Monthly Weather Review

Robustness of the stochastic parameterization of sub-grid scale wind variability in sea-surface fluxes

--Manuscript Draft--

Manuscript Number:	MWR-D-22-0319
Full Title:	Robustness of the stochastic parameterization of sub-grid scale wind variability in sea-surface fluxes
Article Type:	Article
Corresponding Author:	Kota Endo University of Victoria Victoria, BC CANADA
Corresponding Author's Institution:	University of Victoria
First Author:	Kota Endo
Order of Authors:	Kota Endo
	Adam H. Monahan
	Julie Bessac
	Hannah M. Christensen
	Nils Weitzel
Abstract:	High-resolution numerical models have been used to develop statistical models of the enhancement of sea surface fluxes resulting from spatial variability of sea-surface wind. In particular, studies have shown that the flux enhancement is not a deterministic function of the resolved state. Previous studies focused on single geographical areas or used a single high-resolution numerical model. This study extends the development of such statistical models by considering six different high-resolution models, four different geographical regions, and three different ten-day periods, allowing for a systematic investigation of the robustness of both the deterministic and stochastic parts of the data-driven parameterization. Results indicate that the deterministic part, based on regressing the unresolved normalized flux onto resolved scale normalized flux and precipitation, is broadly robust across different models, regions, and time periods. The statistical features of the stochastic part of the model (spatial and temporal autocorrelation and parameters of a Gaussian process fit to the regression residual) are also found to be robust and not strongly sensitive to the underlying model, modelled geographical region, or time period studied. Best-fit Gaussian process parameters display robust spatial heterogeneity across models, indicating potential for improvements to the statistical model. These results illustrate the potential for the development of a generic, explicitly stochastic parameterization of sea-surface flux enhancements dependent on wind variability.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

LaTeX File (.tex, .sty, .cls, .bst, .bib)
Click here to access/download;LaTeX File (.tex, .sty, .cls, .bst, .bib);draft3.tex
Generated using the official AMS LaTeX template v6.1

Robustness of the stochastic parameterization of sub-grid scale wind variability in sea-surface fluxes

Kota Endo,[a] Adam H. Monahan,[a] Julie Bessac,[b] Hannah M. Christensen,[c] Nils Weitzel[d,e]

[a] *School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada*

[b] *Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, Illinois*

[c] *Department of Physics, University of Oxford, Oxford, United Kingdom*

[d] *Department of Geosciences, University of Tübingen, Tübingen, Germany*

[e] *Institute of Environmental Physics, Heidelberg University, Heidelberg, Germany*

Corresponding author: Kota Endo, kotae@uvic.ca

Julie Bessac's current affiliation: National Renewable Energy Laboratory, Golden, Colorado

ABSTRACT: High-resolution numerical models have been used to develop statistical models of the enhancement of sea surface fluxes resulting from spatial variability of sea-surface wind. In particular, studies have shown that the flux enhancement is not a deterministic function of the resolved state. Previous studies focused on single geographical areas or used a single high-resolution numerical model. This study extends the development of such statistical models by considering six different high-resolution models, four different geographical regions, and three different ten-day periods, allowing for a systematic investigation of the robustness of both the deterministic and stochastic parts of the data-driven parameterization. Results indicate that the deterministic part, based on regressing the unresolved normalized flux onto resolved scale normalized flux and precipitation, is broadly robust across different models, regions, and time periods. The statistical features of the stochastic part of the model (spatial and temporal autocorrelation and parameters of a Gaussian process fit to the regression residual) are also found to be robust and not strongly sensitive to the underlying model, modelled geographical region, or time period studied. Best-fit Gaussian process parameters display robust spatial heterogeneity across models, indicating potential for improvements to the statistical model. These results illustrate the potential for the development of a generic, explicitly stochastic parameterization of sea-surface flux enhancements dependent on wind variability.

1. Introduction

While numerical models of the atmosphere have seen remarkable improvement in the last 70 years, such models still require approximations of sub-grid scale processes such as surface fluxes that are not fully resolved. Traditionally, sea surface fluxes have been parameterized as a deterministic function of the resolved state. However, such an assumption requires a scale separation between the unresolved sub-grid scale processes and the resolved atmospheric model grid-box scale processes, an assumption not necessarily valid nor appropriate when the length scale of the unresolved process approaches the length scale of the grid. In the absence of scale separation, stochastic parameterization that includes an explicitly random component is appropriate (Berner et al. 2017; Palmer 2019).The present study expands on previous works of Bessac et al. (2019, 2021), which used a coarse-graining approach to investigate the difference between 'true' (at the scale of a high-resolution simulation) and 'resolved' fluxes for a range of coarsening scales. These previous studies considered nine days of output from a single regional convection-permitting model simulation of the Indo-Pacific Warm Pool region. To quantitatively assess our confidence in the use of such a parameterization in operational weather and climate models, the present study investigates how sensitive the parameterization is to the spatial domain, simulation period, and high-resolution model used to develop it.

a. Parameterized sea surface fluxes

Sea surface fluxes of e.g. energy, momentum, and gases are normally represented in weather and climate models using the bulk flux formulation

$$\mathcal{F}_x = \overline{\rho_a}\,\overline{s}\,c_x(\overline{s})(\overline{X_s} - \overline{X_a}) \tag{1}$$

with $\mathcal{F}_x$ the surface flux for some quantity x (in amount of x per unit surface area per unit time), ρ_a the air density, s the wind speed, c_x the dimensionless exchange coefficient for x at wind speed $\overline{s}$, and X_s, X_a the amount of x per atmospheric mass at the surface and at some specified height, respectively. The functional dependence of the exchange coefficient on wind speed is generally obtained empirically; examples of these functional forms are presented in Lin and Sheng (2020). The overbars in Equation 1 represent Reynolds averaging.

Jensen's inequality for the mean wind speed $\overline{s}$ and the norm of the means of the vector components of wind s_v implies that

$$s_v = \sqrt{\overline{u}^2 + \overline{v}^2} \leq \overline{\sqrt{u^2 + v^2}} = \overline{s} \tag{2}$$

where u, v, are the horizontal wind components. As such, the norm of the mean vector field is always an underestimate of the true mean wind speed. Since global and regional models represent Reynolds averaged variables, numerical models must use Reynolds averaged wind vector components instead of scalar mean wind $\overline{s}$ when computing surface flux. Compensation for such errors arising from Reynolds averaging in time associated with wind speeds computed at model resolution has traditionally been achieved by introduction of the so-called wind gustiness s_g(e.g. Jabouille et al. (1996))

$$\overline{s}^2 = s_v^2 + s_g^2 \tag{3}$$

where s_g must be parameterized using quantities computed by the numerical model, such as the turbulent surface heat flux and precipitation (Godfrey and Beljaars 1991; Zeng et al. 2002; Redelsperger et al. 2000; Miller et al. 1992; Deardorff 1970; Mahrt et al. 2021; Blein et al. 2022).

In fact, the spatial averaging of wind over model gridboxes implied by spatial discretization also results in underestimation of surface fluxes. The gustiness scheme to correct for time averaging can be expanded to include sub-grid scale velocity contributions to correct for spatial averaging as discussed in Mahrt and Sun (1995), expressing time averaged fine-scale wind components $\overline{u}$, $\overline{v}$, as a sum of space-time averaged wind components $\langle\overline{u}\rangle$, $\langle\overline{v}\rangle$ and their local deviations u_*, v_*. Hence

$$\overline{u} = \langle\overline{u}\rangle + u_* \tag{4}$$

$$\overline{v} = \langle\overline{v}\rangle + v_* \tag{5}$$

$$\langle\overline{u}^2 + \overline{v}^2\rangle = \langle\overline{u}\rangle^2 + \langle\overline{v}\rangle^2 + s_*^2 \tag{6}$$

with $s_*^2 = \langle u_*^2\rangle + \langle v_*^2\rangle$, since $\langle u_*\rangle = \langle v_*\rangle = 0$. Jensen's inequality implies that

$$\sqrt{\langle\overline{u}\rangle^2 + \langle\overline{v}\rangle^2} \leq \overline{\langle\sqrt{u^2 + v^2}\rangle}. \tag{7}$$

The right-hand side of this inequality is the desired true space- and time-mean wind speed; the left-hand side is the quantity directly available from the numerical simulation. Such inequalities

hold in general for fluxes which are convex functions of wind speed and the difference between the two quantities constitute what we wish to recover using a flux enhancement.

As with the gustiness, parameterizations of flux enhancements by subgrid-scale (SGS) velocity variability in a numerical model must use resolved model state variables. This SGS flux enhancement can also be represented using a gustiness velocity, as in Blein et al. (2022). Using a formal model selection approach, this study found that the optimal set of predictors of s_*^2 uses information related to deep convection and the local resolved flow. Following Bessac et al. (2019), the present study considers a different approach to this parameteriztaion problem.

b. Stochastic parameterizations

Deterministic parameterization of the difference between 'true' and 'resolved' fluxes assumes sufficient scale separation such that the parameterized sub-grid scale processes are 'in the thermodynamic limit,' where the large number of small-scale fluctuations around the mean largely average out, such that their net effect is negligible. This assumption is not generally appropriate when unresolved atmospheric processes have length scales that approach model resolution, and as such cannot be deterministically parameterized (e.g. Bessac et al. 2019; Christensen 2020; Dorrestijn et al. 2013). The lack of scale separation can, however, be accounted for by use of a stochastic parameterization.

It is known that stochastic corrections to air-sea fluxes are capable of causing noise-induced drift upon climate, as a consequence of the asymmetry in the influence rendered by positive and negative surface buoyancy fluxes (Williams 2012). As a result, the climate of a model with stochastic parameterizations of sub-grid scale processes can be expected to be different from that of a comparable model with deterministic parameterizations (even if the stochastic corrections have zero mean). Thus for the reduction of systematic errors at the climate-scale the development and evaluation of stochastic parameterizations is of importance to the Earth System modelling community, as discussed by Zadra et al. (2018). Furthermore, the potential increase in ensemble spread introduced by explicit stochasticity may be of interest for the improvement of prediction capabilities (e.g. Palmer et al. 2009) thereby presenting an avenue for further improvements to contemporary numerical weather prediction and climate modelling. Davini et al. (2017); Dawson

and Palmer (2015); and Vidale et al. (2021) discuss the use of stochastic model physics as a cost efficient substitute to increasing model resolution.

c. *Idealized flux formulation*

As discussed above, it is known that gustiness and SGS velocity variability contribute to the coarse resolution flux in a manner that is not generally negligible (e.g. Bessac et al. 2019; Blein et al. 2020). While SGS variability of other variables affect surface fluxes, SGS velocity variations are generally the dominant effect (Blein et al. 2020). It is therefore beneficial to focus on that part of the sea surface flux enhancement due entirely to SGS velocity variations. In the present work, the surface wind dependent part of some sea surface flux is expressed using the idealized, normalized, and dimensionless form

$$F(s) = \left(\frac{s}{s_0}\right)^n \tag{8}$$

where $s_0 = 1$ m/s is a scaling factor. Throughout the rest of this study, it will be assumed that all model variables are Reynolds averaged and so the overbar will not be used. The non-dimensional quantity expressed as Equation 8 will be referred to as 'the normalized flux' to distinguish this quantity from the full flux. Neglecting the wind speed dependence of the exchange coefficient, the normalized flux of energy or water vapour is represented with $n = 1$ and momentum, with $n = 2$. A range of exponents have been proposed for fluxes of different gases (e.g. $n = 2$ in Wanninkhof 2014).

We investigate effects of spatial averaging by using convection permitting model outputs to approximate the state of a lower-resolution model by averaging fields on a lower-resolution spatial grid (coarse-graining), and finding the difference between the gridbox-mean normalized fluxes computed from the high-resolution (taken to be the 'truth') and coarsened fields. In particular we consider the quantity

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

the base 10 logarithm of the normalized flux enhancement due to sub-grid scale wind variations, where N denotes the averaging scale for coarsening of the fine-resolution model output. In Equation 9

$$F^{(T)}_{N,n} = \left\langle \left(\frac{\sqrt{u^2+v^2}}{s_0}\right)^n \right\rangle_N \tag{10}$$

and

$$F_{N,n}^{(R)} = \left(\frac{\sqrt{\langle u \rangle_N^2 + \langle v \rangle_N^2}}{s_0} \right)^n \tag{11}$$

are the 'true' and 'resolved-scale' normalized fluxes, respectively (with $<>_N$ denoting spatial averaging on a grid of resolution N). Note that Equation 9 implies

$$F_{N,n}^{(T)} = F_{N,n}^{(R)} + 10^{\varepsilon_{N,n}} \tag{12}$$

Negative values of $\varepsilon_{N,n}$ correspond to normalized SGS flux enhancements of less than one. By virtue of Jensen's inequality, $F_{N,n}^{(T)}$ will always exceed $F_{N,n}^{(R)}$ so $\varepsilon_{N,n}$ is real.

Following Bessac et al. (2019, 2021) we parameterize $\varepsilon_{N,n}$ in terms of resolved variables with the linear regression

$$\varepsilon_{N,n} = \underbrace{(A_{N,n})_0 + \sum_{k=1}^{3} (A_{N,n})_k \left(\log_{10} F_{N,n}^{(R)} \right)^k + \sum_{l=1}^{4} (B_{N,n})_l P_N^{l/4}}_{= \mu_{N,n}} + \psi_{N,n} \tag{13}$$

with regression coefficients A, B, and the coarse-grained resolved scale precipitation P_N (the fine-scale precipitation rate averaged over the coarse grid box). The quantity $\mu_{N,n}$ denotes the deterministic part of $\varepsilon_{N,n}$, computed entirely from resolved-scale variables $F_{N,n}^{(R)}$, P_N, whereas $\psi_{N,n}$ is the statistical residual, modelled to be stochastic and independent of resolved-scale variables. This statistical model thus assumes that the regression fit captures all of the dependence of $\varepsilon_{N,n}$ on resolved variables $F_{N,n}^{(R)}$ and P_N.

d. Objectives

Data-driven climate model development exists at an interesting intersection wherein the statistical and computational methods employed advance very quickly, but at the same time testing and implementing a novel model is an expensive process both in terms of time and in terms of computational resources. Furthermore, the parameterizations that are employed in weather and climate modelling are oftentimes developed using output from one high-resolution numerical model, leading to questions about the model-dependence of these results. As such, it is important

to systematically study the robustness of a data-driven parameterizations, as is done in this study for the context of sea surface flux parameterization.

Previous studies of the flux enhancement by SGS wind variations using a coarse-graining approach have considered output from single models in a limited number of geographical regions over short time periods. For example, the deterministic parameterization of Blein et al. (2022) considered single months of output from the AROME model in two domains (the eastern Indian Ocean around La Réunion and the Antilles region of the Carribean). Bessac et al. (2019, 2021) used nine days of output over the Indo-Pacific Warm Pool from the Cascade simulation of the United Kingdom Met Office Unified Model (Holloway et al. 2012) during an active Madden-Julian Oscillation event. The question of robustness of these parameterizations to variations in the model, geographical domain, and time period used for their construction remains largely open. The object of this study is to expand on Bessac et al. (2019) by applying their proposed statistical model to thirty days of model output in a variety of geographical regions, extending from the equatorial regions to the extratropics, simulated by a set of global convection permitting atmospheric models participating in the DYAMOND (DYnamics of the Atmospheric general circulation Modeled on Non-hydrostatic Domains) initiative (Stevens et al. 2019). Such variety in model output allows for an investigation of the robustness of the parameterization of sea-surface flux enhancement due to subgrid-scale wind speed variability. Section 2 provides an overview of the models and regions considered, as well as the statistical tools used in the analysis. Results are presented in Section 3, the Discussion in Section 4, and the Conclusions in Section 5.

2. Convection-permitting models and statistical methods

a. Convection-permitting models

The DYnamics of the Atmospheric general circulation Modeled on Non-hydrostatic Domains (DYAMOND) initiative described by Stevens et al. (2019) provides forty days (August 1 – September 9, 2016) of high-resolution global numerical model output for a number of different models, produced from identical initial (atmospheric analysis) and boundary (sea-surface temperatures) conditions. Thus, it can be used to investigate the robustness of the statistical parameterization to the models used in its construction. The models considered in this study were selected based on availability in the DYAMOND data archive at the time the analysis started. The grid used for the

archive was retained until analysis, i.e. no re-gridding was performed prior to analyzing the output. The first ten days of the DYAMOND simulations are discarded as spin-up, following the protocols of Stevens et al. (2019).

To allow direct comparison with the results of Bessac et al. (2019), we also consider the Cascade experiment (Holloway et al. 2012; Love et al. 2011) although it is not part of DYAMOND, but a separate project using the UK Met Office Unified Model. The fact that the Cascade project modeled a time-period different from DYAMOND (April 6 – 15, 2009) allows for a further test of robustness with respect to time periods. For brevity and consistency, hereinafter 'Cascade' will denote the Cascade experiment whereas the shorthand 'UM' will denote the UM contribution to DYAMOND. Note also that the Cascade simulation considered only the Indo-Pacific Warm Pool region, and that in the present analysis the first day of the Cascade simulation is discarded as spin-up, as in Bessac et al. (2019, 2021).

A detailed study of the similarities and differences of the numerical models studied is presented in Endo (2022). The models FV3 and UM have relatively similar parameterization schemes; NICAM explicitly simulates rather than parameterizes cumulus convection or fractional cloudiness; and GEOS uses noticeably different parameterization schemes compared to the other models. A summary of model physics and dynamics is presented as Table 1.

b. Geographical regions studied

To assess the robustness of the regression model across various geographical domains with different meteorological properties, we consider four distinct regions, the Indo-Pacific Warm Pool, the subtropical Northeast Pacific, the subarctic Northeast Pacific, and a region of the Southern Ocean encompassing the Antarctic Peninsula and Patagonia (Figure 1).

The Indo-Pacific Warm Pool region is bounded by 20° N and 20° S and by 42° E and 177° E. Land masses within the region include the Maritime Continent, as well as northern Australia, the Horn of Africa, and South Asia. Being a region with substantial convection and resulting small-scale wind variability, the region is expected to have relatively large values of $\varepsilon_{N,n}$ (Figure 1). Furthermore, consideration of this region allows for direct comparison with the results of Bessac et al. (2019, 2021), since the domain is identical to the domain used in the Cascade experiment.

193 TABLE 1. DYAMOND model configurations, synthesizing Stevens et al. (2019); Satoh et al. (2019) and the
194 technical documentation applicable for each model.

Model	FV3	GEOS	IFS	NICAM	UM	Cascade
Technical documentation	Lin and Rood (1996, 1997); Lin (2004); Harris et al. (2020); Zhao et al. (2018)	Putman and Lin (2007); Putman and Suarez (2011); Molod et al. (2015)	Wedi (2014); Malardel et al. (2016); European Centre for Medium-Range Weather Forecasts (2020a)	Satoh et al. (2008, 2014); Kodama et al. (2015, 2021)	Walters et al. (2017); Wood et al. (2014)	Davies et al. (2005)
Grid	Cubed sphere	Cubed sphere	Octahedral reduced Gaussian	Icosahedral	Latitude-Longitude	Latitude-Longitude
Resolution	3.3 km	3.3 km	4.8 km	3.5 km	7.8 km	4 km
Time integration	Forward-in-time finite-volume, vertically Lagrangian	Forward-in-time finite-volume, vertically Lagrangian	Semi-implicit semi-Lagrangian	Split-explicit, vertically implicit	Semi-implicit semi-Lagrangian	Semi-implicit semi-Lagrangian
Hydrostatic model	No	No	Yes	No	No	No
Top height	39 km	80 km	80 km	50 km	85 km	40 km
Cumulus parameterization	Shallow	Full	Shallow	None	Shallow	Suppressed
Fractional cloudiness parameterized	Rotstayn et al. (2000); Smith (1990)	Bacmeister et al. (2006)	Räisänen et al. (2004); Tiedke (1993)	None	Wilson et al. (2008); Smith (1990)	Wilson and Ballard (1999); Smith (1990)
Orographic gravity waves	Garner (2005)	McFarlane (1987)	Lott and Miller (1997)	McFarlane (1987)	Lott and Miller (1997); Vosper (2015)	Webster et al. (2003)
Non-orographic gravity waves	Alexander and Dunkerton (1999); Lindzen (1981)	Garcia and Boville (1994)	Scinocca (2003); Warner and McIntyre (1999)	None	Scaife et al. (2002); Warner and McIntyre (1999)	None reported
Planetary boundary layer	Lock et al. (2000)	Lock et al. (2000)	European Centre for Medium-Range Weather Forecasts (2020b)	Mellor and Yamada (1982); Nakanishi and Niino (2006, 2009)	Lock et al. (2000)	Lock et al. (2000)

206 The subtropical Northeast Pacific region is bounded by 5° N and 30° N and by 120° W and
207 180° W. The only substantial land mass in this region is Hawaii. Low precipitation rates, along
208 with the strong, steady easterlies in this region contrast with the weak winds of the Warm Pool and

the strong but variable winds of the midlatitude regions. Furthermore, air-sea fluxes in the region are well studied in the literature (e.g. Fagan and Mackenzie 2007).

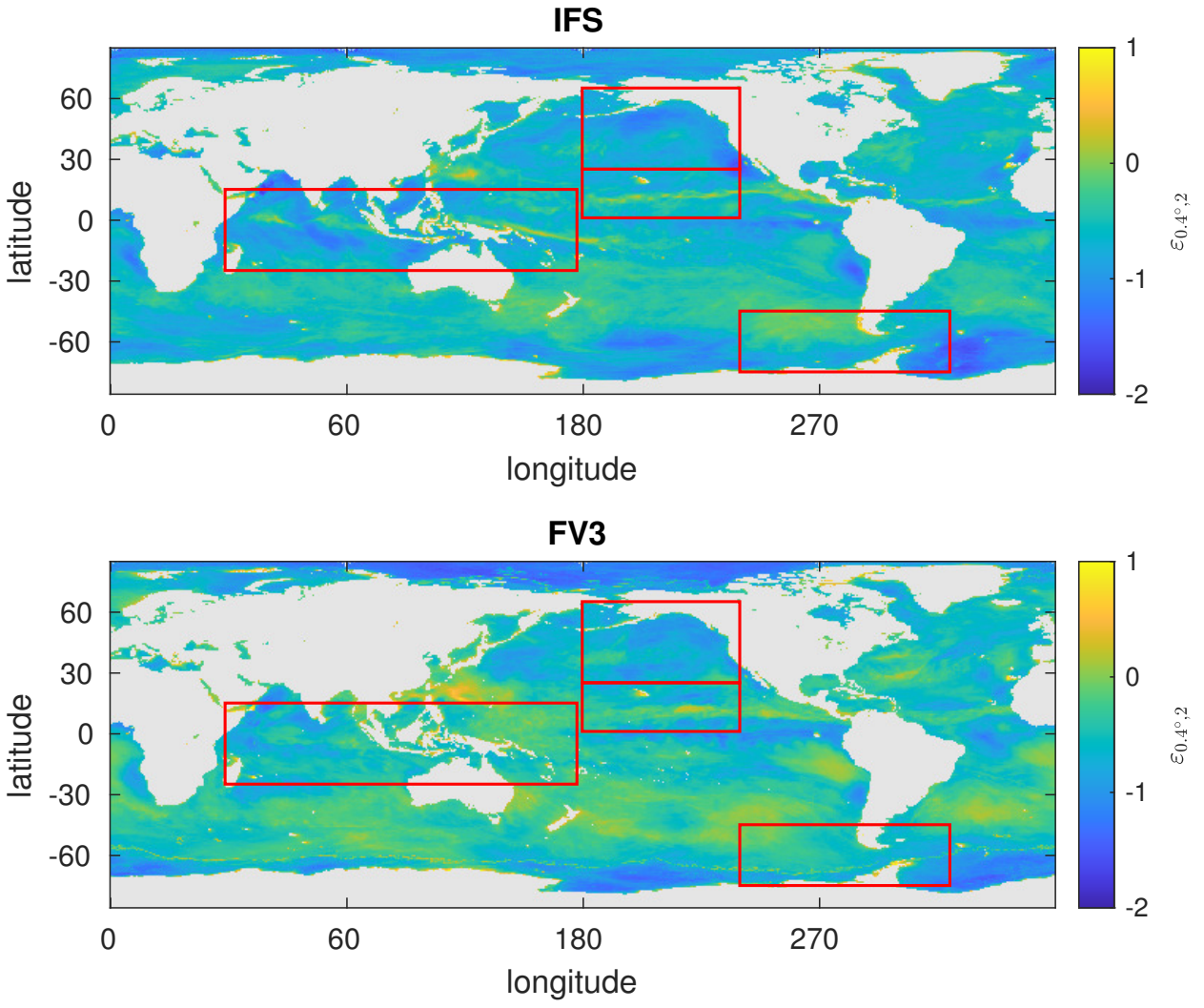
The subarctic Northeast Pacific domain is adjacent to the subtropical Northeast Pacific domain, bounded by 30° N and 70° N and by 120° W and 180°. The region includes the west coast of North America with Alaska and the Aleutian Islands, as well as the Russian Far East. This region contains the Pacific storm track and has different wind and precipitation climatologies compared to the Indo-Pacific Warm Pool and the subtropical Northeast Pacific, with highly variable winds due to synoptic-scale storm activity (albeit relatively weak in the simulation month of August). Quantification of $\varepsilon_{N,n}$ in this well observed region is potentially useful for analysis of observational data, such as those from Station P or Line P (e.g. Franco et al. 2021).

The region of the Southern Ocean considered is bounded by 40° S and 70° S and by 40° W and 120° W and includes Patagonia, the Antarctic Peninsula, and the Drake Passage region. This is a region of strong and variable winds (the 'roaring forties') and may be useful for statistical analysis and uncertainty quantification of datasets originating from the Drake Passage, such as the observations studied in Rees et al. (1997).

c. Coarse graining

The present study uses fine resolution numerical model output as 'truth' and uses a coarse graining approach to generate a proxy of the resolved scale model. This is done by setting a coarse-resolution grid and determining which fine-resolution grid boxes fall in each coarse-resolution box. The 'true' normalized flux contribution from the contributing fine-scale grid boxes are then computed and averaged on the coarse grid. The coarse-scale wind is obtained by averaging the zonal and meridional wind from the contributing fine-scale grid boxes and taking the norm of the averaged wind vectors. This norm is used to compute the 'resolved-scale' normalized flux. The difference between the 'true' and 'resolved-scale' normalized flux is used to compute $\varepsilon_{N,n}$. Resolved scale grid boxes which contain land are excluded from the analysis.

The present work will report results primarily for $N = 0.4°$ (approximately 45 km × 45 km in the Warm Pool, down to approximately 22 km × 45 km in the Southern Ocean) and $n = 2$, to allow for comparison with Bessac et al. (2021). The choice of this exponent means that the results presented in this study will relate directly to the parameterization of momentum fluxes. Fluxes of e.g. water



224 FIG. 1. Global $\varepsilon_{0.4°,2}$ means for IFS and FV3, with land masked out in grey and regions considered in this
225 study outlined in red. Note that the magnitude and sign of $\varepsilon_{N,n}$ are dictated by the choice of the scale factor s_0
226 in Equation 8.

241 and energy, for which $n = 1$, are considered in Endo (2022). The same qualitative conclusions are
242 found for these fluxes as for momentum fluxes.

243 A detailed analysis of the magnitude of $\varepsilon_{N,n}$ in the Warm Pool was presented in Bessac et al.
244 (2019). The overall magnitude of the SGS flux correction increases for larger coarsening scales
245 N. The relative error was found to be largest for the smallest values of resolved flux $F^{(R)}_{N,n}$. For
246 an averaging scale of 0.5° and a flux exponent $n = 2$, the relative error of the 50th percentile
247 resolved fluxes exceeds 10 percent approximately 15 percent of the time. Maps of time-mean

$\varepsilon_{0.4^\circ,2}$ over a ten-day subset of the simulation period for the IFS and FV3 models are broadly similar in distribution and magnitude (Figure 1). The fact that values of $\varepsilon_{0.4^\circ,2}$ simulated by the FV3 model are generally larger than those from IFS are consistent with the higher resolution of the FV3 model, since smaller-scale variability will manifest in the computed $\varepsilon_{N,n}$ values. These mean fields will also differ as a result of the fact that the 40-day simulation period is sufficiently long for each model to generate its own realization of internal weather variability.

d. Modelling stochastic enhancement with Gaussian fields

Estimation of the coefficients of the deterministic part of $\varepsilon_{N,n}$ in Equation 13 is done by means of least squares regression. The stochastic residual $\psi_{N,n}$ is modelled as a space-time random field with a specified covariance structure, separate from the aforementioned deterministic part. Once the parameters of the space-time covariance have been estimated, generation of realizations of $\psi_{N,n}$ is a straightforward process by use of a 'square root' of covariance matrices. Following Bessac et al. (2019, 2021), we consider the covariance function

$$k(x,x',y,y',t,t') = \sigma \exp\left(-d(x,x',y,y',t,t')^{\gamma}\right) \tag{14}$$

with

$$d(x,x',y,y',t,t') = \sqrt{\left(\frac{x-x'}{\theta_x}\right)^2 + \left(\frac{y-y'}{\theta_y}\right)^2 + \left(\frac{t-t'}{\theta_t}\right)^2} \tag{15}$$

as a covariance model that defines characteristic zonal, meridional and temporal lengthscales θ_x, θ_y, and θ_t, as well as variance parameter σ and roughness parameter $\gamma \in]0,2]$. Note that the nugget parameter δ included in Bessac et al. (2019) has been omitted, as was done in Bessac et al. (2021), due to its small contributions in Bessac et al. (2019). In this study, parameters of the covariance model are estimated in locally stationary fashion via maximum likelihood as in Bessac et al. (2019, 2021). In practice, obtaining a correlated random field with known covariance is not anticipated to be computationally expensive, especially in light of recent developments by Chen and Stein (2021) and Tsyrulnikov and Gayfulin (2017), the former improving computation cost of covariance matrices and the latter being an alternative spatio-temporal pseudo-random Gaussian field generator.

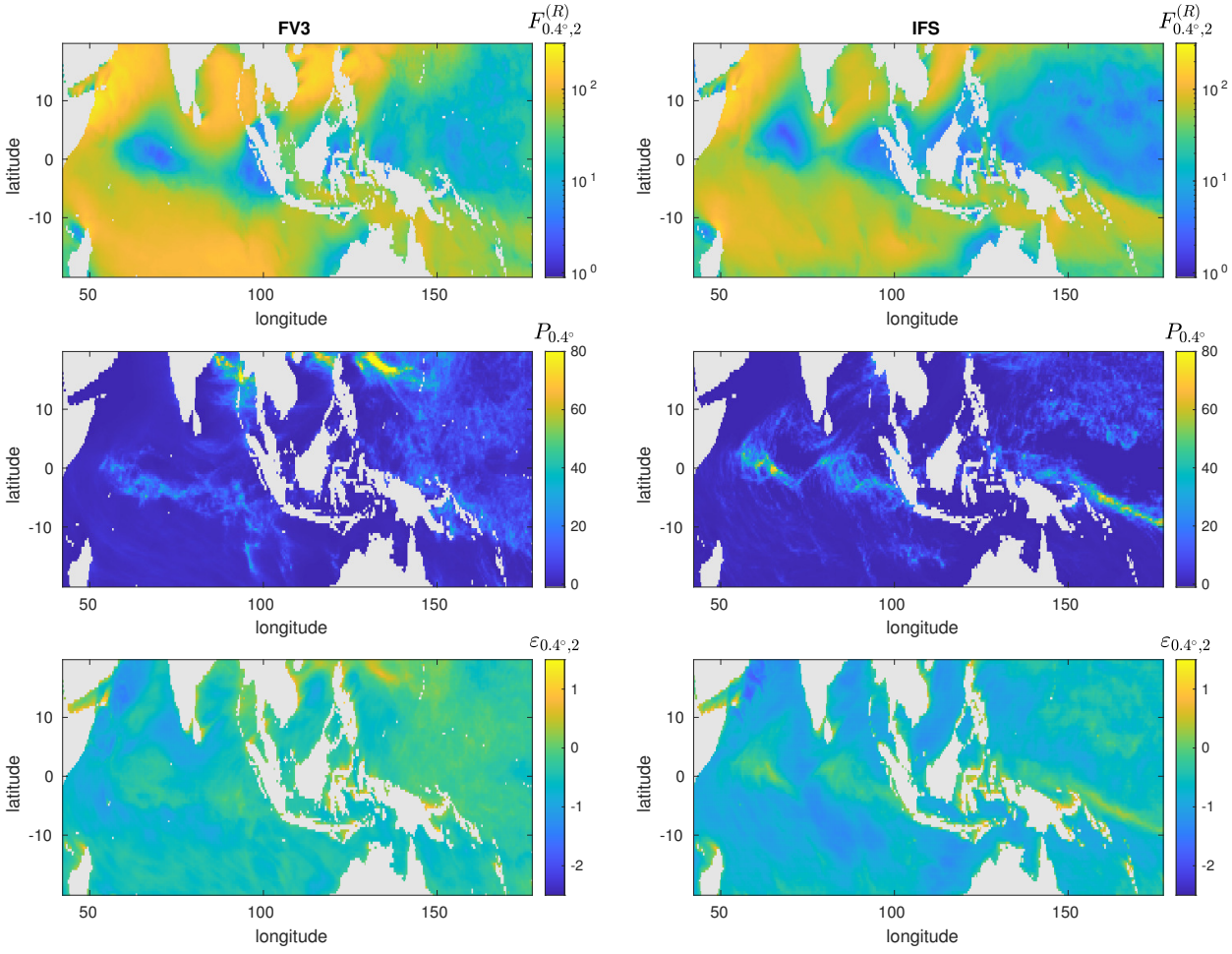


FIG. 2. Time means of resolved flux $F^{(R)}_{0.4^\circ,2}$ (top), precipitation $P_{0.4^\circ}$ (middle), and normalized flux enhancement $\varepsilon_{0.4^\circ,2}$ (bottom) for FV3 (left) and IFS (right) for the Indo-Pacific Warm Pool. The grey mask denotes land, the vertical axis denotes latitude, and the horizontal axis denotes longitude.

3. Analysis

a. Statistical features of resolved-scale normalized flux and precipitation

We first consider the statistics of resolved-scale normalized flux and precipitation, the explanatory variables utilized in the parameterization of the deterministic part of $\varepsilon_{N,n}$. These statistics may be different across models, spatial domains, and time periods because of model differences, intrinsic temporal non-stationarity or spatial heterogeneity, or sampling variability.

Figure 2 shows maps of the time mean resolved-scale normalized flux, precipitation, and residuals $\varepsilon_{0.4^\circ,n=2}$ for the Indo-Pacific Warm Pool for two models (IFS and FV3). It can be observed from

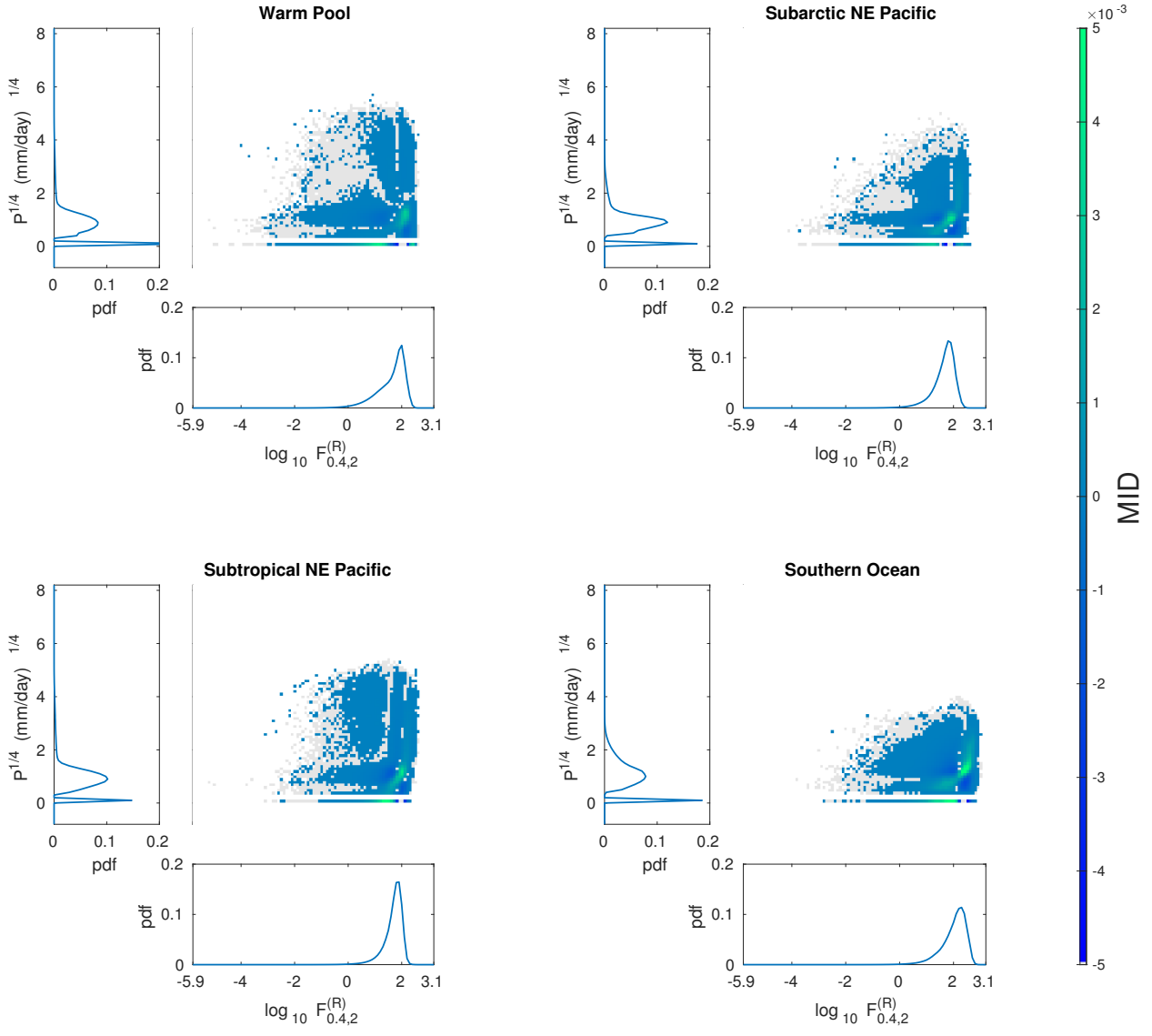


FIG. 3. Distribution of precipitation and resolved-scale normalized flux for the four regions considered for IFS, plotted as a mutual information density and respective marginal densities, with $N = 0.4°$, $n = 2$. Significant values of the mutual information density (at the five percent significance level with a two-tail test) are shown in color, whereas grey indicates that the null hypothesis could not be rejected.

Figure 2 that there are similarities between the geographical distribution of time mean resolved-scale normalized flux, precipitation, and $\varepsilon_{N,n}$, demonstrating some form of interdependence between the variables. Such a dependence reflects the fact that smaller values $F^{(R)}_{N,n}$ (e.g. blue regions in the top row of Figure 2) can result from increased cancellation of vector winds within a grid-box and

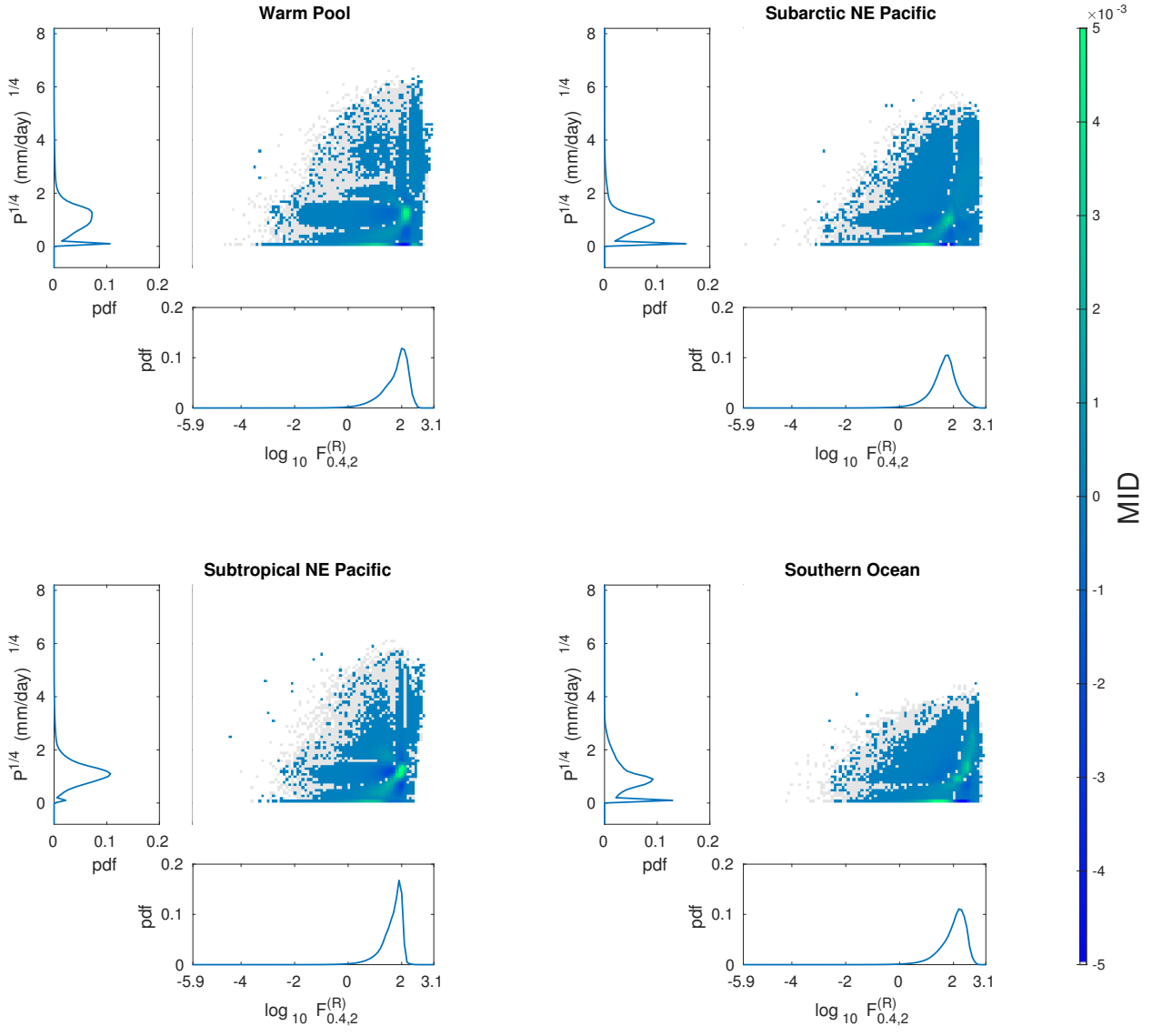


FIG. 4. As in Figure 3 for FV3.

stronger convection (higher P_N, cyan and yellow in the middle row) mirroring larger $\varepsilon_{N,n}$ (yellow in the bottom row). Small-scale wind variations due to proximity to nearby land masses are also evident in the distribution of $\varepsilon_{N,n}$ (Blein et al. 2020). Such relationships between $F^{(R)}_{N,n}$, P_N, and $\varepsilon_{N,n}$ are prominent along the equator in the Indian Ocean.

The marginal probability densities for precipitation and resolved-scale normalized flux, $p(P_{N,n})$ and $p(F^{(R)}_{N,n})$ respectively, and their joint probability density $p(P_{N,n}, F^{(R)}_{N,n})$, can be used to compute

the mutual information density

$$\mathrm{MID}(P_{N,n}, F^{(R)}_{N,n}) = p(P_{N,n}, F^{(R)}_{N,n}) \log\left(\frac{p(P_{N,n}, F^{(R)}_{N,n})}{p(P_{N,n}) \cdot p(F^{(R)}_{N,n})}\right) \tag{16}$$

which characterizes statistical dependence by indicating whether a joint probability density is larger or smaller than the density resulting from the product of the marginal probabilities, a feature which arises when random variables are dependent. In particular, a positive mutual information density in a region indicates a higher probability than if the variables were independent, whereas a negative mutual information density indicates a smaller probability than would be the case for independent variables.

The marginal probability distributions and mutual information densities for the regions considered and two models (IFS and FV3) are shown respectively in Figures 3 and 4. The densities are estimated by aggregating data over the geographical region and the first ten days used in the analysis. A 10-day subset of the DYAMOND simulations is used to allow direct comparision to the CASCADE results, and to investigate sensitivity to the time period considered. Statistical significance of the MID is assessed at the five percent significance level, using the following two-tailed test. Realizations from the null hypothesis distribution of independent $(P_{N,n}, F^{(R)}_{N,n})$ are generated via resampling, i.e. from 2000 random permutations of the sets P_N and $F^{(R)}_{N,n}$. Such resampling preserves marginal distributions while removing dependence between variables. The null hypothesis of independence is rejected when the observed $\mathrm{MID}(P_N, F^{(R)}_{N,n})$ is below the 2.5 percentile or above the 97.5 percentile of the MID of the null hypothesis distribution. Such statistically significant MID values are indicated in color.

From the MID shown in Figures 3 and 4, the two predictors P_N, $F^{(R)}_{N,n}$ are not linearly dependent. Although the absence of such dependence between the two predictors is not strictly necessary for their use in the regression model, we nevertheless are able to confirm that either of the two input variables are not simple functional transformations of the other variable. For the Indo-Pacific Warm Pool, the MID exhibits lower probability densities for bands of $(F^{(R)}_{N,n}, P_N)$ with moderate resolved-scale normalized flux and high precipitation rate as well as for moderate precipitation rate and small resolved-scale normalized flux. While this fact indicates that precipitation and resolved-scale normalized flux are not completely independent, the fact that the joint probability

distribution indicates the two variables are also not linearly dependent justifies their construction as separate predictors for the purpose of constructing a regression model for $\mu_{N,n}$. The structure of the statistical dependence between these two predictors is similar in the IFS and FV3 models, as well as in all other models analyzed (Endo 2022).

Turning to the marginal distribution of precipitation for the Indo-Pacific Warm Pool, of particular note are the two peaks indicating a strong separation between a large number of space-time points with no precipitation and a large number of points with appreciable precipitation. That is, the models consistently distinguish between precipitating and non-precipitating states. Despite broad similarity in the marginal distributions of P_N between the IFS and FV3 models, differences are evident. For example, the fraction of space/time points with zero precipitation is smaller in FV3 than IFS. Again, these differences could result from true model differences or from different weather realizations in the two models. The marginal distributions of resolved-scale normalized flux are similar between these models in the Indo-Pacific Warm Pool.

For both IFS and FV3, the MID for the subtropical Northeast Pacific is similar to that of the Indo-Pacific Warm Pool, demonstrating comparable dependence between wind and precipitation despite the different dominant meteorological processes in the regions reflected in the marginal distributions. For example, the steady wind in this region manifests itself in the resolved-scale normalized flux marginal distribution, presenting a narrower probability distribution when compared against the resolved-scale normalized flux marginal probability distribution in the Indo-Pacific Warm Pool. As with the Warm Pool, a large number of grid points are absent of precipitation. Again, the MID of $F_{N,n}^{(R)}$ and P_N is similar between models despite differences in the marginal distributions.

During the simulation period, the FV3 subarctic Northeast Pacific region is largely absent of precipitation (as indicated by the 'spike' of zero precipitation less prominent in the other two regions), unlike what is observed in the Warm Pool and subtropical Northeast Pacific. Despite the different marginal distributions, in particular smaller values of the precipitation rate and narrower resolved-scale normalized flux marginal distribution, the mutual information density plot is again similar to that of the other domains and between models.

Finally, the stronger winds in the Southern Ocean region are reflected in the marginal distribution for resolved-scale normalized flux, shifted to larger values when compared against the other regions.

Relative to the other regions in the FV3 output, the larger precipitation rate mode corresponds to smaller values, but this feature is not seen in the IFS model output. Once again, despite these differences in the marginal distribution, the mutual information density maintains the features similar to those of the other regions' mutual information densities.

The four regions selected for this study, viz. the Indo-Pacific Warm Pool, subtropical Northeast Pacific, subarctic Northeast Pacific, and a region of the Southern Ocean, offer a variety of wind and precipitation marginal distributions. Despite this fact, the mutual information densities for the regions are remarkably similar across models and regions (for example, a positive MID appears in the vicinity of $\log_{10} F^{(R)}_{N,n} = 2,\ P^{1/4} = 1$ in a consistent manner across all models and regions studied, alongside regions of negative MID flanking the positive signal), indicative of the similar statistical dependence between resolved-scale normalized flux and precipitation. The relationship between the two predictor variables is robust. This robustness is part of the justification of the use of the same predictors in all regions. Inclusion of these regions in the analysis will ensure that a wide range of possible precipitation and resolved-scale normalized flux marginal distributions are considered for the study, thus allowing for a comprehensive evaluation of the robustness of the statistical model.

b. Robustness of the deterministic regression model $\mu_{N,n}$

We now consider the dependence of the SGS normalized flux enhancement $\varepsilon_{N,n}$ on resolved-scale normalized flux and precipitation. A regression fit using Equation 13 generates regression coefficients $A_0, \ldots, B_4$ which are specific to the realized values of the predictor-predictand set $P, F^{(R)}, F^{(T)}$ used in the regression. We examine the robustness of these regression fits across models, geographical regions, and time periods.

The form of the regression model Equation 13, a sum of a quartic polynomial of $P_N^{1/4}$ and a cubic polynomial of $\log_{10} F^{(R)}_{N,n}$, results in the possibility of compensation between regression terms corresponding to different powers. As a result, similar regression functions can be produced with quite different regression coefficients. Therefore, it is not particularly meaningful to directly compare regression coefficients between fits. It is more meaningful to compare computed $\mu_{N,n}(F^{(R)}, P)$ values, as defined in Equation 13, between different regression fits to determine if a regression fit

obtained from one model, region, and time period produces similar values of $\mu_{N,n}$ as a different regression fit.

We measure the robustness of the regression model using the correlation coefficient

$$\mathrm{corr}(\mu(I_i, I_j), \mu(I_j, I_j)) = \frac{\mathrm{mean}((\mu(I_i, I_j) - \mathrm{mean}(\mu(I_i, I_j))) \cdot (\mu(I_j, I_j) - \mathrm{mean}(\mu(I_j, I_j))))}{\mathrm{std}(\mu(I_i, I_j))\ \mathrm{std}(\mu(I_j, I_j))} \tag{17}$$

between pairs of deterministic regression models $\mu(I_i, I_j)$, each fit to a set of inputs I_i, and then applied using a potentially different set of inputs I_j, where I_i, I_j consist of sets of resolved-scale normalized flux and precipitation from different models, regions, and time periods. The second argument of the correlation represents prediction of μ obtained using the model/region/time period used to estimate the statistical model, while the first argument represents the predictions using on a different model/region/time period. A correlation value closer to unity corresponds to a greater degree of robustness between the regression models, since this indicates that a statistical model developed in one setting can be successfully applied to model the relationship between $F^{(R)}_{N,n}$, P, and $\varepsilon_{N,n}$ in another setting. Therefore, high correlation between a wide variety of numerical models, geographical regions, and time periods indicate robustness of the statistical model. In Equation 17 the correlation is evaluated over all spatial and temporal points in the region. The correlation values between deterministic regression values computed using output from different models, different geographical regions, and different 10-day sets of model output, are presented as Figures 5 to 7. The correlation analysis across models is conducted for the Warm Pool region in order to allow comparison with the Cascade simulation. For comparison with Cascade, it should be noted that this simulation was conducted for a different time period than the DYAMOND simulations.

A number of inferences can be made from inspection of the correlations of the regression fits. Correlations between different models in the Warm Pool (Figure 5) indicate the following:

1. The majority of correlation values across models range between 0.70 and 0.98, indicating a reasonably high degree of robustness between models, albeit imperfect.

2. FV3 and GEOS have the highest mutual correlations. These models share the same dynamical core, and the same planetary boundary layer parameterization. It is not surprising to find that regression fits are most robust between different models with somewhat different parameterizations but identical dynamical cores.

3. NICAM and Cascade have the lowest correlation coefficients with other models (between 0.35 and 0.78) although their mutual correlation values are rather large (in the 0.95 range). This fact is noteworthy given the different time periods of the two simulations, as well as the fundamental differences between the two models. There appears to be no obvious similarity between the models' dynamical cores or their physics parameterizations.

Correlations between regression models' fits for different models over different geographical locations are reported in Figure 6. The correlations indicate a reasonably high degree of robustness across regions. Note that the correlations for given model pairs are not symmetric, i.e. model i variables used to calculate normalized flux enhancements using model j and vice versa are expected to have different correlation values in general.

Again focusing on the Indo-Pacific Warm Pool, Figure 7 presents the correlation coefficients for different time periods of the DYAMOND simulation (August 11–20, 21–30, and August 31 – September 9). All correlation values are above 0.96, indicating a high degree of robustness in time, and indicate that any ten days' worth of hourly model output is sufficient to produce a regression fit for the Warm Pool domain. Such high correlations are similarly observed in the other regions as well (Endo 2022).

Graphical representations (in the form of joint probability distributions) of the correlation matrices presented as Figures 5–7 are useful when evaluating the magnitude and spread of $\mu_{N,n}$ parameterizations, and are included as Supplementary Figures S1–S3. While the probability density distribution contours in estimate-target space are not in all cases symmetric across the 1:1 line, the deviation of estimated $\mu_{N,n}$ from the 'true' $\mu_{N,n}$ appears to be largely limited. As well, comparable ranges of values are predicted by the $\mu_{N,n}$ functions. The fact that a number of the joint distributions cluster around low-dimensional curves is suggestive of a possible imperfection in the specific form of regression that was used to generate the $\mu_{N,n}$ estimates. These results indicate a reasonable, albeit imperfect, robustness of the regression model of Bessac et al. (2019, 2021).

c. Covariance structures for $\psi_{N,n}$

We now consider the statistics of the regression residual $\psi_{0.4^\circ,2}$, representing that part of the variability of $\varepsilon_{0.4^\circ,2}$ that is not captured by the regression on resolved normalized flux and precipitation and which we model as a space-time stochastic process. As discussed in detail in Bessac

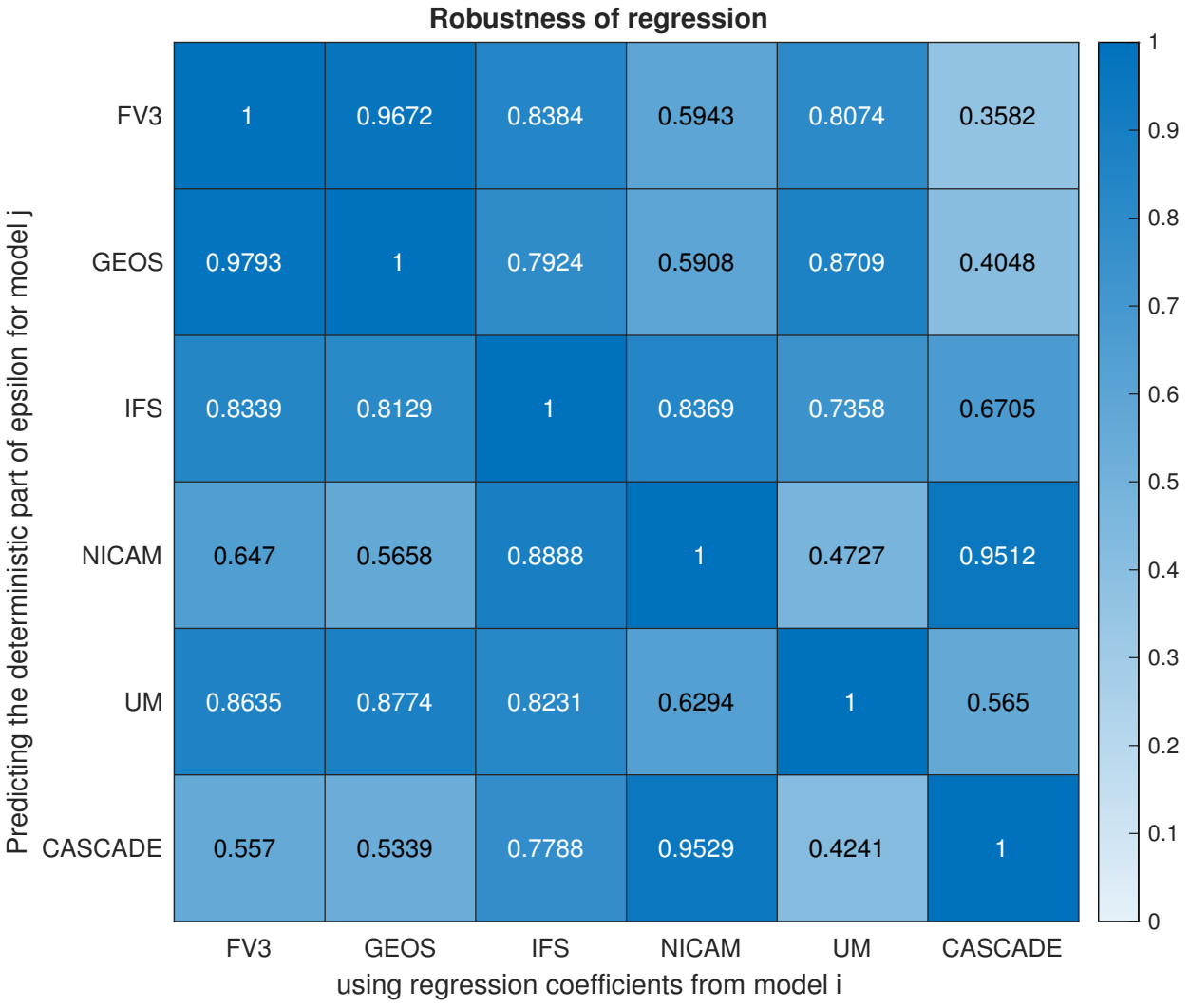
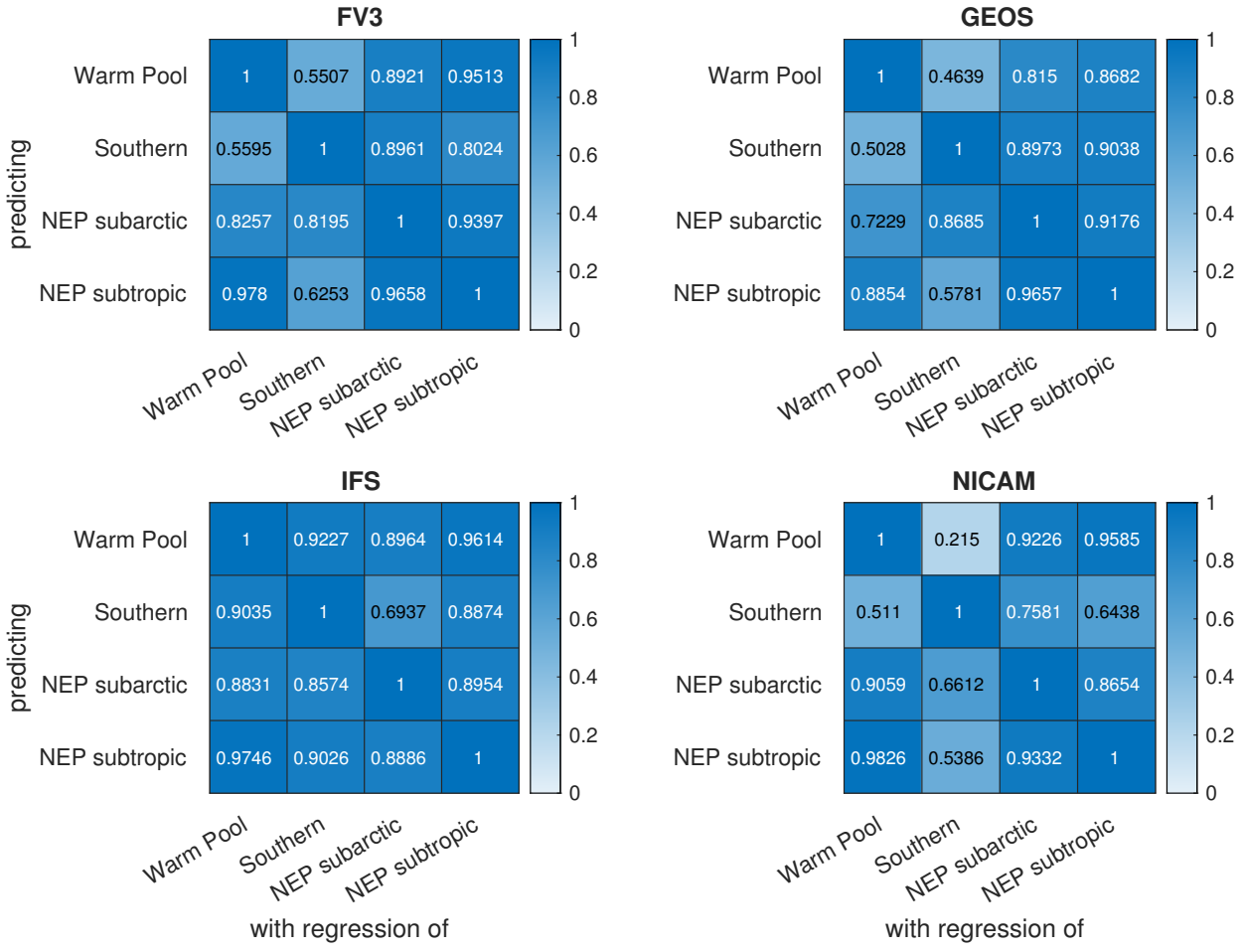


FIG. 5. Correlations computed over the included grid-boxes, of the deterministic part of the normalized flux enhancement for the Indo-Pacific Warm Pool, for $N = 0.4°$ and $n = 2$. The regression fit obtained from the model in the horizontal axis is applied to the variables obtained from the model in the vertical axis, and is compared to $\mu_{0.4°,2}$ obtained from the regression and variables of the model in the vertical axis. Black type is used for correlations lower than 0.73 and white for those higher than 0.73.

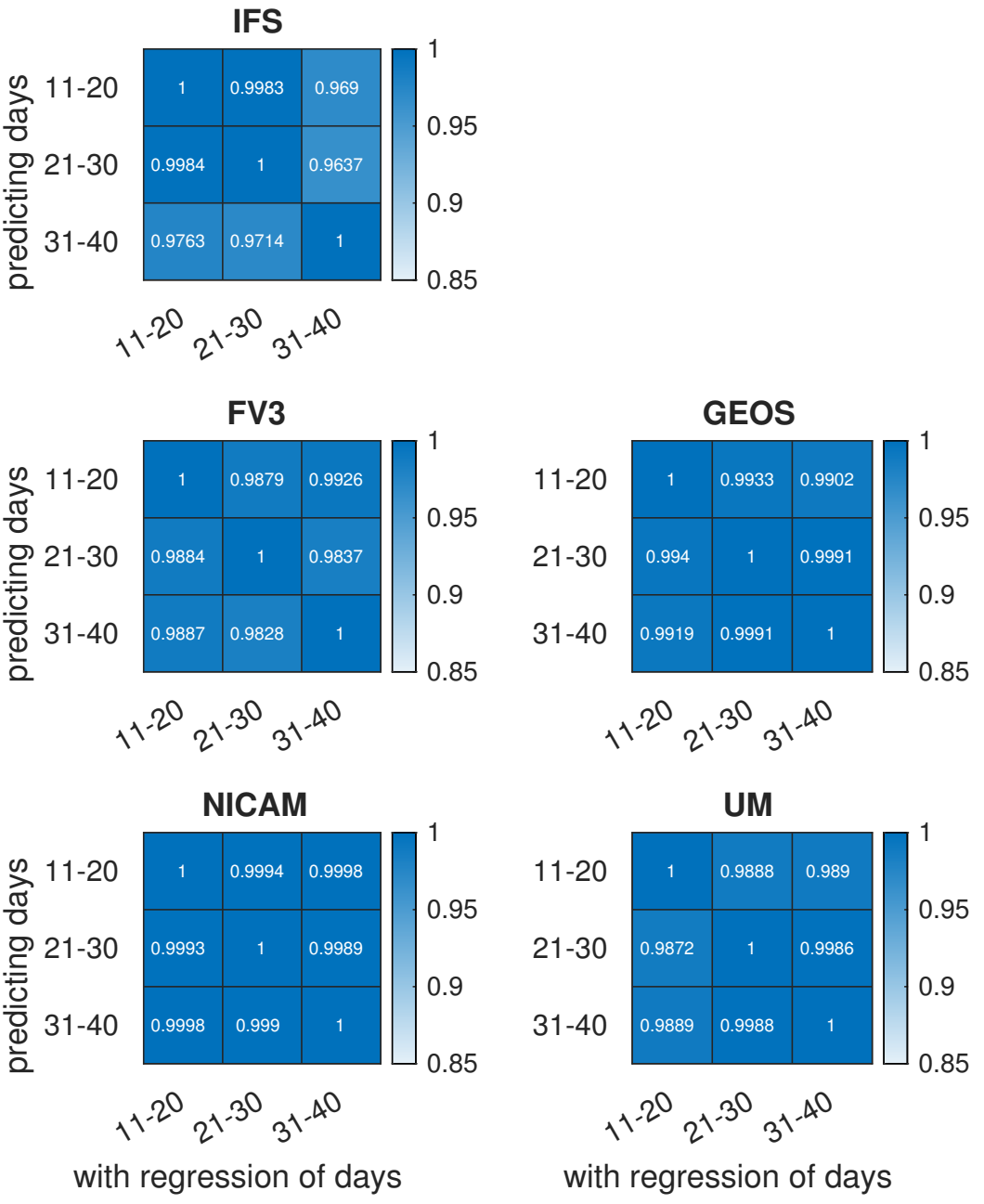
et al. (2019), the $\psi_{0.4°,2}$ field will display correlations in both space and time. Before fitting this field to the covariance structure described in Section 2d, we consider the space/time structure of the correlations in the different regions and models.

Domain-mean spatial autocorrelation functions for the models studied are presented in Figure 8. Rather than showing the entire correlation function, zonal and meridional cross-sections are shown.



439 FIG. 6. Correlations computed over the included grid-boxes, of the deterministic part of the normalized flux
440 enhancement for robustness across geographical regions, for each global model space with $N = 0.4°$ and $n = 2$.
441 The regression fit obtained from the region in the horizontal axis is applied to the variables obtained from the
442 region in the vertical axis, and is compared to $\mu_{0.4°,2}$ obtained from the regression and variables of the region in
443 the vertical axis.

459 As found in Bessac et al. (2019, 2021) the spatial autocorrelation shows a zonal elongation. This
460 feature is present in all geographical regions studied, with larger decay scales in the Southern
461 Ocean. Notwithstanding modest quantitative differences, general profiles of decay are similar
462 across models, with no other identifiable feature present in the other regions studied. Characteristic
463 length scales for spatial decay appear to be smaller for IFS and NICAM in the Indo-Pacific
464 Warm Pool and the subtropical Northeast Pacific, but are comparable with FV3 and GEOS in



444 FIG. 7. Correlations computed over the included grid-boxes, of the deterministic part of the normalized flux
445 enhancement for three different ten-day periods, for Indo-Pacific Warm Pool with $N = 0.4°$ and $n = 2$. The
446 regression fit obtained from the time period in the horizontal axis is applied to the variables obtained from the
447 time period in the vertical axis, and is compared to $\mu_{0.4°,2}$ obtained from the regression and variables of the time
448 period in the vertical axis. Note that the colorbar has a higher lower bound compared to Figures 5 and 6, for the
449 purpose of making the lower-correlation combinations more readily identifiable.

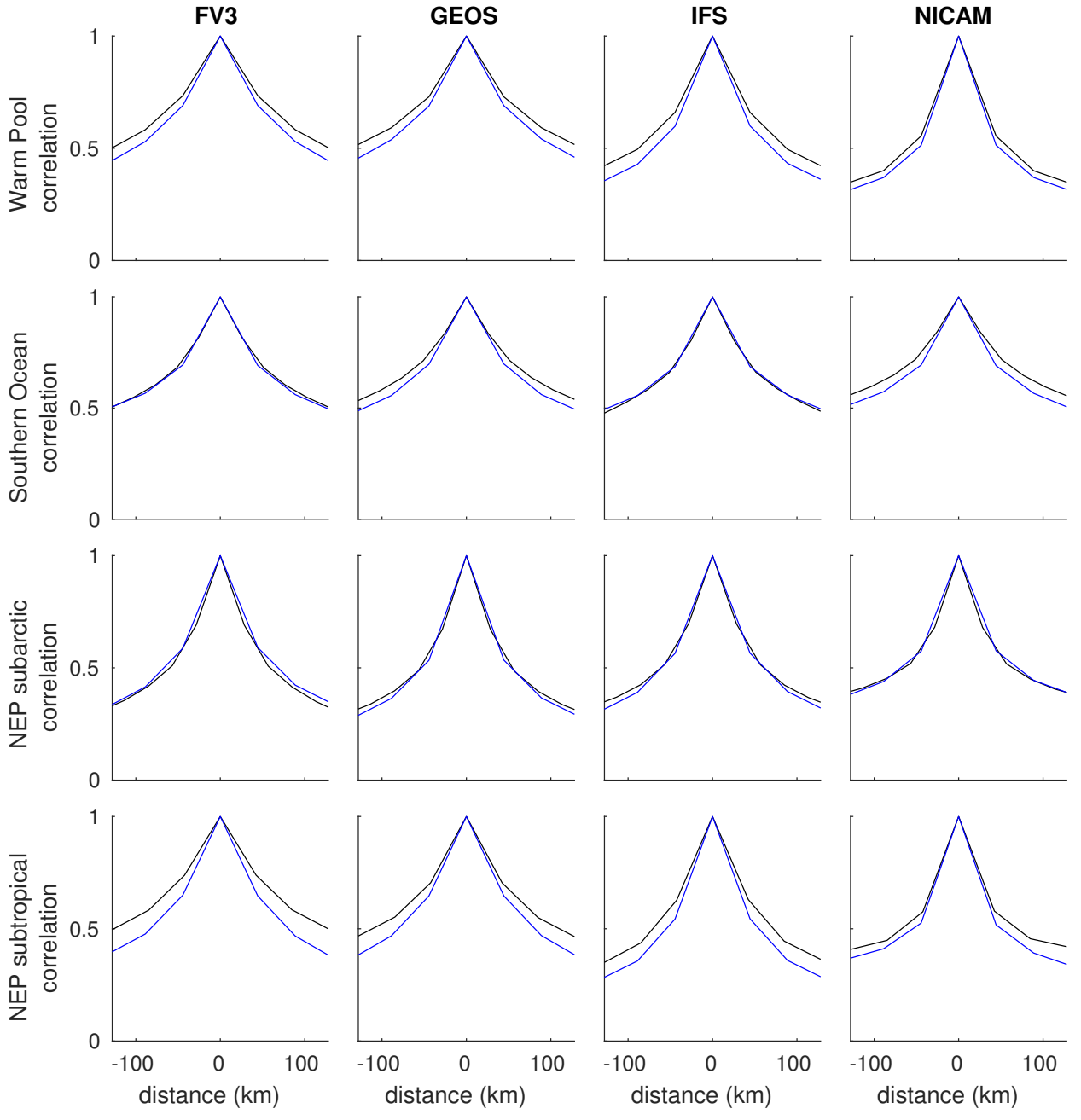


FIG. 8. Domain mean spatial correlations for ψ in the zonal (black) and meridional (blue) directions, with $N = 0.4°$, $n = 2$. Note that the correlation functions are not necessarily symmetric about zero.

the subarctic Northeast Pacific and the Southern Ocean domains, for the choice of regional and temporal boundaries chosen for the present study.

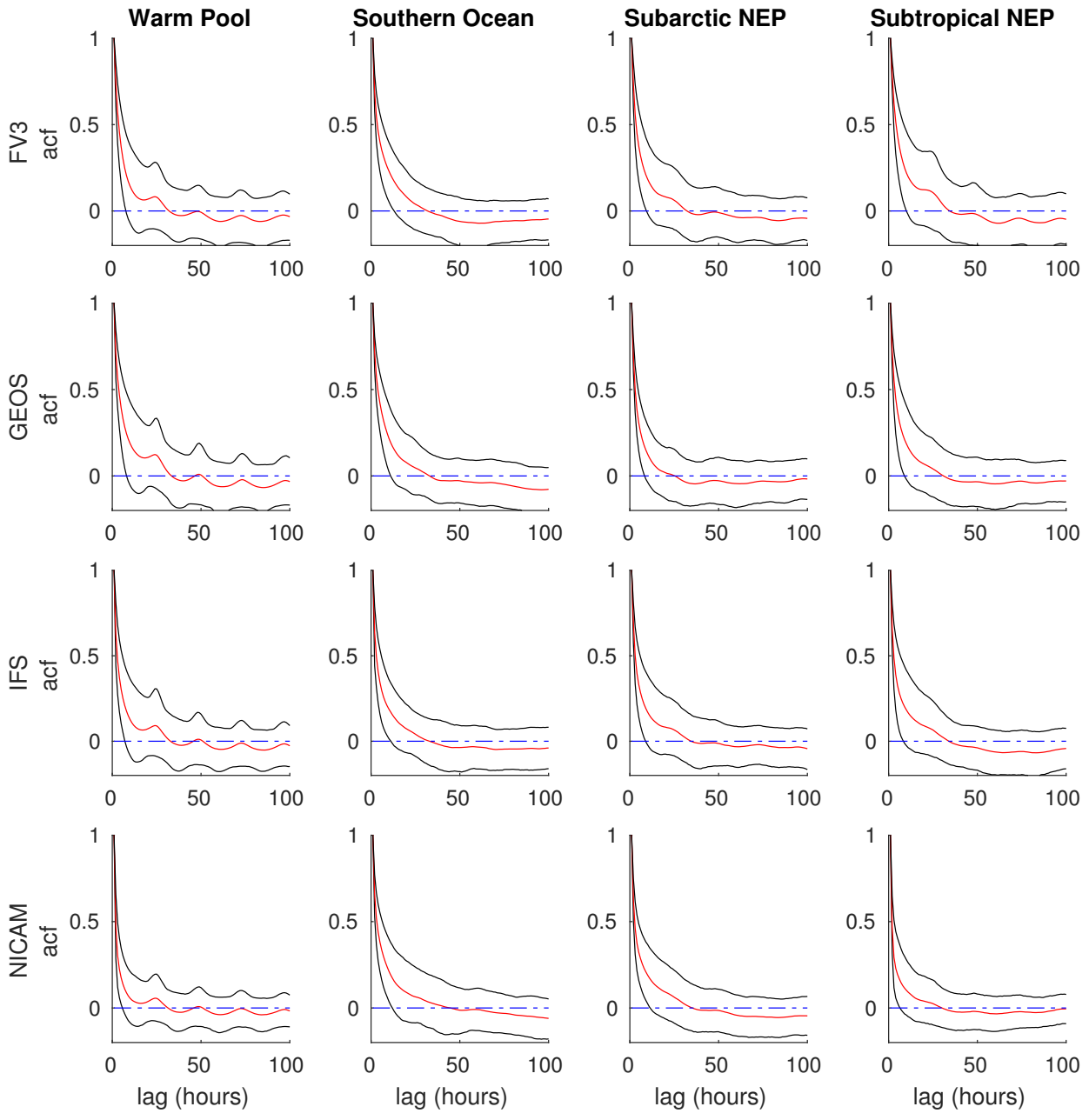


FIG. 9. Temporal autocorrelation functions for various ψ fields with $N = 0.4°$, $n = 2$. Red indicates the median, while the black lines indicate the interdecile range (10 and 90 percentiles) for points across the domain.

The temporal autocorrelation function (acf) of the ψ field is the correlation between $\psi_{N,n}$ at some given time and $\psi_{N,n}$ at some lag time, normalized such that the acf at zero lag is unity. As with the spatial correlation function, we present the acf averaged over all base points in the domain

(Figure 9). We also present the 10th and 90th percentiles of the acf for points across the domain. A diurnal cycle is evident in the Warm Pool domain but largely absent in all other regions, suggesting the presence of a missing physical predictor of $\varepsilon_{N,n}$. We will discuss this point in greater detail later in Section 4. There is no other major substantial difference between temporal autocorrelation functions for different geographical regions, with the decay scale and interdecile range comparable in scale across models and regions.

We model $\psi_{N,n}$ as a Gaussian random field with a covariance model that captures the three-dimensional decay in the correlation function. Following Bessac et al. (2019, 2021), we consider the covariance function defined as per Equation 14. We do not model a diurnal cycle in the covariance function as it appears consistently across models only in the Warm Pool region. First guess values used to obtain the fit for the Gaussian process regression θ_i coefficients used the zonal distance, meridional distance, and temporal lag time such that the correlation at the distance or lag is equal to e^{-1}. The coefficients γ and σ were initialized with constants 1 and 0.3 respectively. Parameter fits were carried out in small, non-overlapping windows moved across the spatial domain. Estimates of θ_x, θ_y, and θ_t, and γ for the Warm Pool region from FV3 and IFS output are presented as Figure 10.

Despite the general similarity of values across the domain, the estimated Gaussian process parameters display spatial variability. This spatial variability of the parameters may result from sampling variability, or may reflect actual spatial inhomogeneity of the parameters. By comparing the spatial patterns of the estimated Gaussian process parameters obtained from different numerical models, it is possible to investigate if any true spatial heterogeneity in the statistics is present, since sampling variations due to internal atmospheric variability should not be correlated across fits for different numerical models. In this regard, it should be noted that the DYAMOND contributions are forced with identical SST and as such this forcing may introduce systematic, correlated inhomogeneities if the regression model does not completely account for the deterministic dependence of $\varepsilon_{N,n}$ on the resolved state. Spatial correlation coefficients between models for the Gaussian parameter fit fields in the Warm Pool are presented in Figure 11, and indeed demonstrate the presence of systematic spatial heterogeneity: this observation originates from the relatively high correlations seen across different model combinations presented in Figure 11, with correlation values for most combinations being over 0.5. Such heterogeneity is also found in the subarctic Northeast Pacific,

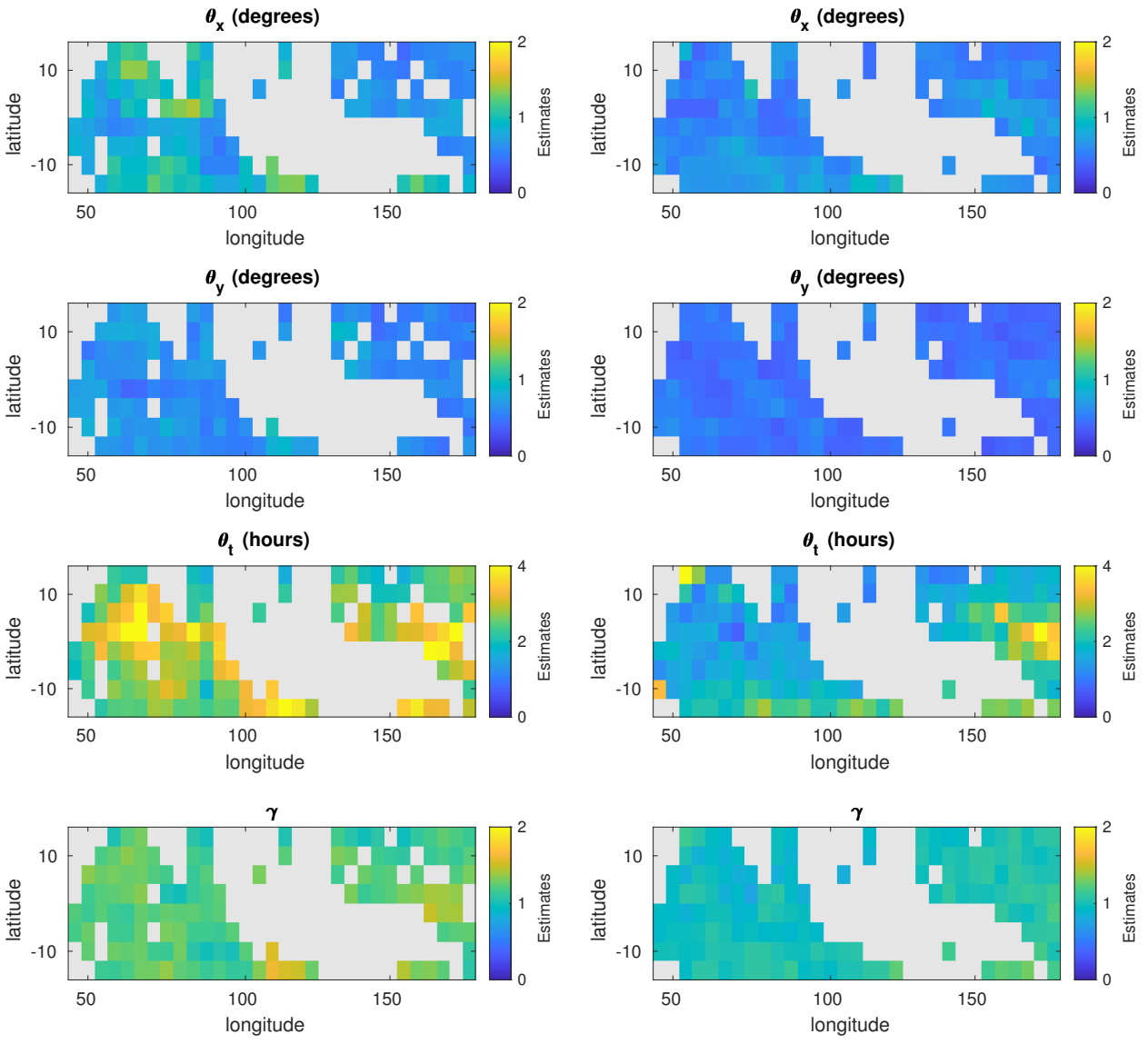


FIG. 10. Maps of Gaussian process parameter fits for FV3 (left) and IFS (right) in the Indo-Pacific Warm Pool with $N = 0.4°$ and $n = 2$, for August 11–20. The grey mask denotes land, the vertical axis denotes latitude, and the horizontal axis denotes longitude.

although absent in the subtropical Northeast Pacific and the Southern Ocean (Endo 2022). The roughness parameter displays higher correlations than the decay length scales, indicating consistent spatial heterogeneity across different models.

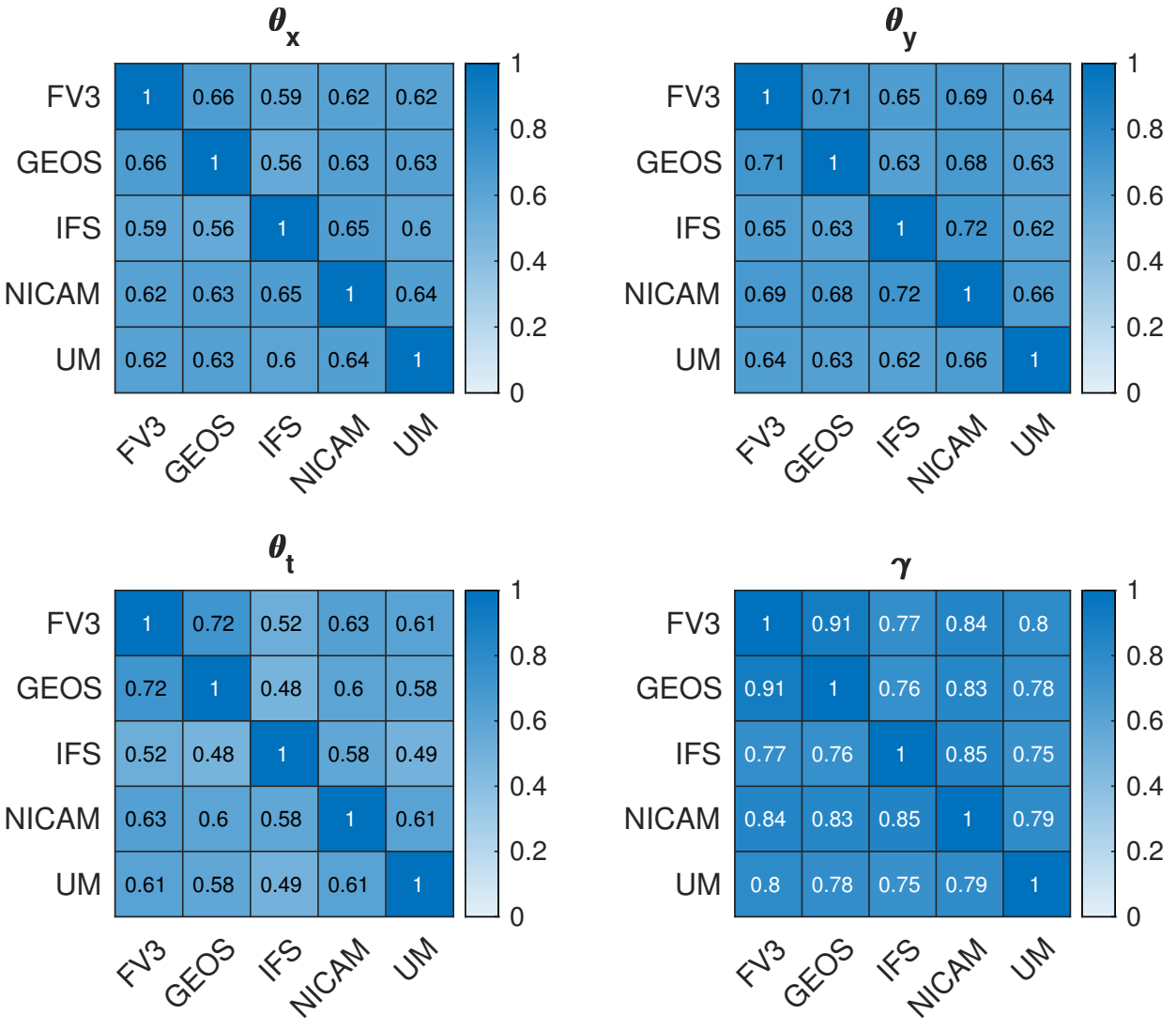


FIG. 11. Correlations between estimated Gaussian process fit parameters for different numerical models. Results for the Indo-Pacific Warm Pool are presented.

d. Different resolutions N and normalized flux exponents n

Endo (2022) reports similarities between the spatial and temporal covariance structures for ψ with $n = 1$ and 2, as well as a striking similarity between results for $N = 0.4°$ and $N = 0.76°$. Consistent with the results of Bessac et al. (2019), the correlation between the deterministic part of the normalized flux enhancement is higher for $n = 1$ than for $n = 2$. These results further support the robustness of data driven stochastic parameterization for operational weather and climate models.

4. Discussion

a. Implications

This study has considered the robustness to changes of base model, geographic region, and time period of a data-driven parameterization of subgrid-scale velocity enhancements of sea surface fluxes. A regression analysis using resolved-scale normalized flux and precipitation was used to model the field

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

measuring the difference between true ($F^{(T)}_{N,n}$) and resolved-scale ($F^{(R)}_{N,n}$) sea surface normalized fluxes due to subgrid-scale velocity variations. We have investigated the robustness of both the regression model as well as the covariance structure of the stochastic residual $\psi_{N,n}$. A Gaussian process was fit to $\psi_{N,n}$, in order to model the spatial and temporal correlation structures present in this field. We have shown that the deterministic part of the normalized flux enhancement regression is broadly robust to the high-resolution model output used to fit the parameterization. It is similarly broadly robust to the use of data from different geographical regions and time periods. However, the robustness of the statistical model is weaker for some numerical models, as was the case with NICAM, and for some geographical regions, as was observed for the Southern Ocean. Similarly, the statistical features of the stochastic residual are also robust and are largely independent of the underlying model, region, or time period.

Although the fine-scale model output is by construction treated as if they generate an ideal 'true' surface wind speed, in fact these fields will be influenced by the specific parameterizations used in the models (particularly the boundary layer schemes). Furthermore, because of the finite resolution of the models, there are SGS wind variations on smaller scales that are not accounted for in the coarse graining approach. The relatively high degree of robustness of the parameterization across models indicates a modest sensitivity to the details of the boundary layer schemes and differences in model resolution.

The robustness of the regression model was quantified by use of correlations between regression model predictions based on the data used to estimate the model and predictions using input data from other models, regions, or time periods. It was found that the deterministic part of the normalized flux enhancement $\mu_{N,n}$ can be estimated with a reasonable degree of imperfect

robustness for many different models while using a statistical model developed using one numerical model. The presence of some relatively small correlation values (less than 0.5) indicates that the regression fit does not generalize perfectly in all cases. In particular, the regression fit encountered specific difficulties with generalizability involving the Southern Ocean, albeit this result is model-dependent. Probability density functions of the data used for these $\mu_{N,n}$ regressions are suggestive that limitations of generalizability are partly a consequence of the form of regression used in the present analysis. Regression models based on different 10-day periods are close to identical.

The random spatiotemporal field $\psi_{N,n}$ has similar but not identical statistical features across different models and regions. Similarly the temporal autocorrelation functions in all models and regions attenuate with comparable time scales but only the Warm Pool region displays a diurnal cycle. Such a deviation is suggestive of a missing predictor for the regression model, providing an indication of potential for further regression model improvement.

If the random spatiotemporal field $\psi_{N,n}$ truly has spatially homogeneous statistics, as would be expected if the regression model completely captured the deterministic dependence of the SGS normalized flux enhancement on the resolved-scale state, any apparent spatial variations in the obtained Gaussian process fits will be due to sampling variability. However, it was determined that in the Indo-Pacific Warm Pool and the subarctic Northeast Pacific, spatial variations in Gaussian process fits between different models were correlated and thus there exists some spatial structure that may not be caused by sampling variability. It is possible that the functional form of the regression model does not completely represent the dependence of $\varepsilon_{N,n}$ on $F^{(R)}_{N,n}$ and P_N. Such change to the regression, be it by the addition of a predictor or the use of a different parametric form of the regression fit, will change the spatial and temporal structure of the residual field $\psi_{N,n}$.

Despite the regression model being most robust between the two models that share dynamical cores, different models with different dynamical cores appeared to have high correlation coefficients as well. The DYAMOND model producing the least robust regression model among other DYAMOND models was NICAM, which differs the most from the other models by virtue of explicitly representing convective processes parameterized in the other numerical models. In contrast, the correlation coefficients between NICAM and Cascade, another simulation with convection (mostly) represented at resolved scale rather than by use of parameterization, were notably high. The degree of parameterization of cumulus convection may be a contributing factor to model

sensitivity of the regression fit. This interpretation is consistent with the fact that the statistical fit utilizes precipitation as a proxy of deep convection and small-scale surface convergence, and this quantity is expected to be influenced by the cumulus parameterization of a given numerical model.

The fact that the empirical parameterization is broadly - but not perfectly - robust across models, regions, and time periods indicates that the appropriate approach to developing 'operational' parameterizations would be to combine these various data sources to produce a single, unified empirical fit.

b. Future extensions

This study investigated the robustness of the statistical model of Bessac et al. (2019) used to represent the SGS wind variation enhancement of normalized sea surface flux. A number of questions have emerged that are relevant to future research on introducing stochastic sea surface flux enhancements into operational weather and climate models. The important fact uncovered by the present study is that an empirical parameterization of sea surface flux enhancement obtained from some numerical model can be applied to a different numerical model or to a different geographical region with the expectation that the resulting normalized flux enhancement will be reasonably accurate. This finding is useful when envisioning the next steps for flux representation in contemporary numerical weather prediction and climate modelling.

The use of a combination of polynomial fits, as was done in the present study, may not be the best statistical model for capturing the dependence of $\varepsilon_{N,n}$ on $F_{N,n}^{(R)}$ and P_N, and may have contributed to the limitations identified in the present work. As such, the investigation of different forms of the statistical model, as well as investigation into other predictors for the regression fit, may improve the statistical model. For example, Blein et al. (2022) use a formal predictor-selection procedure to determine optimal regression models among a large range of possible predictors, including ones that would not necessarily be available in a coarse resolution model. They note the inability of the parameterization to generate values of low gustiness velocities implied by the parameterization of Bessac et al. (2019), where low gustiness requires strongly negative $\varepsilon_{N,n}$ values: by virtue of the Gaussian distribution of $\varepsilon_{N,n}$, occurrences of such low $\varepsilon_{N,n}$ are expected to be few in number. This feature again supports exploration of additional parameters to improve the regression, or alternatively, a different distribution for $\psi_{N,n}$ with fatter tails. The use of DYAMOND output

may be beneficial for this purpose, due to the large number of variables reported by participating models. In addition to the benefits discussed above, a model selection process similar to what is employed by Blein et al. (2022) can also be used to simplify the functional form of any regression model that may be used in a given parameterization. An improved statistical model for $\varepsilon_{N,n}$ should reduce the systematic spatial structure of the Gaussian process fits, resulting in a spatially uniform model such as would be needed for use in operational weather and climate models.

An alternative approach which avoids the need to specify a parametric form of the regression fit would be the use of machine learning techniques to effectively and efficiently represent the statistical dependences needed for parameterization of the normalized sea surface flux enhancement. Generative adversarial networks (GANs), for example, may be utilized in this direction (Gagne et al. 2020). In particular, the ability of GANs to draw realizations from high dimensional distributions without explicit expressions for the probability distribution should allow for investigation of a better functional form of the regression fit, without the need to specify its form explicitly. An important consideration for such an approach would be the necessity to ensure consistent generation of statistical models and realizations, since a number of fluxes, each with different (and possibly noninteger) values of n, are computed simultaneously by the numerical model.

Further numerical simulations at even higher resolution may also be beneficial to identifying better statistical fits. Use of a finer scale model can determine at which point unresolved flow ceases to contribute meaningfully to the surface fluxes. Larger training sets should also reduce risk due to numerical model extrapolation, since the regression model fit has the potential to produce unreasonable results outside of input variable values encountered during the model fit. Nevertheless, the results obtained in this study with the use of a regression fit to one model output applied to a different model output finds no evidence of consequential extrapolation errors.

Finally, it is important to note that surface flux parameterization is not an issue only for the sea surface. As such, an equivalent study should be conducted on the land surface as well. The number of additional parameters that may be involved in such parameterization, however, can be expected to make such study nontrivial due to additional factors such as topography and other fixed surface heterogeneities.

5. Conclusions

The present study utilized high-resolution global atmospheric model output from the DYAMOND initiative to statistically model the enhancement of normalized sea surface flux due to sub-grid scale wind variations, expanding on the works by Bessac et al. (2019, 2021). In particular, the high resolution vector wind field was coarsened to obtain a wind vector field approximating that in a operational global atmospheric model. Then, the differences between the sea-surface wind dependent part of the normalized sea surface flux in high resolution and coarse resolution wind fields were studied, in order to determine how the 'true' normalized flux from the higher resolution model field can be represented using the 'resolved-scale' coarse resolution field. The resolved-scale normalized surface flux and the resolved-scale precipitation were used as variables for a polynomial fit to compute the deterministic part of the normalized flux enhancement and the regression residual was fit to a space-time Gaussian process. We investigated the robustness of the statistical model across a range of different high-resolution models, geographical regions, and time periods.

We demonstrated that the features of both the regression fit and the stochastic residual are broadly robust, albeit imperfect. This finding supports the application of the methods in a robust manner across all oceans on the globe without major region specific or model specific features. As such, data-driven development of an explicitly stochastic flux enhancement model is shown to be an attractive method for the climate modelling community. A robust spatial heterogeneity in the statistics of the Gaussian process fit was found to be present in the Indo-Pacific Warm Pool and the subarctic Northeast Pacific. This fact suggests potential improvement of the regression fit by the use of additional predictors or by use of a different form of statistical model.

Acknowledgments. The authors thank the coordinators of the DYAMOND project for allowing us access to the model output used in the present study. DYAMOND data management was provided by the German Climate Computing Center (DKRZ) and supported through the projects ESiWACE and ESiWACE2. The projects ESiWACE and ESiWACE2 have received funding from the European Union's Horizon 2020 research and innovation program under grant agreements No 675191 and 823988. This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project IDs bk1040 and bb1153. The authors thank Roberta Hamme, Bill Merryfield, and Judith Berner for their comments on an earlier version of this manuscript. Kota Endo and Adam Monahan acknowledge the support of the Natural Sciences and

Engineering Research Council of Canada (NSERC) (funding reference RGPIN-2019-204986). The effort of Julie 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, and under Scientific Discovery through Advanced Computing (SciDAC) program through the FASTMath Institute under Contract DE-AC02-06CH11357. Hannah M. Christensen was funded by UK Natural Environment Research Council grant number NE/P018238/1. Nils Weitzel acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project no. 395588486). The authors thank three anonymous reviewers whose comments greatly improved this manuscript.

Data availability statement. Cascade data may be obtained from the Centre for Environmental Data Analysis (CEDA) Archives, part of the United Kingdom Natural Environment Research Council (NERC) Environmental Data Service (EDS), viz. from the dataset published as Lister and Woolnough (2008). DYAMOND data (Stevens et al. 2019) may be obtained from the ESiWACE coordination team.

References

Alexander, M., and T. Dunkerton, 1999: A spectral parameterization of mean-flow forcing due to breaking gravity waves. *J. Atmos. Sci.*, **56 (24)**, 4167–4182.

Bacmeister, J., M. J. Suarez, and F. R. Robertson, 2006: Rain reevaporation, boundary layer–convection interactions, and Pacific rainfall patterns in an AGCM. *J. Atmos. Sci.*, **63 (12)**, 3383–3403.

Berner, J., and Coauthors, 2017: Stochastic parameterization: Toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98 (3)**, 565–588.

Bessac, J., H. M. Christensen, K. Endo, A. H. Monahan, and N. Weitzel, 2021: Scale-aware space-time stochastic parameterization of subgrid-scale velocity enhancement of sea surface fluxes. *J. Adv. Model. Earth Syst.*, **13 (4)**, e2020MS002 367.

Bessac, J., A. H. Monahan, H. M. Christensen, and N. Weitzel, 2019: Stochastic parameterization of subgrid-scale velocity enhancement of sea surface fluxes. *Mon. Wea. Rev.*, **147 (5)**, 1447–1469.

Blein, S., R. Roehrig, and A. Voldoire, 2022: Parameterizing the meso-scale enhancement of oceanic surface turbulent fluxes: a physical-statistical approach. *Quart. J. Roy. Meteor. Soc.*, **148 (745)**, 1683–1708.

Blein, S., R. Roehrig, A. Voldoire, and G. Faure, 2020: Meso-scale contribution to air–sea turbulent fluxes at GCM scale. *Quart. J. Roy. Meteor. Soc.*, **146 (730)**, 2466–2495.

Chen, J., and M. L. Stein, 2021: Linear-cost covariance functions for Gaussian random fields. *J. Amer. Stat. Assoc.*, 1–18.

Christensen, H. M., 2020: Constraining stochastic parametrisation schemes using high-resolution simulations. *Quart. J. Roy. Meteor. Soc.*, **146 (727)**, 938–962.

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

Davini, P., and Coauthors, 2017: Climate SPHINX: evaluating the impact of resolution and stochastic physics parameterisations in the EC-Earth global climate model. *Geosci. Model Dev.*, **10 (3)**, 1383–1402.

Dawson, A., and T. Palmer, 2015: Simulating weather regimes: Impact of model resolution and stochastic parameterization. *Climate Dyn.*, **44 (7)**, 2177–2193.

Deardorff, D. W., 1970: Preliminary results from numerical integrations of the unstable planetary boundary layer. *J. Atmos. Sci.*, **27 (8)**, 1209–1211.

Dorrestijn, J., D. T. Crommelin, A. P. Siebesma, and H. J. Jonker, 2013: Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theor. Comput. Fluid Dyn.*, **27 (1)**, 133–148.

Endo, K., 2022: Robustness of the parameterization of sub-grid scale wind variability on sea-surface fluxes. M.Sc. thesis, School of Earth and Ocean Sciences, University of Victoria, 240 pp., http://hdl.handle.net/1828/13809.

European Centre for Medium-Range Weather Forecasts, 2020a: *IFS Documentation CY47R1 Part III: Dynamics and numerical procedures*. Reading, England, European Centre for Medium-Range Weather Forecasts.

European Centre for Medium-Range Weather Forecasts, 2020b: *IFS Documentation CY47R1 Part IV: Physical processes*. Reading, England, European Centre for Medium-Range Weather Forecasts.

Fagan, K. E., and F. T. Mackenzie, 2007: Air–sea CO2 exchange in a subtropical estuarine-coral reef system, Kaneohe Bay, Oahu, Hawaii. *Mar. Chem.*, **106 (1-2)**, 174–191.

Franco, A. C., and Coauthors, 2021: Anthropogenic and climatic contributions to observed carbon system trends in the Northeast Pacific. *Global Biogeochem. Cycles*, **35 (7)**, e2020GB006 829.

Gagne, D. J., H. M. Christensen, A. C. Subramanian, and A. H. Monahan, 2020: Machine learning for stochastic parameterization: Generative adversarial networks in the Lorenz '96 model. *J. Adv. Model. Earth Syst.*, **12 (3)**, e2019MS001 896.

Garcia, R. R., and B. A. Boville, 1994: "Downward control" of the mean meridional circulation and temperature distribution of the polar winter stratosphere. *J. Atmos. Sci.*, **51 (15)**, 2238–2245.

Garner, S. T., 2005: A topographic drag closure built on an analytical base flux. *J. Atmos. Sci.*, **62 (7)**, 2302–2315.

Godfrey, J., and A. Beljaars, 1991: On the turbulent fluxes of buoyancy, heat and moisture at the air-sea interface at low wind speeds. *J. Geophys. Res.: Oceans*, **96 (C12)**, 22 043–22 048.

Harris, L., X. Chen, L. Zhou, and J.-H. Chen, 2020: *The Nonhydrostatic Solver of the GFDL Finite-Volume Cubed-Sphere Dynamical Core*. National Oceanic and Atmospheric Administration, Geophysical Fluids Dynamics Laboratory.

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

Jabouille, P., J. Redelsperger, and J. Lafore, 1996: Modification of surface fluxes by atmospheric convection in the TOGA COARE region. *Mon. Wea. Rev.*, **124 (5)**, 816–837.

Kodama, C., and Coauthors, 2015: A 20-year climatology of a NICAM AMIP-type simulation. *J. Meteor. Soc. Japan*, **93 (4)**, 393–424.

Kodama, C., and Coauthors, 2021: The Nonhydrostatic ICosahedral Atmospheric Model for CMIP6 HighResMIP simulations (NICAM16-S): experimental design, model description, and impacts of model updates. *Geosci. Model Dev.*, **14 (2)**, 795–820.

Lin, S., and J. Sheng, 2020: Revisiting dependences of the drag coefficient at the sea surface on wind speed and sea state. *Continental Shelf Research*, **207**, 104 188.

Lin, S.-J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132 (10)**, 2293–2307.

Lin, S.-J., and R. B. Rood, 1996: Multidimensional flux-form semi-Lagrangian transport schemes. *Mon. Wea. Rev.*, **124 (9)**, 2046–2070.

Lin, S.-J., and R. B. Rood, 1997: An explicit flux-form semi-Lagrangian shallow-water model on the sphere. *Quart. J. Roy. Meteor. Soc.*, **123 (544)**, 2477–2498.

Lindzen, R. S., 1981: Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.: Oceans*, **86 (C10)**, 9707–9714.

Lister, G., and S. Woolnough, 2008: Cascade - Scale interactions in the tropical atmosphere model runs. NCAS British Atmospheric Data Centre. URL "https://catalogue.ceda.ac.uk/uuid/20981e3052a66ca71c2ba92b94760150".

Lock, A., A. Brown, M. Bush, G. Martin, and R. Smith, 2000: A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Wea. Rev.*, **128 (9)**, 3187–3199.

Lott, F., and M. J. Miller, 1997: A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quart. J. Roy. Meteor. Soc.*, **123 (537)**, 101–127.

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

Mahrt, L., D. Belušić, and O. Acevedo, 2021: Small-scale spatial variation of the nocturnal wind field. *Bound.-Layer Meteor.*, **180 (2)**, 225–245.

Mahrt, L., and J. Sun, 1995: The subgrid velocity scale in the bulk aerodynamic relationship for spatially averaged scalar fluxes. *Mon. Wea. Rev.*, **123 (10)**, 3032–3041.

Malardel, S., N. Wedi, W. Deconinck, M. Diamantakis, C. Kühnlein, G. Mozdzynski, M. Hamrud, and P. Smolarkiewicz, 2016: A new grid for the IFS. *ECMWF Newsletter*, **146 (321)**, 23–28.

McFarlane, N., 1987: The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. *J. Atmos. Sci.*, **44 (14)**, 1775–1800.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20 (4)**, 851–875.

Miller, M. J., A. C. M. Beljaars, and T. N. Palmer, 1992: The sensitivity of the ECMWF model to the parameterization of evaporation from the tropical oceans. *J. Climate*, **5 (5)**, 418–434.

Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8 (5)**, 1339–1356.

Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119 (2)**, 397–407.

Nakanishi, M., and H. Niino, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87 (5)**, 895–912.

Palmer, T., 2019: Stochastic weather and climate models. *Nat. Rev. Phys.*, **1 (7)**, 463–471.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: *Stochastic parametrization and model uncertainty*. Reading, England, European Centre for Medium-Range Weather Forecasts.

Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227 (1)**, 55–78.

Putman, W. M., and M. Suarez, 2011: Cloud-system resolving simulations with the NASA Goddard Earth Observing System global atmospheric model (GEOS-5). *Geophys. Res. Lett.*, **38 (16)**, L16 809.

Räisänen, P., H. W. Barker, M. F. Khairoutdinov, J. Li, and D. A. Randall, 2004: Stochastic generation of subgrid-scale cloudy columns for large-scale models. *Quart. J. Roy. Meteor. Soc.*, **130 (601)**, 2047–2067.

Redelsperger, J.-L., F. Guichard, and S. Mondon, 2000: A parameterization of mesoscale enhancement of surface fluxes for large-scale models. *J. Climate*, **13 (2)**, 402–421.

Rees, A. P., N. J. P. Owens, and R. C. Upstill-Goddard, 1997: Nitrous oxide in the Bellingshausen Sea and Drake Passage. *J. Geophys. Res.: Oceans*, **102 (C2)**, 3383–3391.

Rotstayn, L. D., B. F. Ryan, and J. J. Katzfey, 2000: A scheme for calculation of the liquid fraction in mixed-phase stratiform clouds in large-scale models. *Mon. Wea. Rev.*, **128 (4)**, 1070–1088.

Satoh, M., T. Matsuno, H. Tomita, H. Miura, T. Nasuno, and S.-i. Iga, 2008: Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *J. Comput. Phys.*, **227 (7)**, 3486–3514.

Satoh, M., B. Stevens, F. Judt, M. Khairoutdinov, S.-J. Lin, W. M. Putman, and P. Düben, 2019: Global cloud-resolving models. *Curr. Climate Change Rep.*, **5 (3)**, 172–184.

Satoh, M., and Coauthors, 2014: The Non-hydrostatic ICosahedral Atmospheric Model: Description and development. *Prog. Earth Planet. Sci.*, **1 (18)**.

Scaife, A., N. Butchart, C. Warner, and R. Swinbank, 2002: Impact of a spectral gravity wave parameterization on the stratosphere in the Met Office Unified Model. *J. Atmos. Sci.*, **59 (9)**, 1473–1489.

Scinocca, J. F., 2003: An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *J. Atmos. Sci.*, **60 (4)**, 667–682.

Smith, R., 1990: A scheme for predicting layer clouds and their water content in a general circulation model. *Quart. J. Roy. Meteor. Soc.*, **116 (492)**, 435–460.

820 Stevens, B., and Coauthors, 2019: DYAMOND: the DYnamics of the Atmospheric general circu-
821 lation Modeled On Non-hydrostatic Domains. *Prog. Earth Planet. Sci.*, **6 (61)**.

822 Tiedke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121 (11)**,
823 3040–3061.

824 Tsyrulnikov, M., and D. Gayfulin, 2017: A limited-area spatio-temporal stochastic pattern generator
825 for simulation of uncertainties in ensemble applications. *Meteor. Z.*, **26 (5)**, 549–566.

826 Vidale, P. L., and Coauthors, 2021: Impact of stochastic physics and model resolution on the
827 simulation of tropical cyclones in climate GCMs. *J. Climate*, **34 (11)**, 4315–4341.

828 Vosper, S., 2015: Mountain waves and wakes generated by South Georgia: Implications for drag
829 parametrization. *Quart. J. Roy. Meteor. Soc.*, **141 (692)**, 2813–2827.

830 Walters, D., and Coauthors, 2017: The Met Office Unified Model Global Atmosphere 6.0/6.1 and
831 JULES Global Land 6.0/6.1 configurations. *Geosci. Model Dev.*, **10 (4)**, 1487–1520.

832 Wanninkhof, R., 2014: Relationship between wind speed and gas exchange over the ocean revisited.
833 *Limnol. Oceanogr.: Methods*, **12 (6)**, 351–362.

834 Warner, C., and M. McIntyre, 1999: Toward an ultra-simple spectral gravity wave parameterization
835 for general circulation models. *Earth Planets Space*, **51 (7-8)**, 475–484.

836 Webster, S., A. Brown, D. Cameron, and C. Jones, 2003: Improvements to the representation of
837 orography in the Met Office Unified Model. *Quart. J. Roy. Meteor. Soc.*, **129 (591)**, 1989–2010.

838 Wedi, N. P., 2014: Increasing horizontal resolution in numerical weather prediction and climate
839 simulations: illusion or panacea? *Philos. Trans. Roy. Soc., A*, **372 (2018)**, 20130 289.

840 Williams, P. D., 2012: Climatic impacts of stochastic fluctuations in air–sea fluxes. *Geophys. Res.*
841 *Lett.*, **39 (10)**, L10 705.

842 Wilson, D. R., and S. P. Ballard, 1999: A microphysically based precipitation scheme for the UK
843 Meteorological Office Unified Model. *Quart. J. Roy. Meteor. Soc.*, **125 (557)**, 1607–1636.

844 Wilson, D. R., A. C. Bushell, A. M. Kerr-Munslow, J. D. Price, and C. J. Morcrette, 2008: PC2:
845 A prognostic cloud fraction and condensation scheme. I: Scheme description. *Quart. J. Roy.*
846 *Meteor. Soc.*, **134 (637)**, 2093–2107.

Wood, N., and Coauthors, 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quart. J. Roy. Meteor. Soc.*, **140 (682)**, 1505–1520.

Zadra, A., and Coauthors, 2018: Systematic errors in weather and climate models: Nature, origins, and ways forward. *Bull. Amer. Meteor. Soc.*, **99 (4)**, ES67–ES70.

Zeng, X., Q. Zhang, D. Johnson, and W. Tao, 2002: Parameterization of wind gustiness for the computation of ocean surface fluxes at different spatial scales. *Mon. Wea. Rev.*, **130 (8)**, 2125–2133.

Zhao, M., and Coauthors, 2018: The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *J. Adv. Model. Earth Syst.*, **10 (3)**, 735–769.