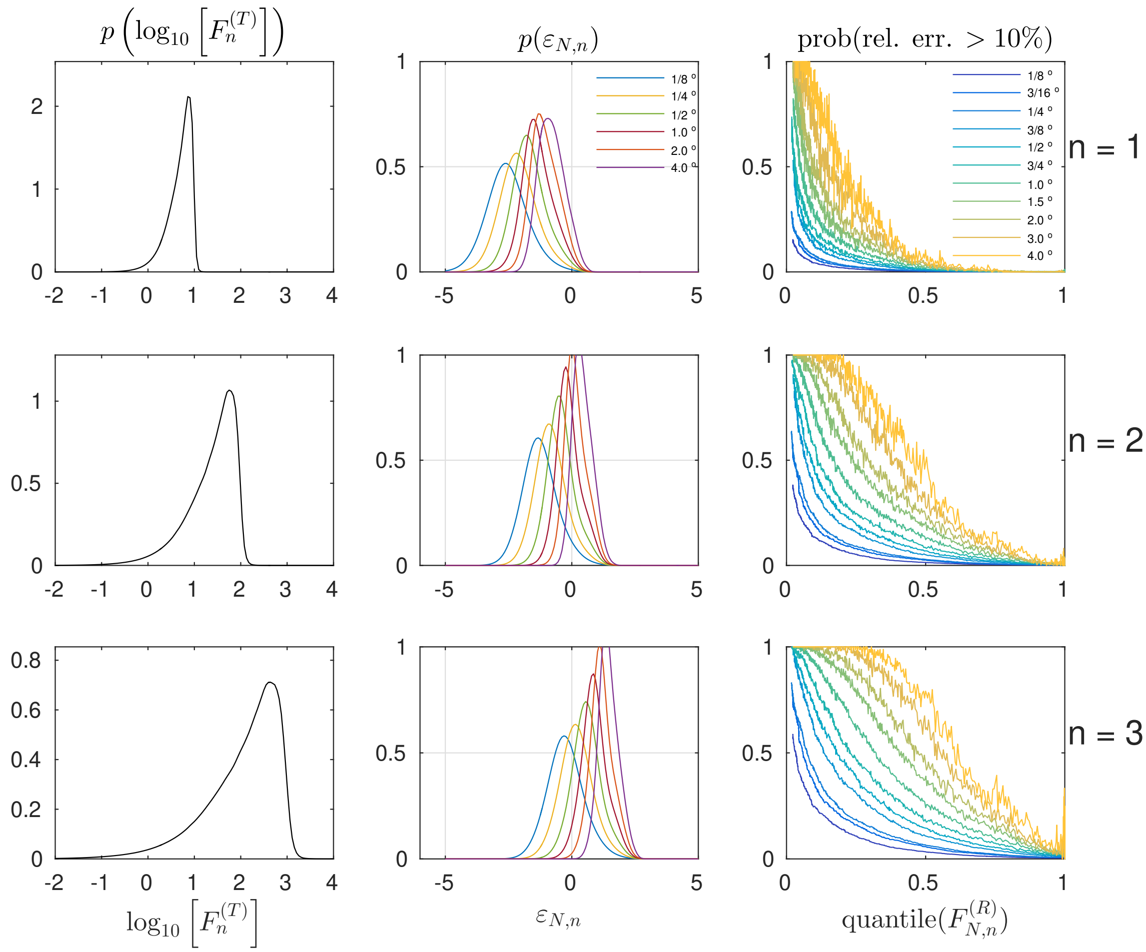


Figure 2: Left column: estimated pdfs of the true flux $F_n^{(T)}$ without coarse-graining. Center column: estimated pdfs of the log-10 error process $\varepsilon_{N,n}$ for a range of averaging scales N. Right column: fraction of times that the relative error exceeds 10% as a function of resolved flux quantile for a range of averaging scales N. For each column, results are shown for the exponents $n = 1, 2$ and 3.

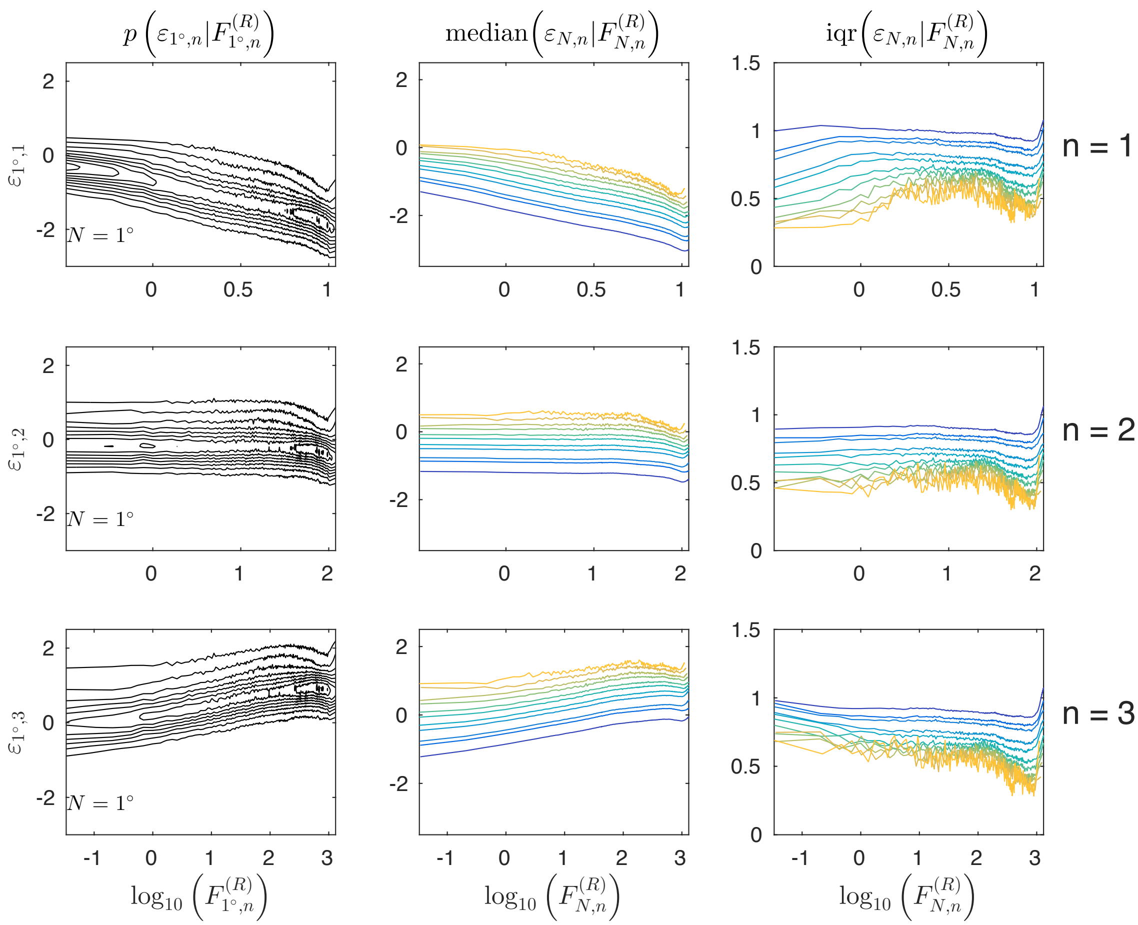


Figure 3: Statistics of the log-10 error process $\varepsilon_{N,n}$ conditioned on the resolved flux $F^{(R)}_{N,n}$ for $n = 1$ (upper row), $n = 2$ (middle row), and $n = 3$ (lower row). Left column: kernel density estimate of the probability density function of $\varepsilon_{1^\circ,n}$ conditional on $F^{(R)}_{1^\circ,n}$ for the coarsening scale $N = 1^\circ$. Center column: medians of $\varepsilon_{N,n}$ conditioned on $F^{(R)}_{N,n}$ for a range of coarsening scales N. Right column: interquartile ranges of $\varepsilon_{N,n}$ conditioned on $F^{(R)}_{N,n}$ for a range of coarsening scales N. The color scheme indicating N is as in the right column of Figure 2.

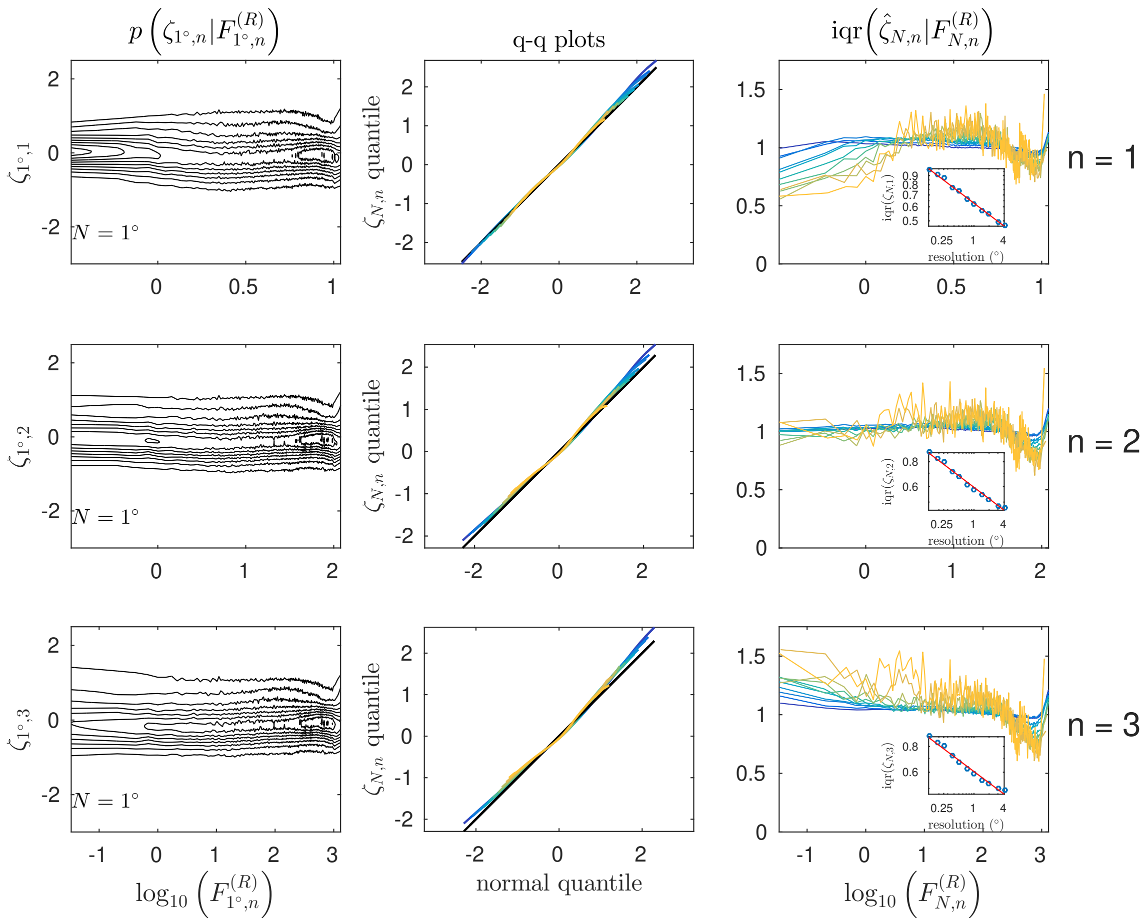


Figure 4: Statistics of the residual process $\zeta_{N,n}$ (Eq. 10) conditioned on the resolved flux $F^{(R)}_{N,n}$ for $n = 1$ (upper row), $n = 2$ (middle row), and $n = 3$ (lower row). Left column: kernel density estimates of the pdf of $\zeta_{1^\circ,n}$ conditioned on $F^{(R)}_{1^\circ,n}$ for a coarsening scale $N = 1^\circ$. Center column: quantile-quantile plots of $\zeta_{N,n}$ and a normal distribution for a range of coarsening scales N. The 1:1 line is indicated in black. Right column: interquartile range of $\hat{\zeta}_{N,n} = \zeta_{N,n}/(\gamma_n N^{\alpha_n})$. The coefficients α_n, γ_n are determined from a regression fit of $\log_{10}(\mathrm{iqr}(\zeta_{N,n}))$ to $\log_{10}(N)$ (inset). In the center and right columns, the color scheme indicating N is as in the right column of Figure 2.

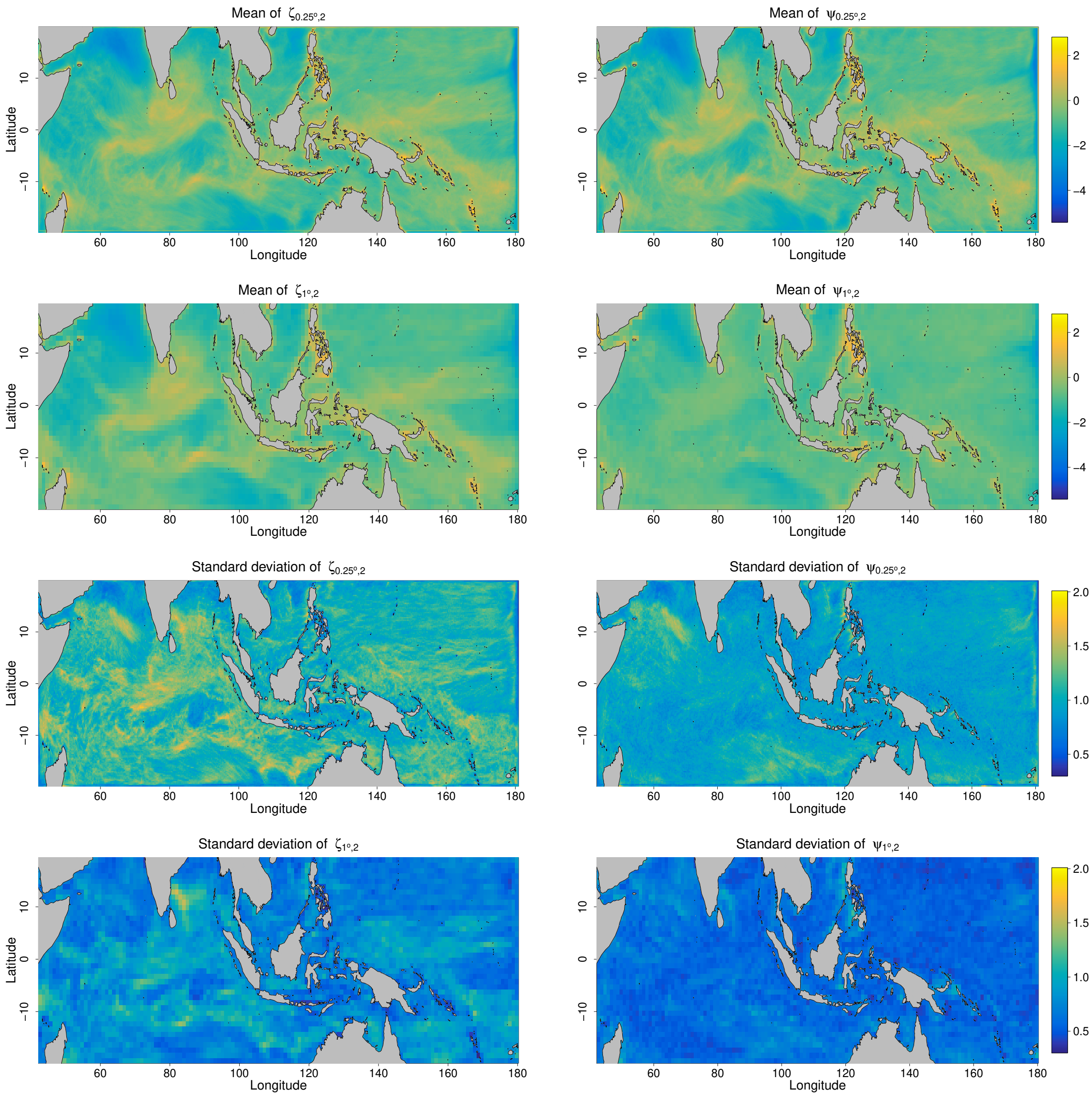


Figure 5: Mean (upper rows) and standard deviation (lower rows) of the residuals $\zeta_{N,2}$ (left) and $\psi_{N,2}$ (right) at the coarsening scales N=0.25° (first and third rows) and N=1° (second and fourth rows).

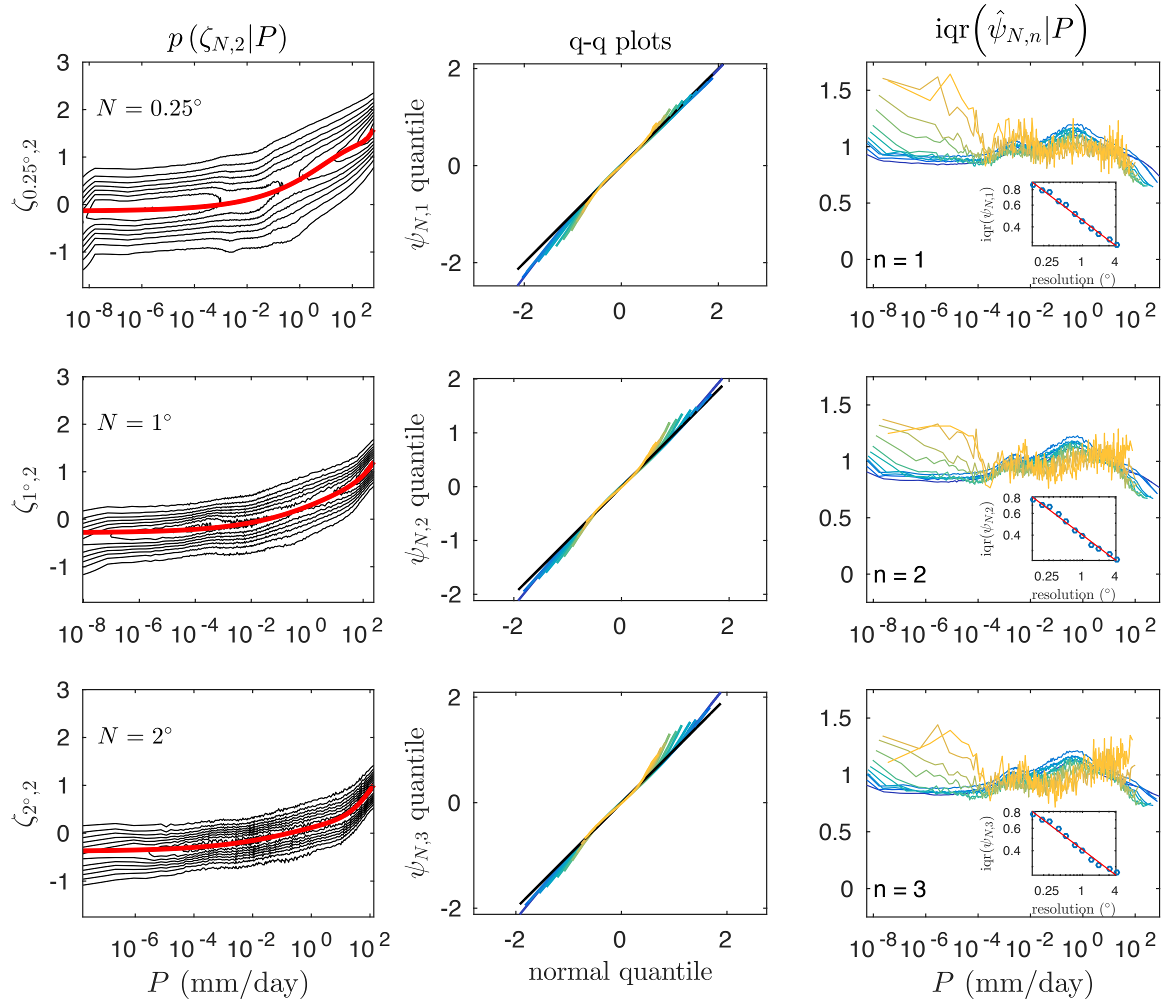


Figure 6: Left column: distributions of the residual process $\zeta_{N,2}$ conditioned on the precipitation rate P for $N = 0.25°$ (top), $N = 1°$ (middle), and $N = 2°$ (bottom). The red curves show the regression relationship Eq. (13) for each value of N. Center column: quantile-quantile plots of $\psi_{N,n}$ and a normal distribution for a range of coarsening scales N. The 1:1 line is indicated in black. Right column: interquartile range of $\hat{\psi}_{N,n} = \psi_{N,n}/\left(\mu_n N^{\lambda_n}\right)$ conditioned on P, for a range of coarsening scales N. The coefficients μ_n, λ_n are obtained as regressions of $\log_{10}\left(\mathrm{iqr}(\psi_{N,n})\right)$ on $\log_{10}(N)$ (inset). The color scheme in the center and left columns is as in the right column of Figure 2.

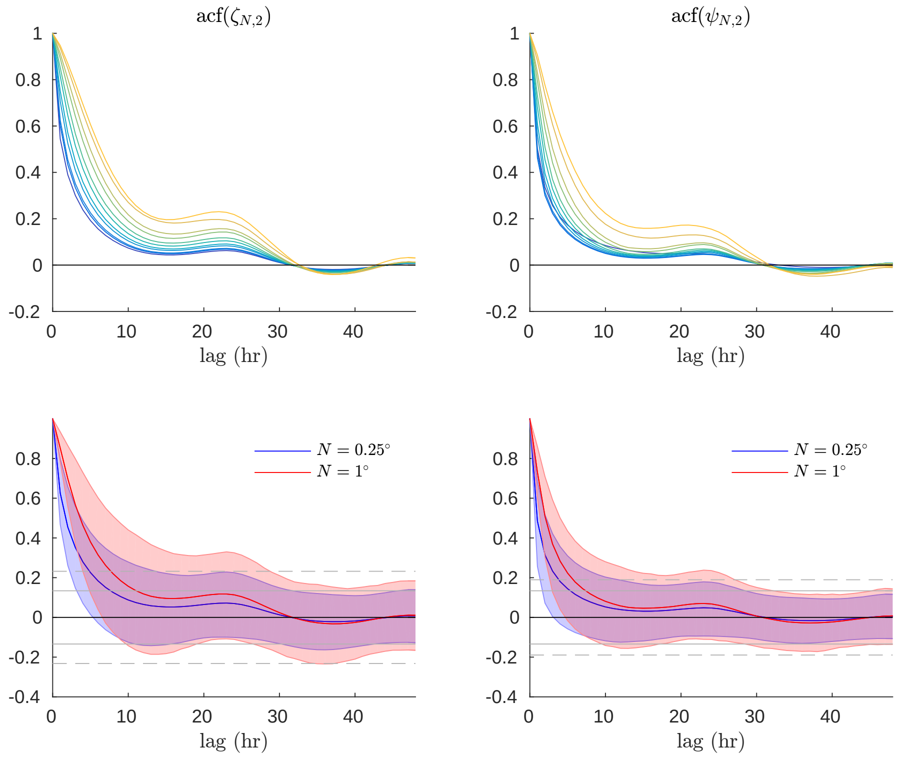


Figure 7: Temporal autocorrelation functions of $\zeta_{N,2}$ and $\psi_{N,2}$. Upper row: Composite acfs across all points in the domain, for a range of coarsening scales N. The color scheme is as in the right column of Figure 2. Lower row: Composite (solid curves) and interdecile range (shaded band) of acfs across all points in the domain, for $N = 0.25°$ and $N = 1°$. The solid grey lines show the 95% correlation coefficient confidence interval estimated as $\pm 1.96/\sqrt{N}$ with $N = 216$ (the raw number of degrees of freedom). The dashed grey lines indicate the confidence ranges reducing N by a factor of 3 (left panel) or 2 (right panel) to account approximately for the serial dependence of the time series.

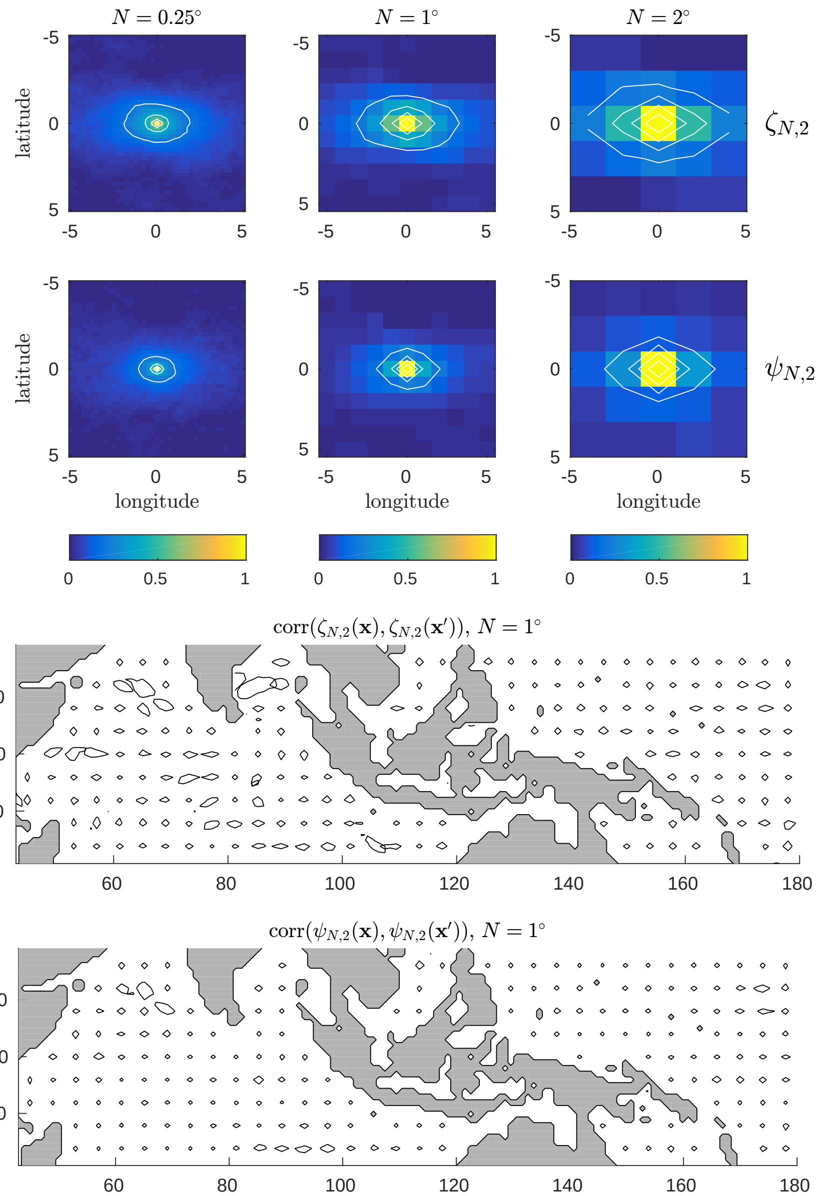
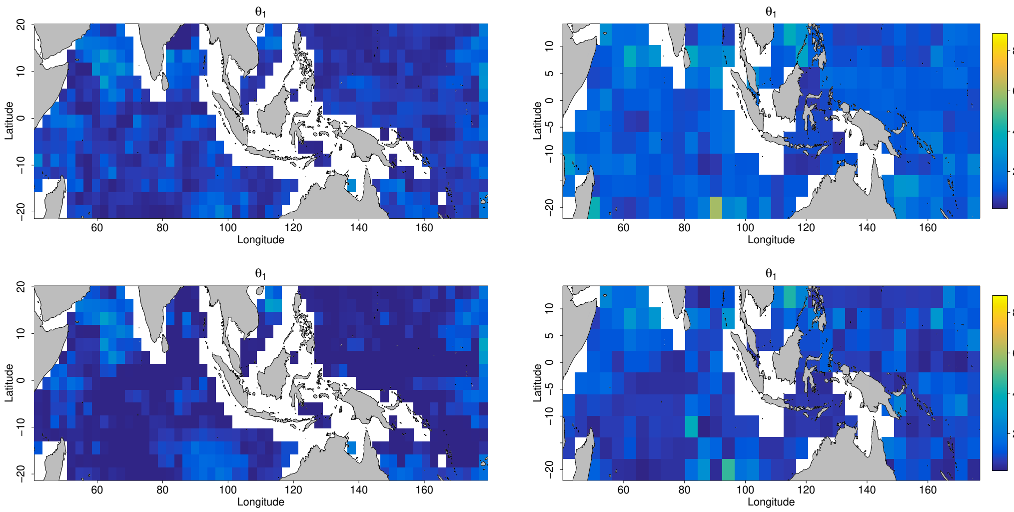
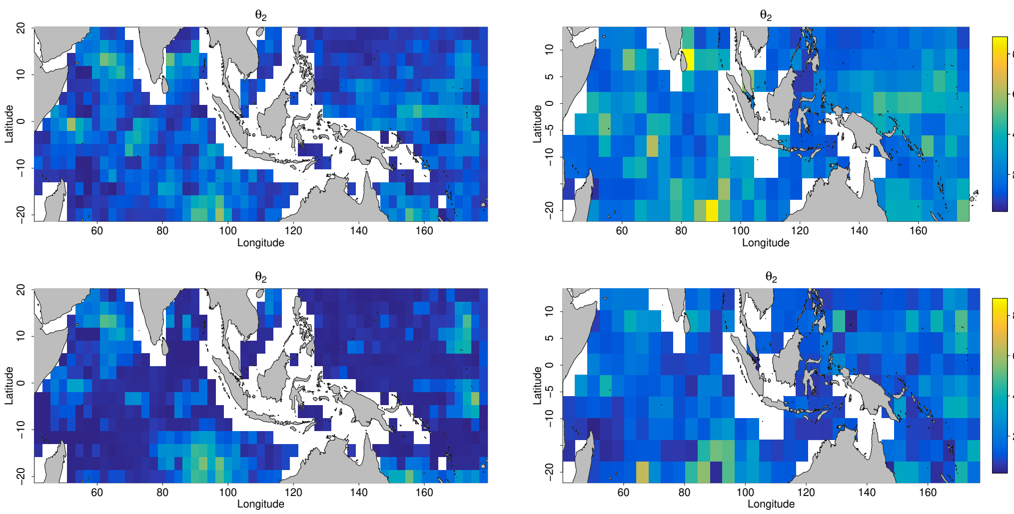


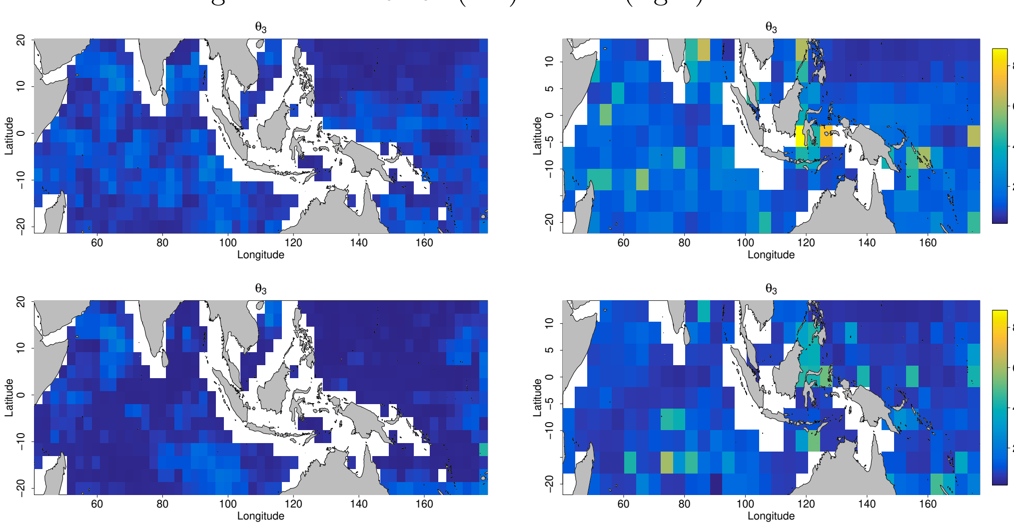
Figure 8: Spatial correlation functions of $\zeta_{N,2}$ and $\psi_{N,2}$. Upper two rows: composite spatial correlation functions for base points across the domain, for coarsening scales $N = 0.25^\circ, 1^\circ$, and 2°. Contour intervals: 0.2,0.4,0.6,0.8,1. Lower two rows: contours surrounding the regions for which the squared correlation values with base points (on a coarse $4^\circ \times 4^\circ$ grid) exceed 0.5.



(a) Anisotropic meridional scale θ_1 (in degrees) of the covariance Eq. (16) fit to $\zeta_{N,2}$ (upper) and $\psi_{N,2}$ (lower) for two coarsening scales $N = 0.25°$ (left) and $1°$ (right).



(b) Anisotropic zonal scale θ_2 (in degrees) of the covariance Eq. (16) fit to $\zeta_{N,2}$ (upper) and $\psi_{N,2}$ (lower) for two coarsening scales $N = 0.25°$ (left) and $1°$ (right).



(c) Temporal range θ_3 (in hours) of the covariance Eq. (16) fit to $\zeta_{N,2}$ (upper) and $\psi_{N,2}$ (lower) for two coarsening scales $N = 0.25°$ (left) and $1°$ (right).

Figure 9: Maps of estimated covariance parameters θ_1, θ_2, and θ_3 for $\zeta_{N,2}$ and $\psi_{N,2}$.

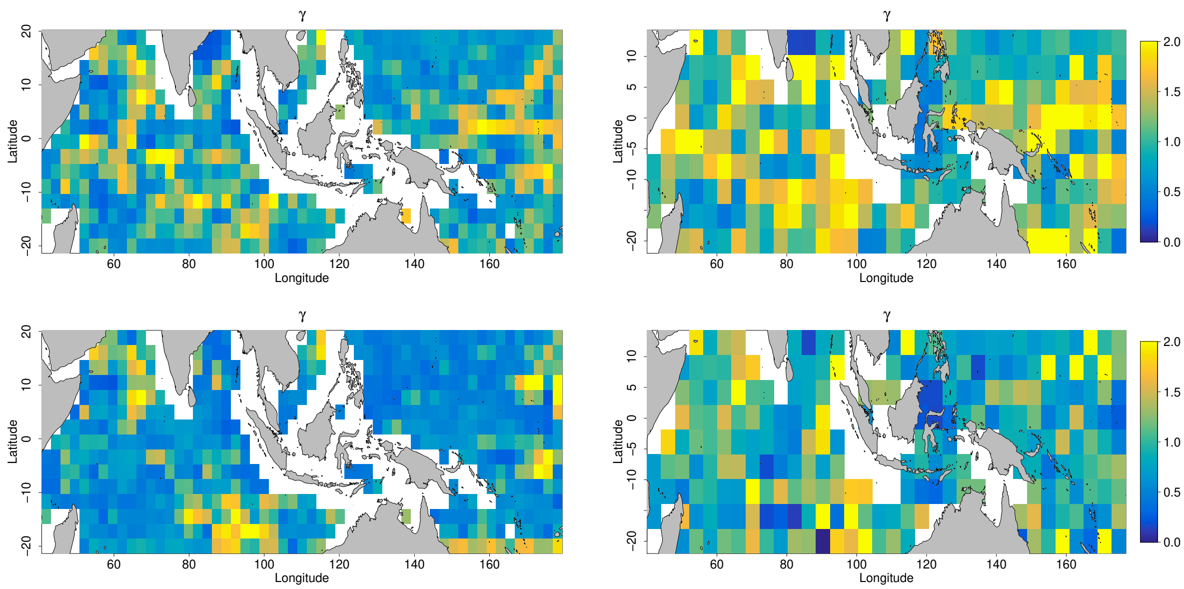


Figure 10: Exponent parameter γ of the covariance Eq. (16) fit to $\zeta_{N,2}$ (upper) and $\psi_{N,2}$ (lower) for different coarsening scales $N = 0.25^\circ$ (left) and 1° (right).

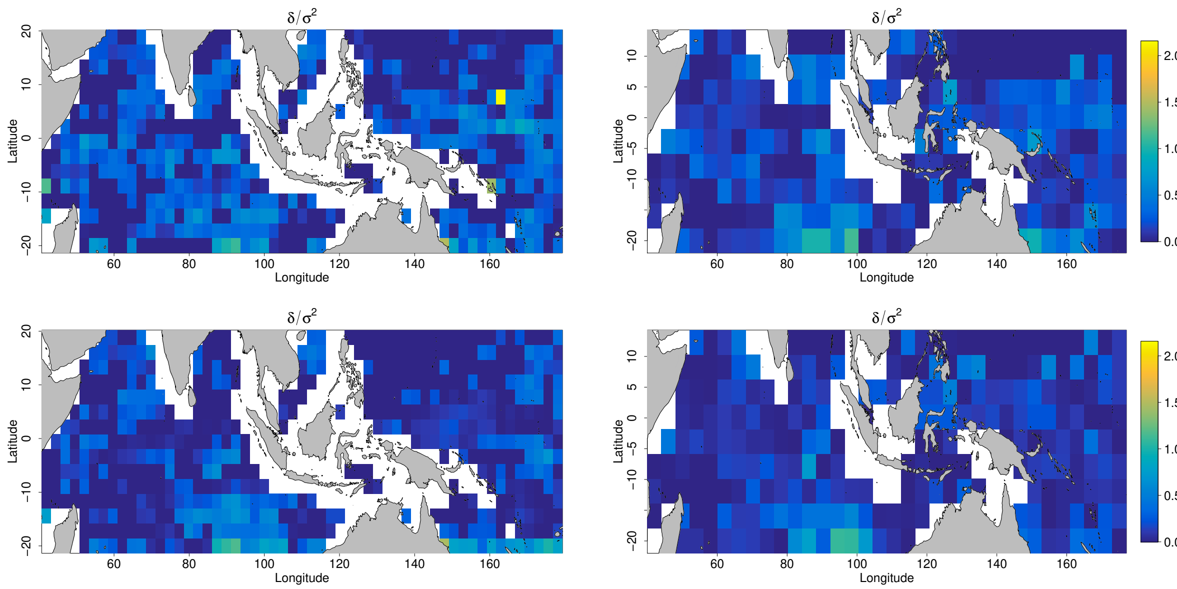


Figure 11: Ratio of nugget parameter δ from the fit covariance Eq. (16) over the empirical variance of the observed error process $\epsilon_{N,2}$, model for $\zeta_{N,2}$ (upper) and $\psi_{N,2}$ (lower) for different coarsening scales $N = 0.25°$ (left) and $1°$ (right).

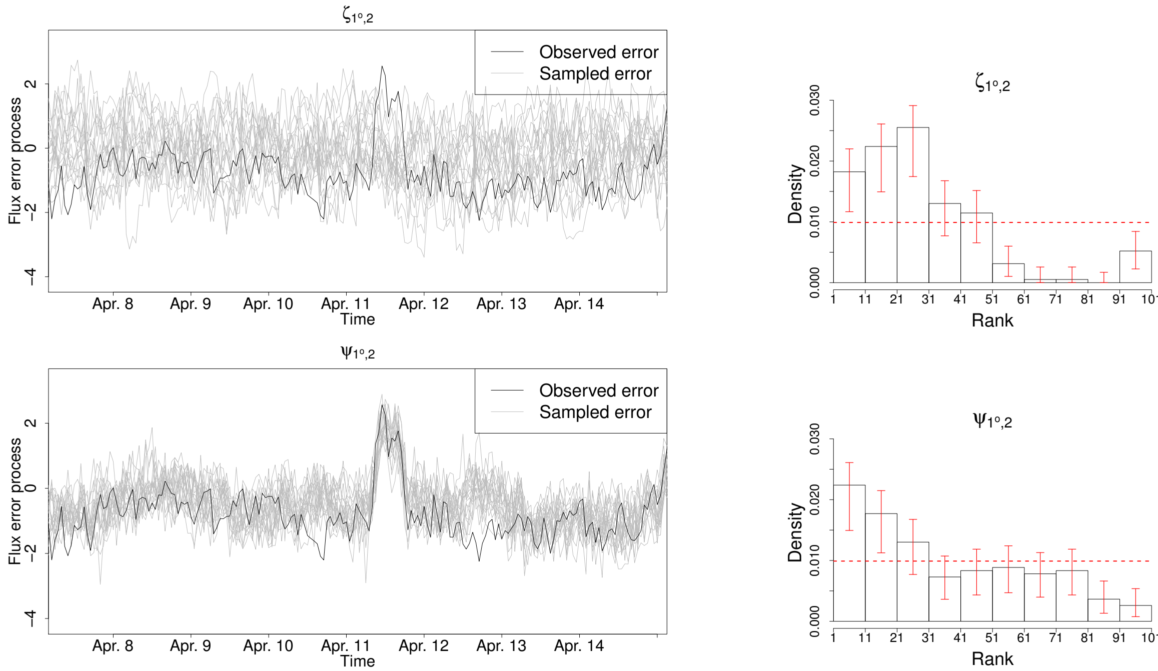


Figure 12: Left: Times series of the error process $\varepsilon_{1^\circ,2}$ at (161.9°E,0.83°S) (black line) and synthetic samples (grey lines) obtained from realizations of $\zeta_{1^\circ,2}$ (upper panel) using Eq. (10) and from realizations of $\psi_{1^\circ,2}$ (lower panel) using Eq. (13). Right: rank histograms between the observed error process $\varepsilon_{1^\circ,2}$ and error process reconstructed from the samples generated from $\zeta_{1^\circ,2}$ (upper panel) and from $\psi_{1^\circ,2}$ (lower panel). The red error bars correspond to 95%-confidence intervals associated with each estimated count of the histogram, the horizontal red line corresponds to the uniform histogram expected under perfect match between observed and simulated error process.

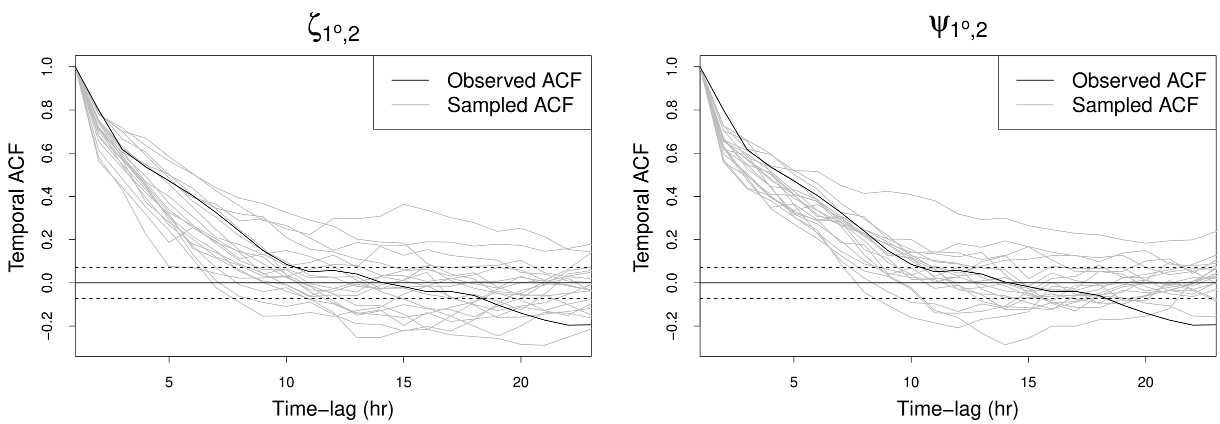


Figure 13: Temporal autocorrelation functions (acf) of $\varepsilon_{1^\circ,2}$ at (161.9°E,0.83°S). Black: observed autocorrelation function, Grey: autocorrelation functions from synthetic samples based on realizations of $\zeta_{1^\circ,2}$ (left) and $\psi_{1^\circ,2}$ (right). The dashed lines correspond to 95%-confidence intervals for a white noise process.

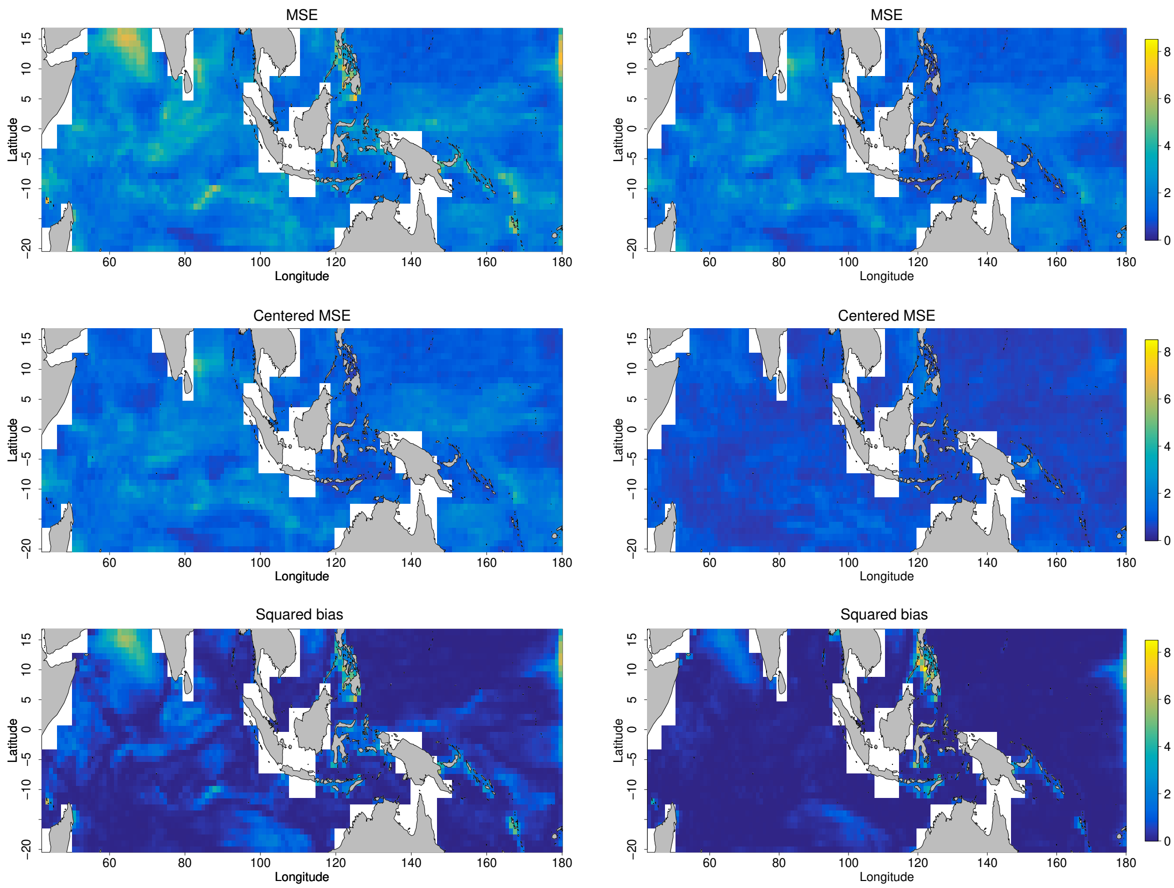


Figure 14: Top panel: Total Mean Square Error (MSE), Central panel: centered MSE, Bottom panel: Squared bias between the observed error process $\epsilon_{1^\circ,2}$ and its samples generated from $\zeta_{1^\circ,2}$ (left) and $\psi_{1^\circ,2}$ (right). The total MSE can be decomposed between the centered MSE and the squared bias, in order to assess the contribution of bias and of fluctuations to the total MSE.

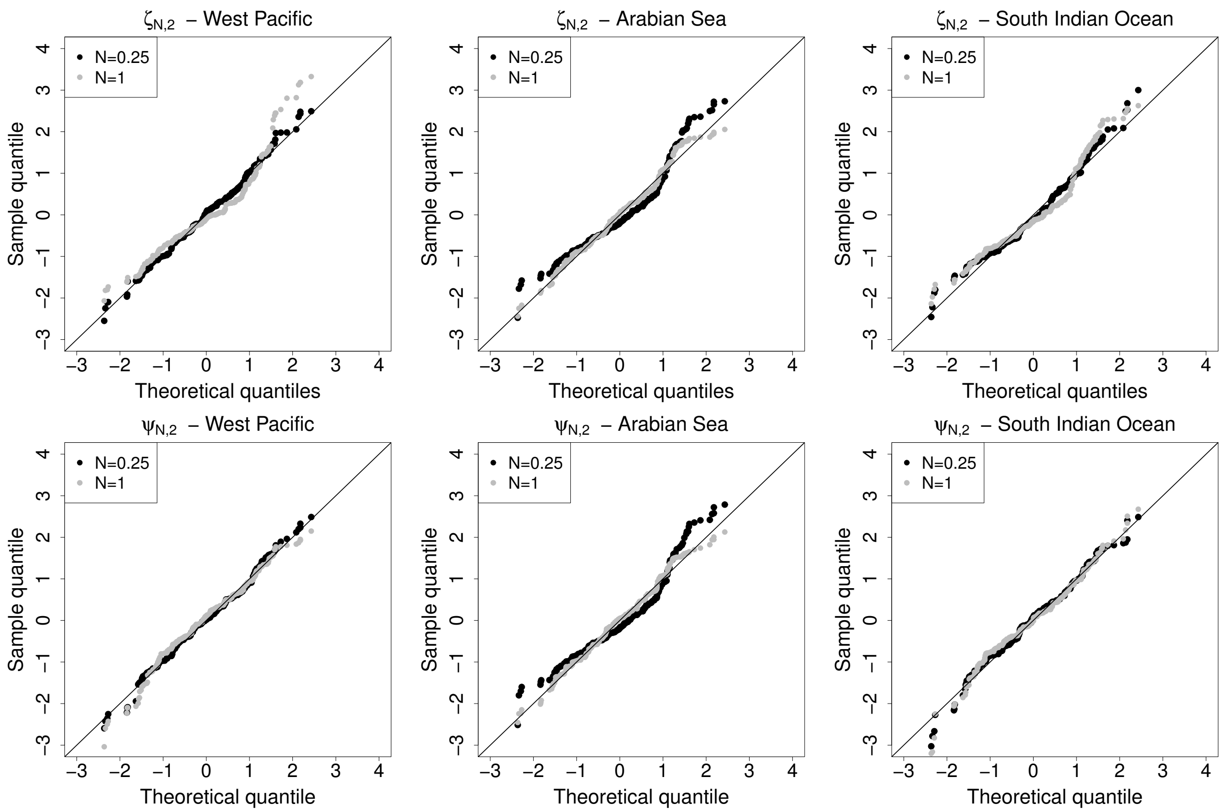


Figure 15: Normal quantile-quantile plot for $\zeta_{N,2}$ (top) and $\psi_{N,2}$ (bottom) at the central location of the three subregions: Western Pacific (left), Arabian Sea (center), and Southern Indian Ocean (right). QQ-plots are depicted for two coarsening scales varies: $N = 0.25°$ (black) and $N = 1°$ (grey).

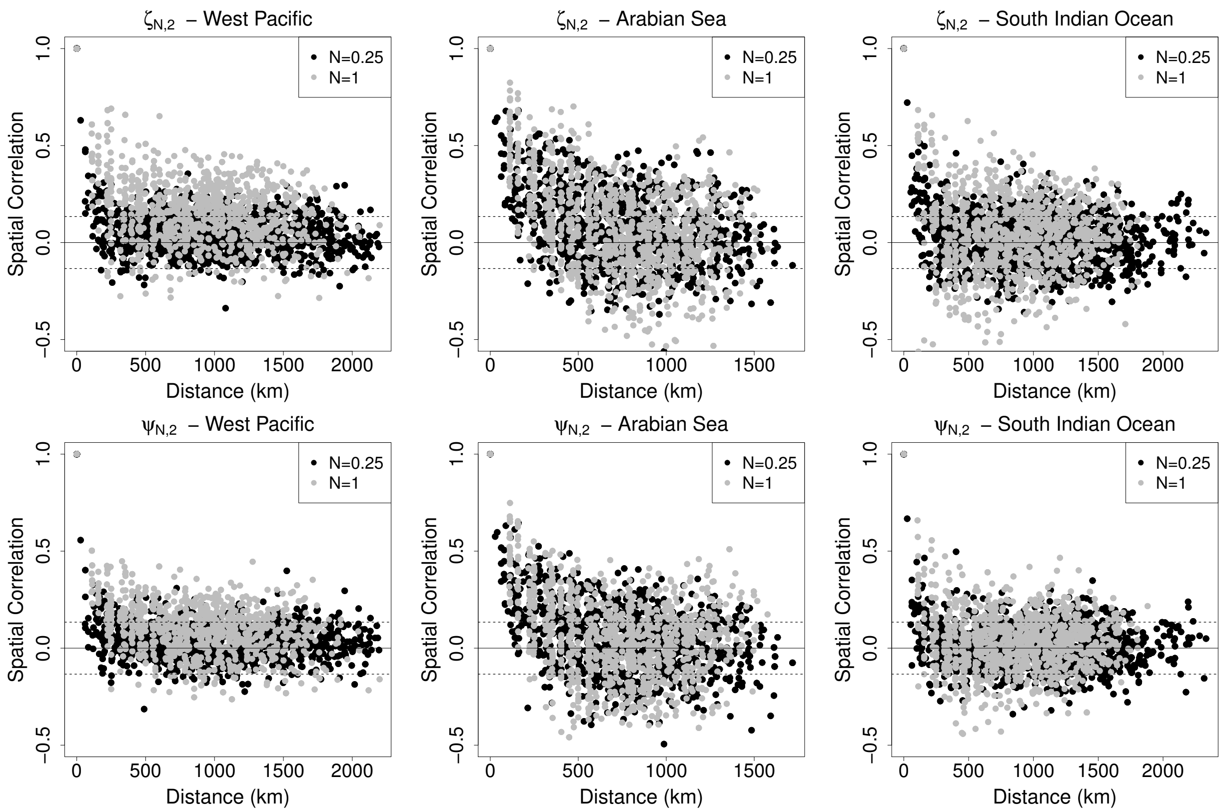


Figure 16: Spatial correlation of $\zeta_{N,2}$ (top) and $\psi_{N,2}$ (bottom) against the distance in km for the three subregions: Western Pacific (left), Arabian Sea (center), and Southern Indian Ocean (right) (only 50 random points are depicted). Correlations are depicted for two coarsening scales varies: $N = 0.25°$ (black) and $N = 1°$ (grey).