



**HAL**  
open science

## Using an ensemble data set of turbulent air-sea fluxes to evaluate the IPSL climate model in tropical regions

Alina Găinușă-Bogdan, Pascale Braconnot, Jérôme Servonnat

### ► To cite this version:

Alina Găinușă-Bogdan, Pascale Braconnot, Jérôme Servonnat. Using an ensemble data set of turbulent air-sea fluxes to evaluate the IPSL climate model in tropical regions. *Journal of Geophysical Research: Atmospheres*, 2015, 120 (10), pp.4483 - 4505. 10.1002/2014JD022985 . hal-01806181

**HAL Id: hal-01806181**

**<https://hal.science/hal-01806181v1>**

Submitted on 28 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## RESEARCH ARTICLE

10.1002/2014JD022985

## Key Points:

- Database of 14 turbulent air-sea flux products is assembled for model evaluation
- Proposed methodology is suited for systematic large-scale model evaluation
- Robust biases in state variables do not result in robust biases in IPSL fluxes

## Correspondence to:

A. Găinușă-Bogdan,  
alina.gainusa-bogdan@lsce.ipsl.fr

## Citation:

Găinușă-Bogdan, A., P. Braconnot, and J. Servonnat (2015), Using an ensemble data set of turbulent air-sea fluxes to evaluate the IPSL climate model in tropical regions, *J. Geophys. Res. Atmos.*, 120, 4483–4505, doi:10.1002/2014JD022985.

Received 16 DEC 2014

Accepted 1 APR 2015

Accepted article online 15 APR 2015

Published online 19 MAY 2015

## Using an ensemble data set of turbulent air-sea fluxes to evaluate the IPSL climate model in tropical regions

Alina Găinușă-Bogdan<sup>1,2</sup>, Pascale Braconnot<sup>1</sup>, and Jérôme Servonnat<sup>1</sup>

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement, Unité mixte CEA-CNRS-UVSQ, Gif-sur-Yvette Cedex, France,

<sup>2</sup>Laboratoire de Météorologie Dynamique, Institut Pierre Simon Laplace, Paris, France

**Abstract** Ocean-atmosphere interactions represent a key component of the hydrological cycle in tropical regions and their variability has profound influences on low-latitude climate. In order to evaluate how climate models represent these fluxes while taking into account the observational uncertainties, we assemble a comprehensive database of 14 climatological surface flux products, including in situ-based, satellite, hybrid, and reanalysis data sets. We find that the large observational uncertainties are reflected in the climatological magnitudes, as well as in the spatial patterns and seasonal variations and that, for the most part, they do not carry specific signatures of product type. This data ensemble allows us to draw several conclusions on the current representation of the intertropical turbulent air-sea fluxes in the atmospheric component of the Institut Pierre Simon Laplace-Coupled Model 5A, when forced by observed sea surface temperatures. Despite significantly underestimated near-surface wind speeds over the entire tropical oceans domain, the atmospheric model produces generally well represented zonal and meridional wind stress values, and only weak biases in the spatial patterns and seasonality. The simulated latent heat flux develops a bias pattern matching that of the wind speed, but with no systematic underestimation. Compared to the same reference, the sensible heat flux is overestimated over the entire region of interest, in response to a significant overestimation of the sea-air temperature contrast. The observational ensemble and analyses presented in this paper offer a good framework for large-scale model surface flux evaluation.

### 1. Introduction

Ocean-atmosphere heat and water fluxes play a major role in the global climate, as they determine the vertical transport of heat, momentum, and moisture and their availability for meridional transport and exchange with the other components of the climate system.

These fluxes are expected to be modified in response to climate change, introducing a complex feedback in the evolution of the global climate. However, the nature of this response is yet uncertain. Furthermore, due to their coupling role on smaller time scales, they also have an important effect on seasonal and decadal model prediction skill [Gulev *et al.*, 2008]. Their correct representation in global numerical models is thus of chief importance [Barrier, 2001], as is the development of quantitative and objective methods of evaluation of simulated fluxes with observations.

In this paper we are interested in evaluating the representation of the turbulent ocean-atmosphere interactions in the Institut Pierre Simon Laplace-Coupled Model 5A (IPSL-CM5A) at low latitudes. The choice of this region is motivated by the ongoing improvement on tropical convection in the IPSL model [Rio *et al.*, 2013]. A future revision to the model will concern the representation of the turbulent surface fluxes. This is why they represent the focus of the present evaluation, the radiative fluxes being beyond the scope of this study. The IPSL-CM5A model is part of the Coupled Model Intercomparison Project Phase 5 (CMIP5) modeling exercise and is presented in detail by Dufresne *et al.* [2013]. In this study we present the first step of this assessment, focused on its atmospheric component, LMDZ5A. This evaluation considers not only the flux estimates but also the available meteorological variables used in the flux calculations.

In models as well as global gridded observational products turbulent fluxes are based on bulk formulae. Surface wind stress ( $\tau$ ), latent heat flux ( $LH$ ) and sensible heat flux ( $SH$ ) are thus estimated as follows:

$$\vec{\tau} = \rho C_D |\Delta \vec{U}(z)| \Delta \vec{U}(z) \quad (1)$$

$$LH = \Lambda_v \rho C_E (q(z) - q_{\text{sat}}(\theta_0)) |\Delta \vec{U}(z)| \quad (2)$$

$$SH = \rho c_p C_H (\theta(z) - \theta_0) |\Delta \vec{U}(z)| \quad (3)$$

where  $\rho$  is the near-surface air density,  $\Delta \vec{U}$  is the relative surface wind (sometimes approximated to the near-surface wind velocity, the ocean surface current being neglected),  $q$  is the near-surface specific air humidity,  $q_{\text{sat}}$  is the saturation humidity at the sea surface temperature,  $\theta_0$  and  $\theta$  are the surface and the near-surface potential air temperatures, respectively,  $\Lambda_v$  is the latent heat of vaporization and  $c_p$  the specific heat of air. The wind speed, air temperature, and air humidity are attributed to a standard level  $z$ , usually 10 m above the surface. The bulk transfer coefficients for momentum ( $C_D$ ), humidity ( $C_E$ ), and sensible heat ( $C_H$ ) are parameterized as functions of wind speed and atmospheric stability, as well as, in some cases, other factors (e.g., sea state) [Jones and Toba, 2001]. These parameterizations vary considerably from one study to another [Dyer, 1974; Blanc, 1985; Fairall et al., 2010], potentially representing the main source of uncertainty in the flux estimates in many cases, especially for  $LH$  [Brunke et al., 2011].

Evaluating a model naturally implies having a reference against which to compare the model. Several types of flux products are now available. They have been developed from in situ measurements [da Silva et al., 1994; Berry and Kent, 2009; Hughes et al., 2012], satellite observations [Andersson et al., 2010; Tomita et al., 2010; Bentamy et al., 2003; Shie et al., 2010], or meteorological reanalyses [Kalnay et al., 1996; Onogi et al., 2007; Dee et al., 2011]. Hybrid products also consider reanalyses together with other types of observations [Yu and Weller, 2007; Kumar et al., 2012; Large and Yeager, 2009; Brodeau et al., 2010]. Uncertainties in the different types of retrieval of the meteorological state variables, in the parameterization of the transfer coefficients and differences in the assembly of the final-gridded products all add to the uncertainty of surface flux products.

The large uncertainties in the “observational” turbulent flux fields have been widely discussed in the literature and shown to be on the order of 0.01–0.02 N m<sup>-2</sup> for wind stress, 30–60 W m<sup>-2</sup> for the latent heat flux, and 10–15 W m<sup>-2</sup> for the sensible heat flux, or up to 0.04 N m<sup>-2</sup> for wind stress and up to 100 W m<sup>-2</sup> for the surface heat flux [Kumar et al., 2012; Tomita et al., 2010; World Climate Research Programme (WCRP), 2000; Bourassa et al., 2008; Smith et al., 2011; Chaudhuri et al., 2013; Josey and Berry, 2010; Gulev et al., 2010; Röske, 2006]. They need to be accounted for in model evaluation [Braconnot and Frankignoul, 1993] and are a reason why estimating the quality of simulated fluxes and of their feedbacks requires the development of specific strategies. Yet, due to the difficulty in estimating it, observational uncertainty is rarely explicitly included in analyses of model results, as done by Wittenberg et al. [2006] for the tropical Pacific. The most recent large-scale model flux evaluation study of Bates et al. [2012] employs a single-observational reference for most of its analyses. It is only for an assessment of regional averages that this is put in the context of the multiproduct analysis of Röske [2006]. In their key paper on climate model performance, Gleckler et al. [2008] take into account the observational uncertainty by calculating the large-scale metrics against two reference products.

While for particular regions, time periods, or applications some flux products are shown to be superior to others or certain products are shown to be unsuitable [World Climate Research Programme (WCRP), 2000; Smith et al., 2011; Tomita et al., 2010; Yu et al., 2011; Kumar et al., 2012; Röske, 2006], there is currently no consensus on one global data set to be used for large-scale model surface flux validation [World Climate Research Programme (WCRP), 2000; Bourassa et al., 2008; Smith et al., 2011; Gulev et al., 2010; World Climate Research Programme (WCRP), 2012]. Therefore, our approach is to consider the major available climatological gridded surface flux products that we combine into an observational ensemble to compare our model results with.

In a first step, we describe the observational database. We characterize the spread between the different flux products and related atmospheric and oceanic variables, we check the consistency between the products and investigate whether there are systematic biases pertaining to the different product types. The results of this first analysis consolidate the observational ensemble approach for our purposes. We then use this data compilation to evaluate the representation of intertropical ocean-atmosphere turbulent fluxes in the IPSL-CM5A model [Dufresne et al., 2013], starting with its atmospheric component LMDZ5A [Hourdin et al., 2013]. Rather than attempting to calculate a skill score for the model, our focus is on the identification of the robust flux-related biases in the model climatology. Wherever the model results fall outside the so-defined observational envelope, a significant model bias is identified and can be targeted for model improvement

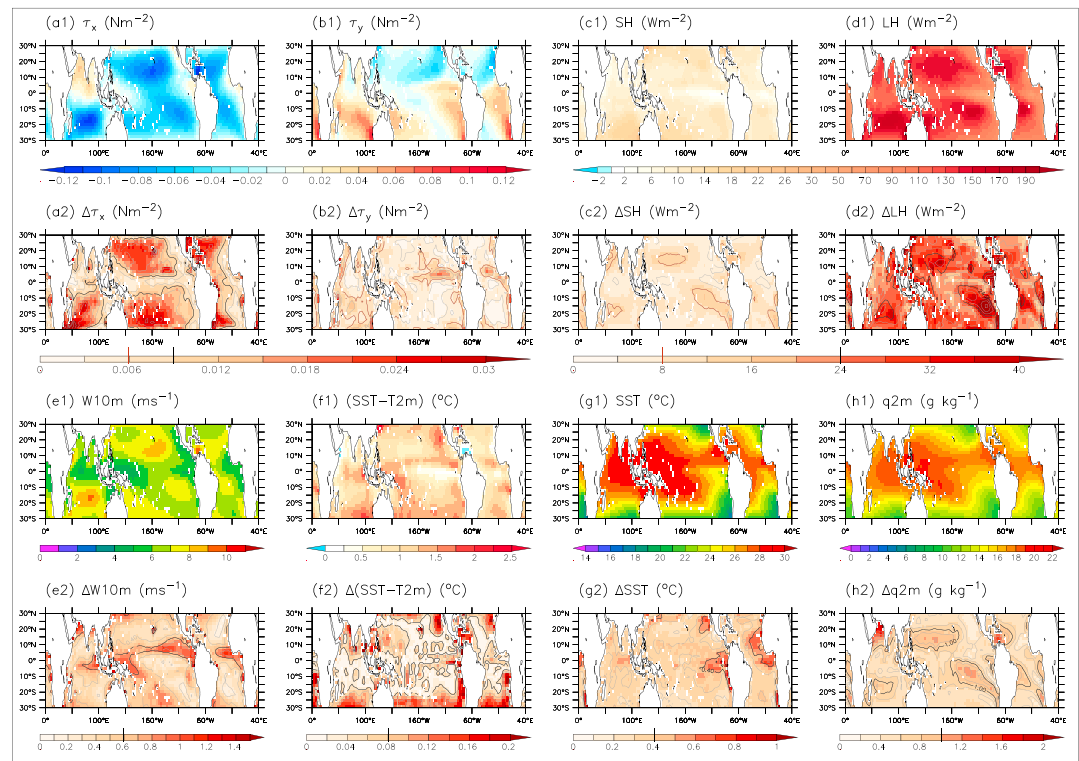
**Table 1.** Summary of Observational Data Sets Used in This Study<sup>a</sup>

Data Category	Data Set	Period Covered	Reference
In situ-based	<u>Da Silva</u>	1945–1989	<i>da Silva et al.</i> [1994]
	<u>NOC2</u> (Version 2 of the National Oceanography Center flux data set)	1973–2011	<i>Berry and Kent</i> [2009]
	<u>FSU3</u> (Florida State University fluxes)	1978–2004	<i>Hughes et al.</i> [2012]
Satellite-based	<u>HOAPS3</u> (Version 3 of the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite data)	1988–2005	<i>Andersson et al.</i> [2010]
	<u>J-OFURO2</u> (Japanese Ocean Flux Data Sets with Use of Remote Sensing Observations)	1988–2007	<i>Tomita et al.</i> [2010]
	<u>IFREMER</u>	1993–2007	<i>Bentamy et al.</i> [2003]
	<u>GSSTF2b</u> (Version 2 of the Goddard Satellite-Based Surface Turbulent Fluxes)	1988–2008	<i>Shie et al.</i> [2010]
Reanalysis	<u>NCEP/NCAR</u> (National Centers for Environmental Prediction/ National Center for Atmospheric Research reanalysis)	1948 to the present	<i>Kalnay et al.</i> [1996]
	<u>JRA25</u> (Japanese 25 year reanalysis)	1979–2005	<i>Onogi et al.</i> [2007]
	<u>ERA-Interim</u> (European Center for Medium-Range Weather Forecasts reanalysis)	1979 to the present	<i>Dee et al.</i> [2011]
Hybrid	<u>OAFlex</u> (Objectively Analyzed air-sea Fluxes for the Global Oceans)	1958–2009	<i>Yu and Weller</i> [2007]
	<u>TropFlux</u>	1989–2011	<i>Kumar et al.</i> [2012]
	<u>CORE2</u> (GFDL version 2 forcing for common ocean-ice reference experiments)	1949–2006	<i>Large and Yeager</i> [2009]
	<u>DFS4</u> (DRAKKAR Forcing Set v4.3)	1958–2006	<i>Brodeau et al.</i> [2010]

<sup>a</sup>The underlined characters represent the short names for the observational data sets. These are the names used throughout the document to refer to observational products. What follows these short names in the table, in parentheses, are, where applicable, the corresponding extended names.

and accounted for in studies using that model. We mainly focus on aspects that can be of interest for the evaluation of other models, such as climatological magnitudes and spatial patterns, shape and amplitude of the seasonal cycle, and coherence with large-scale atmosphere and ocean dynamics.

The manuscript is organized as follows: the observational and model data used in the analysis, as well as the preliminary data treatment, are presented in section 2. In section 3 we first offer a brief overview of the main statistics employed in the analyses. We then describe the observed distribution of turbulent air-sea fluxes in the intertropical region, quantify the observational uncertainty of the fluxes and of the associated surface state variables and address the question of systematic differences between different observational data types. Section 4 presents the LMDZ5A model results in comparison with the observational ensemble, all the while considering the intervariable relationships. We conclude the paper with a discussion and summary of our findings in section 5.



**Figure 1.** Average observed climatological annual mean (a1) zonal wind stress, (b1) meridional wind stress, (c1) sensible heat flux, (d1) latent heat flux, (e1) near-surface wind speed, (f1) sea-air temperature contrast, (g1) sea surface temperature, (h1) near-surface specific air humidity over the intertropical oceans. (a2–h2) The corresponding observational uncertainty, calculated as either the maximum observed departure of a data set from the ensemble mean (equation (4), colors) or the root-mean-square pairwise distance between the observational products (equation (6), contours). The contours coincide with consecutive interval limits on the corresponding color bars, except for (e2), where the contour spacing corresponds to  $0.2 \text{ m s}^{-1}$  and (f2), where it corresponds to  $0.08^\circ\text{C}$ ; in each subfigure, the level of a sample, dark contour, is indicated by a similar tick mark on the color bar. Note that the reanalyses are not considered in the observational ensemble.

## 2. Data Sets: Observations and Simulations

### 2.1. Observational-Based Products

In an intercomparison of observational flux products, *Smith et al.* [2011] have shown that different observational strategies are in order depending on the study aims. To our current knowledge, there is no one ideal climatological data set to be used for general, large-scale validation of the representation of turbulent surface fluxes in climate models. We thus assemble a comprehensive observational database, covering different types of turbulent flux products, which we use to establish a range of reasonable climatological values, spatial patterns, and seasonal variations to compare our model results with (Table 1).

We restrict the database to available gridded data sets over the intertropical oceans for the latent heat flux ( $LH$ ), the sensible heat flux ( $SH$ ), the zonal wind stress ( $\tau_x$ ), the meridional wind stress ( $\tau_y$ ), and the meteorological surface parameters used to calculate these fluxes, namely, the 10 m level wind speed ( $W10m$ ), the sea-air temperature gradient using the sea surface and 2 m level air temperatures ( $\Delta T2m$ ), the sea surface temperature ( $SST$ ), and the 2 m level specific air humidity ( $q2m$ ) (Figure 1). The availability of these variables for each data set (Table 1) is given in Table 2. Note that different products offer the air temperature and humidity at two different standard levels, 2 m and 10 m above the surface, 10 m being the standard altitude generally used in the bulk formulae flux computations. Here we have chosen only the 2 m level temperature and humidity, as they directly correspond to the output of the model to be evaluated.

Our observational database includes three in situ-based data sets (Table 1). In situ measurements (from research or voluntary observing vessels, buoys, and floats) cover longer time periods and include research quality climatological data, but the corresponding global gridded data sets often suffer from poor and inhomogeneous sampling and inherent observational biases and uncertainties, including problems of

**Table 2.** List of Data Sets and Variables Analyzed in This Study<sup>a</sup>

Data Set	Variables							
	LH	SH	$\tau_x$	$\tau_y$	W10m	T2m	SST	q2m
AMIP	✓	✓	✓	✓	✓	✓	✓	✓
Da Silva	✓	✓	✓	✓	-	-	✓	-
NOC2	✓	✓	-	-	✓	-	✓	-
FSU3	✓	✓	✓	✓	✓	-	✓	-
HOAPS3	✓	✓	-	-	✓	-	✓	✓
J-OFURO2	✓	✓	✓	✓	✓	-	-	-
IREMER	✓	✓	✓	✓	✓	-	✓	-
GSSTF2b	✓	✓	✓	✓	✓	-	-	-
NCEP/NCAR	✓	✓	✓	✓	✓	✓	✓	✓
JRA25	✓	✓	✓	✓	-	✓	✓	✓
ERA-Interim	-	-	✓	✓	✓	✓	✓	-
OAFIux <sup>b</sup>	✓	✓	-	-	-	-	✓	-
TropFlux	✓	✓	-	-	✓	✓	✓	✓
CORE2	✓	✓	✓	✓	-	-	-	-
DFS4	✓	✓	✓	✓	✓	✓	✓	✓

<sup>a</sup>Checkmark and dash indicate the availability and absence of the corresponding field, respectively.

<sup>b</sup>OAFIux now provides *W10m*, *q2m*, and *T2m* data. However, at the time the observational data was assembled, technical difficulties concerning data access prevented their inclusion in this study.

variable corrections [Josey *et al.*, 1999; Kent and Berry, 2005; Kent *et al.*, 2007; Gulev *et al.*, 2007a, 2007b; Bourassa *et al.*, 2008; Smith *et al.*, 2011; Brunke *et al.*, 2011]. The older, Da Silva data set, in particular, has been excluded from some analyses [e.g., Smith *et al.*, 2011], as it has been shown to suffer from many caveats in recent years, including having a limited time overlap with other products (Table 1), sampling problems [Kubota *et al.*, 2002, 2003; Chou *et al.*, 2003] and an underestimation of the oceanic heat loss [Gulev *et al.*, 2010]. However, given the extensive past use of this data set [Jones *et al.*, 1999; World Climate Research Programme (WCRP), 2000; Moore and Renfrew, 2002; Kubota *et al.*, 2002; Chou *et al.*, 2003; Gleckler, 2005; Wittenberg *et al.*, 2006; Hughes *et al.*, 2012], we include it in our analysis and test whether it stands out as an outlier compared to more modern products.

The in situ products are complemented, in this study, by four satellite-based observations (Table 1). Satellite measurements provide a good answer to the spatial sampling problem and in some locations to that of temporal sampling, compared to in situ data, and, within each product, they are bound to be more consistent in space and time. However, they have shorter time coverage and their characteristics change with changes in the observing system. They are also subject to a different set of uncertainties linked to the retrieval algorithms for the variables of interest, especially for air temperature and humidity [Bourras, 2006; Bourassa *et al.*, 2008; Smith *et al.*, 2011; Brunke *et al.*, 2011]. Furthermore, they may contain systematic weather or process-related (e.g., rain and atmospheric stratification) biases and uncertainties [Smith *et al.*, 2011], severely hindering their use for general studies at climatological scale.

While not a proper observational product type, another category of data extensively used for climate monitoring, climate model evaluation, and climate studies in general is reanalysis data. For each reanalysis product, these data are obtained through the assimilation of all available observations in a model that remains unchanged over time [see National Center for Atmospheric Research, 2013; Reanalyses.org, 2013]. This provides dynamically consistent, three-dimensional, global gridded data sets of a large array of variables, with typically long time coverage. On the other hand, reanalysis data suffer from model-related problems, like the poor representation of the atmospheric boundary layer and of equatorial winds or various problems with parameterizations [Bourassa *et al.*, 2008; Smith *et al.*, 2011; Moore and Renfrew, 2002]. Importantly, the spatial and temporal distribution, and source of the assimilated data are variable, according to the input data availability, which can introduce inhomogeneities in time, even unphysical behaviors [Josey and Berry, 2010;



Smith et al., 2011; Yu et al., 2011] or significantly time-dependent errors [Kubota et al., 2008; Chaudhuri et al., 2013]. Here we consider three reanalyses, listed in Table 1.

Finally, we added four hybrid flux products (Table 1). Blended or hybrid data sets use various statistical methods to combine data from the categories described above as well as information about the input data errors and proposed corrections, in order to obtain optimized large-scale gridded data sets for specific needs (e.g., global energy balance; forcing of ocean general circulation models—CORE2 and DFS4). By design, this type of product combines some of the strengths, but also some of the weaknesses, of the original fields it is based on [Fairall et al., 2010].

It should be noted that the delimitations between the four flux product categories listed above are not always unequivocal. For example, while constructed mostly from in situ/satellite data, the FSU3, J-OFURO2, and GSSTF2b products each include at least one variable that belongs to a different data type. The SST field used in the FSU3 product is the blended product from the National Meteorological Center [Hughes et al., 2012; Reynolds, 1988]. J-OFURO2 also uses a blended SST field, the Merged satellite and in situ data Global Daily Sea Surface Temperatures but also the National Centers for Environmental Prediction (NCEP)/Department of Energy Reanalysis 2 for the near-surface air temperature [Tomita et al., 2010]. The same reanalysis is used for the air temperature, sea skin temperature, and sea level pressure in GSSTF2b [Shie et al., 2010]. These characteristics could qualify them for the hybrid product category. Here, however, we have considered as proper hybrid products only those products using several types of data and either employing specifically designed merging procedures or further ad hoc corrections to the initially obtained fields.

## 2.2. Atmospheric Model Intercomparison Project Simulations

We consider historical LMDZ5A simulations for the 1979–2010 period with prescribed observational sea surface temperatures and sea ice cover, according to the CMIP5 protocol for Atmospheric Model Intercomparison Project (AMIP) experiments [Taylor et al., 2000; Hurrell et al., 2008].

A complete description of the LMDZ atmospheric model can be found in Hourdin et al. [2006, 2013]. In particular, the atmospheric surface boundary layer is parameterized using the scheme proposed by Louis [1979], based on the Businger-Dyer bulk formula [Businger, 1966, 1988]. At the ocean interface the roughness length is computed following Smith [1988] for momentum and includes the Charnock formulation and a smooth flow term. Since this version of the model uses the same neutral drag coefficient for momentum and heat fluxes, a factor of 0.8, corresponding to a mean ratio between the heat and momentum exchange coefficients [Smith, 1988], is applied to the turbulent heat exchange coefficient to account for a lower drag [Smith, 1988].

The input variables are prognostic variables at the first model level, roughly 35 m above the surface in the tropics. However, these are not directly comparable to available observational meteorological variables. For this purpose we use diagnostic model variables for 10 m level wind speed and 2 m temperature and humidity. We have verified that the standard-level diagnostic variables are representative of the prognostic variables at the first model level, so that their evaluation is still informative for the model.

The model is run with a horizontal resolution that includes 96 grid points in latitude and 96 in longitude, equivalent to 1.875° latitudinal and 3.75° longitudinal resolution. This resolution is relatively coarse compared to some of the observational products (e.g., the JRA25 reanalysis is provided with a resolution of approximately 1.125° in both latitude and longitude). This difference in resolution can lead to differences in the representation of physical processes, especially where nonlinear processes are concerned. Vertically, 39 atmospheric layers are defined on hybrid  $\sigma - P$  coordinates, including 15 levels above 20 km height.

We use an ensemble of five such simulations in order to assess model results spread due to random atmospheric noise. These simulations are initialized with the January 1 fields of five different years (1987–1991) from a previous AMIP simulation of the 1979–1997 period.

## 2.3. Preliminary Data Processing

We chose the 1979–2005 reference period for our climatologies, as it represents the common time frame of the coupled and forced atmospheric historical simulations run with the current standard version (5A) of the IPSL model. For each data set, whenever possible, we selected the available data within this period. However, due to the sole availability of the precomputed climatology, the Da Silva data corresponds to the period 1945–1989. In each case, the climatological monthly mean fields have been obtained by averaging the data

of the corresponding month for all years used. These mean values were then combined into climatological annual cycles. The annual means were calculated as weighted averages of the climatological months according to the calendar corresponding to each individual data set. This corresponded to the Gregorian calendar for all observational data sets, and the “no leap”—Gregorian calendar without any leap years—for the AMIP simulations.

Since we only consider the variables over the ocean, the land data were masked out wherever necessary. Then, to obtain numerically comparable fields suitable for the study of large-scale features, a simple linear interpolation regridding procedure was applied to map all data onto one common grid, corresponding to the low-resolution AMIP runs (3.75° longitude × 1.875° latitude).

The ensemble mean data set for the AMIP simulations has been calculated before the computation of the climatological annual mean and the spatial regridding and masking.

### 3. The Observed Fluxes and Related Variables

We first compare the different climatological products described in section 2. The purpose is to characterize this data ensemble and document the spread among the products over the intertropical region, focusing on mean magnitudes, spatial distributions, and seasonal variations. To avoid including model-specific biases in our “observational reference,” we do not include reanalyses in any ensemble statistics. However, we do include them in our analyses, to assess how they compare with the in situ, satellite, and blended climatological fields.

#### 3.1. Strategy of Intercomparison

In the following we successively consider for each variable  $x$  at grid point  $(i, j)$  its annual mean  $x_{ij}$  and climatological seasonal cycle  $x_{ijt}$ ,  $t = 1, \dots, 12$ .

In a first step, we use two different measures to characterize the spread between different products at each grid point of the domain, considering only the annual mean (Figure 1). The first definition considers the maximum absolute value of the difference between any individual data set and the observational mean:

$$\Delta_1 \text{OBS} = \max_{i=1..n} \left( \left| \text{OBS}_i - \overline{\text{OBS}} \right| \right) \tag{4}$$

where  $n$  is the number of available non-reanalysis observational products,  $\text{OBS}_i$  represents the individual climatological observational estimates,  $\Delta_1 \text{OBS}$  is the observational uncertainty, and  $\overline{\text{OBS}}$  the observational ensemble mean at every grid point

$$\overline{\text{OBS}} = \frac{1}{n} \sum_{i=1}^n \text{OBS}_i \tag{5}$$

The  $\overline{\text{OBS}}$  annual mean fields are shown in Figures 1a1 to 1h1.

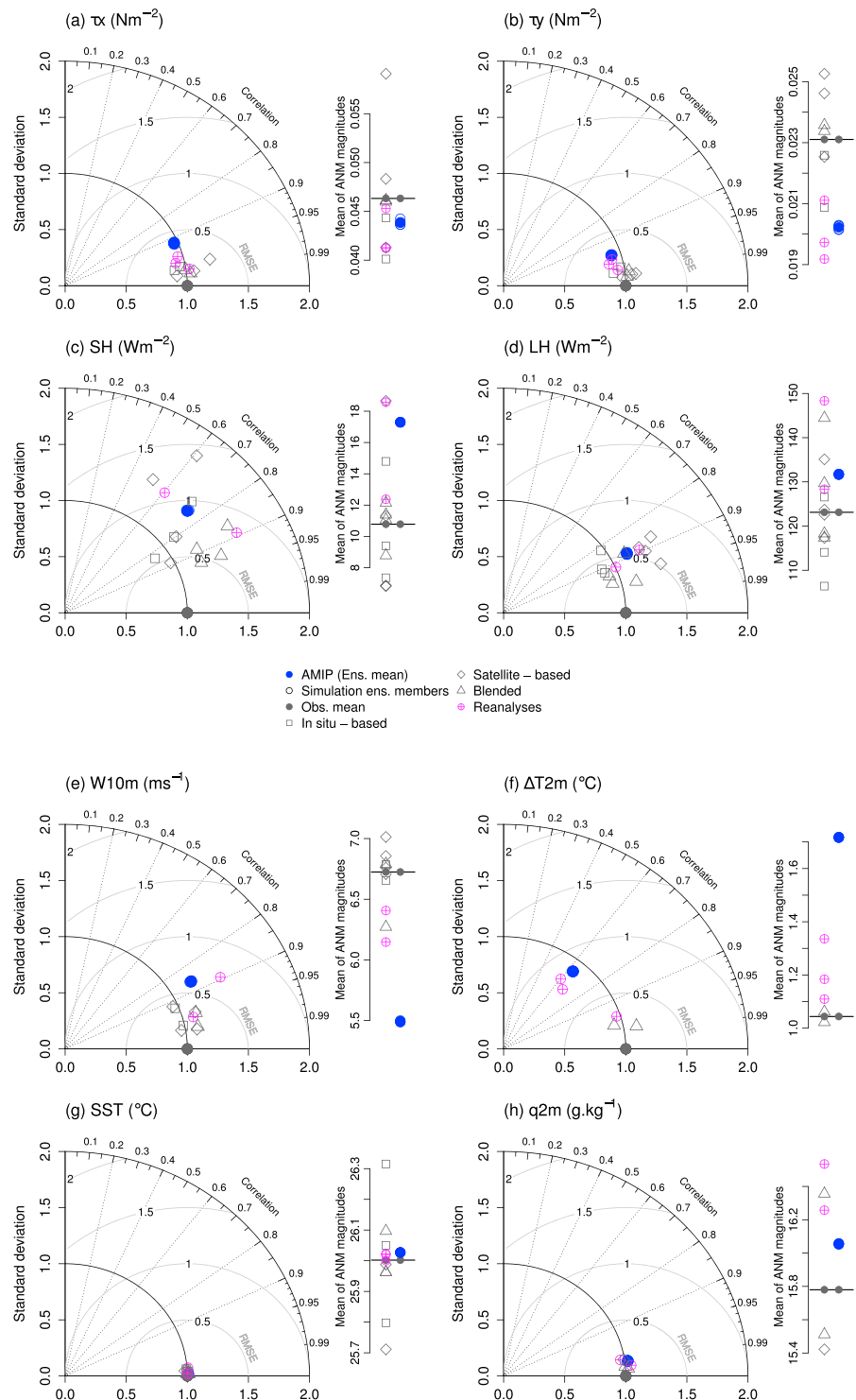
This definition of the observational uncertainty (equation (4)) provides a permissive approach to the observational data, wherein any non-reanalysis observational product is considered possibly representative of reality. The spatial distribution of this “maximum uncertainty” estimate is shown in the color maps in Figures 1a2 to 1h2.

The alternative estimate of the observational uncertainty does not make use of the observational mean. The number of individual observational products is too low to properly sample the underlying statistical distribution. Moreover, all the products cannot be considered strictly independent. This is why the second definition of the observational spread is based on pairwise root-mean-square (RMS) differences between the non-reanalysis observational products. It estimates the typical distance between two products in the ensemble, at every grid point:

$$\Delta_2 \text{OBS} = \left[ \sum_{i=1}^{n-1} \sum_{j=i+1}^n (\text{OBS}_i - \text{OBS}_j)^2 / \binom{n}{2} \right]^{0.5} \tag{6}$$

where  $\binom{n}{2} = \frac{(n-1)n}{2}$  is the number of pair combinations. The results of this second estimate of the observational spread are shown by the contours in Figures 1a2 to 1h2.





**Figure 2.** Taylor diagrams representing the annual mean spatial variability over the intertropical oceans relative to the observational ensemble mean data set for the individual observational data sets and the AMIP simulations for (a) zonal wind stress, (b) meridional wind stress, (c) sensible heat flux, (d) latent heat flux, (e) near-surface wind speed; (f) sea-air temperature contrast; (g) sea surface temperature, and (h) near-surface specific air humidity. These diagrams show, for each individual data set, three statistics of its spatial distribution relative to the reference: the correlation coefficient (equation (A1) in Appendix A; represented by the azimuthal position on the diagram), the standard deviation (equation (A2) in Appendix A; radial distance from the origin) and the root-mean-square difference (equation (A3) in Appendix A; distance from the reference). Side bars compare the absolute climatological annual mean values averaged over the region between individual observational/model data sets and the reference.

Since the degree of agreement among products in terms of mean variable magnitudes is not necessarily correlated with the agreement in terms of their variability patterns, we go one step further and disconnect the typical magnitudes from the spatial patterns by decomposing the annual mean data at every grid point into a mean value over the tropics  $\langle x_{ij} \rangle$  and an annual mean spatial pattern (the spatial anomaly from the spatial mean)  $\delta x_{ij}$  for every product:

$$x_{ij} = \langle x_{ij} \rangle + \delta x_{ij} \quad (7)$$

We synthesize, for each climatological annual mean data set  $x$ , three statistics of its spatial variability relative to that of the observational reference over the intertropical oceans with the aid of the Taylor diagrams in Figure 2 [Taylor, 2001]. Details on notation and statistics can be found in Appendix A. Figure 2 also contains information on the mean magnitudes of the variables in the different products, by comparing the spatial averages of their absolute values over the intertropical oceans,  $\langle |x_{ij}| \rangle$ , along the side bars next to the Taylor diagrams.

Finally, we also consider the seasonal variability around the annual mean, separately, by decomposing the climatological monthly data at every grid point  $x_{ijt}$  into its annual mean component,  $x_{ij}$  and a local seasonal anomaly  $\delta x_{ijt}$  that we treat separately:

$$x_{ijt} = x_{ij} + \delta x_{ijt} \quad (8)$$

We summarize the information related to the climatological seasonal variations  $\delta x_{ijt}$  in a similar fashion to the analysis of the annual mean spatial patterns in Figure 3. For the spatiotemporal variability we consider the 12 months in the calculation of the correlation and RMS. The side bars compare the average amplitudes of the climatological seasonal cycle over the intertropical region,  $\langle \text{OBS}_{SA} \rangle$ , where  $\text{OBS}_{SA}$  is defined at each grid point as the difference between the maximum and minimum climatological monthly data.

To fully assess how large the average uncertainty is, we compared the different spreads with two physically relevant magnitudes: the average climatological annual mean magnitude and the average amplitude of the climatological seasonal cycle of the mean reference  $\overline{\text{OBS}}$ . These averages are representative of a mean value over the tropics. These relative measures are presented in Table 3.

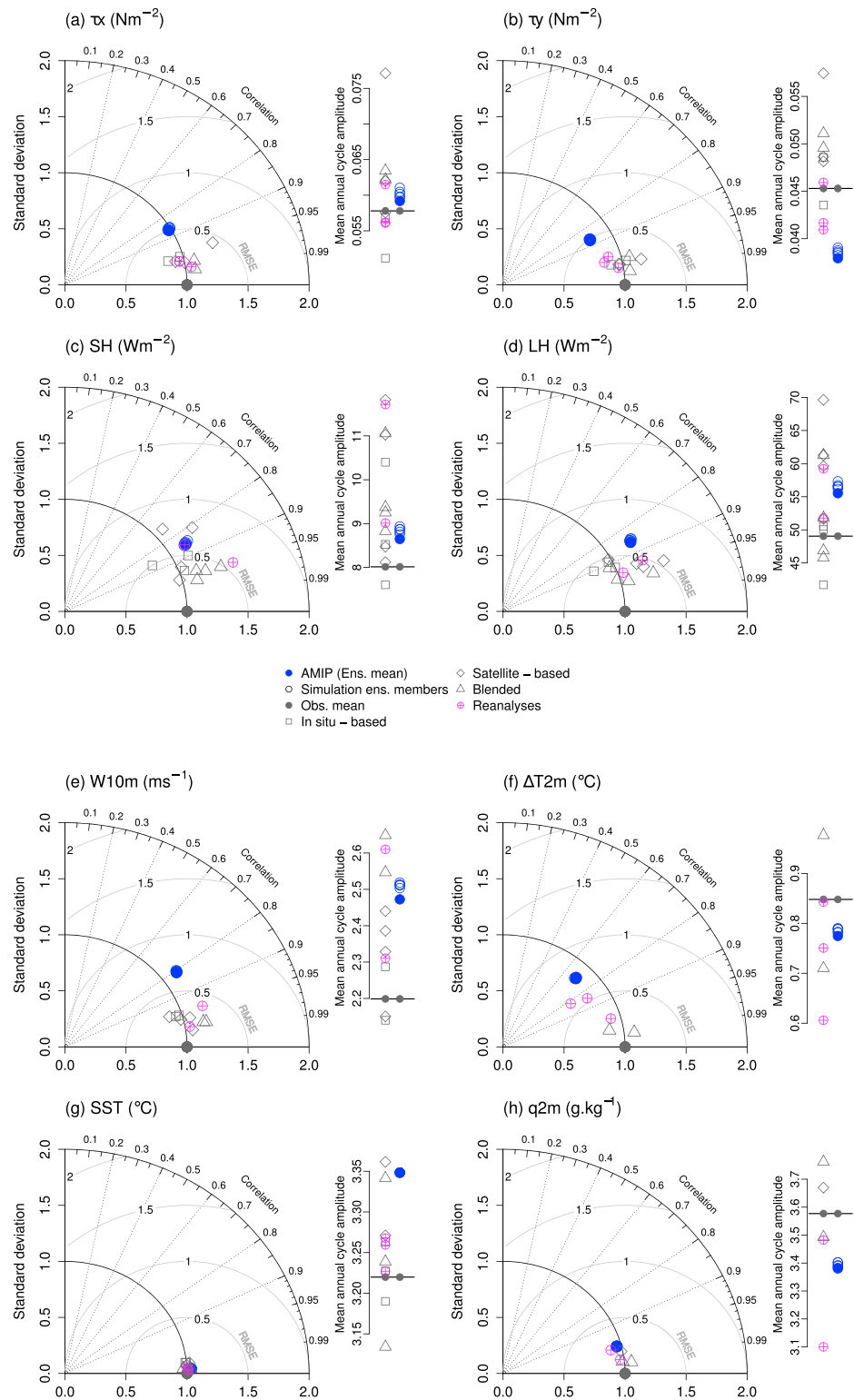
In the following subsections, we present the results of these analyses for the different variables: the wind stress (3.2), the turbulent heat fluxes (3.3), and the meteorological state variables associated to the turbulent fluxes (3.4).

### 3.2. Wind Stress

The maps of the ensemble mean zonal and meridional wind stress in Figures 1a1 and 1b1 are consistent with the well-known features of the wind stress at low latitudes. The tropics are characterized by a predominantly easterly wind stress over the oceans, with maxima located near the central regions of the basins (Figure 1a1). Weak westerly wind stress is found over limited areas such as the Arabian Sea and Bay of Bengal, where it is associated to the annual reversal of the monsoon circulation. The weaker equatorward component of the wind stress, with climatological annual mean values on the order of  $0.02 \text{ N m}^{-2}$  (versus  $0.05 \text{ N m}^{-2}$  for  $\tau_x$ ), is maximal along the continental west coasts, where the wind stress field is influenced by orography (Figure 1b1).

The colors in Figures 1a2 and 1b2 show the maximum departures of the different products from the ensemble mean (the “maximum observational uncertainty,” as defined in section 2.3, equation (4)) for the climatological annual mean zonal and meridional components of the wind stress. The annual mean climatological zonal wind stress  $\tau_x$  has maximum uncertainties of  $0.014 \pm 0.007 \text{ N m}^{-2}$  over the intertropical oceans (note that statistics are given as mean  $\pm$  standard deviation of climatological annual mean values over the intertropical oceans domain throughout the text, unless otherwise specified). Consistent with the lower magnitudes of  $\tau_y$ , the  $\tau_y$  uncertainties are smaller ( $0.006 \pm 0.003 \text{ N m}^{-2}$ ), especially over the central west tropical and subequatorial oceanic regions (Figure 1a2). Note, however, that relative to their respective average magnitudes, the  $\Delta_1 \text{OBS}$  uncertainties in the two components of the wind stress are rather similar: 25% for  $\tau_y$  and 31% for  $\tau_x$  (Table 3).

The typical distances between the different wind stress products,  $\Delta_2 \text{OBS}$  (equation (6)) are given by the contours in Figures 1a2 and 1b2. These pairwise RMS observational uncertainty estimates show roughly the same spatial patterns over the tropics as the maximum uncertainties  $\Delta_1 \text{OBS}$ , but are lower, on the order of 20–25% of the average climatological magnitudes (Table 3).



**Figure 3.** Taylor diagrams representing AMIP-modeled and observational climatological seasonal variability over the intertropical oceans relative to the observational ensemble mean data set (same as equations (A1), (A2), and (A3), Appendix A, with  $\delta x_{ijt}, \delta \overline{\text{OBS}}_{ijt}, N_i N_j N_t$  instead of  $\delta x_{ij}, \delta \overline{\text{OBS}}_{ij}, N_i N_j$ , respectively, and summation on all three dimensions  $i, j$ , and  $t$ ), for (a) zonal wind stress, (b) meridional wind stress, (c) sensible heat flux, (d) latent heat flux, (e) near-surface wind speed, (f) sea-air temperature contrast, (g) sea surface temperature, and (h) near-surface specific air humidity. Side bars compare the spatial averages of the local amplitudes of the climatological seasonal cycle between individual observational/model data sets and the reference.

**Table 3.** Observational Uncertainty Divided by the Observational Ensemble Mean Magnitudes  $|\bar{X}|$  or by the Amplitude of the Ensemble Mean Climatological Seasonal Cycle  $\bar{X}_{SA}$ <sup>a</sup>

Variable $X$	Climatological Annual Mean Values				Seasonal Cycle Amplitude	
	$\frac{\Delta_1 X}{ \bar{X} }$	$\frac{\Delta_2 X}{ \bar{X} }$	$\frac{\Delta_1 X}{\bar{X}_{SA}}$	$\frac{\Delta_2 X}{\bar{X}_{SA}}$	$\frac{\Delta_1 X_{SA}}{\bar{X}_{SA}}$	$\frac{\Delta_2 X_{SA}}{\bar{X}_{SA}}$
SST	1.5%	1.1%	12%	9%	9%	7%
$q2m$ <sup>b</sup>	4%	5%	18%	23%	14%	16%
$\Delta T2m$ <sup>b</sup>	6%	12%	7%	14%	14%	25%
$W10m$	8%	6%	25%	19%	31%	23%
$\tau_y$	25%	21%	12%	10%	33%	24%
$\tau_x$ <sup>c</sup>	31%	24%	24%	19%	41%	30%
$LH$	22%	17%	54%	41%	56%	39%
$SH$	84%	57%	111%	75%	87%	51%

<sup>a</sup>Values are given as percentages of spatially averaged terms over the intertropical oceans. The variables are presented in ascending order of uncertainty. Note that reanalysis products were excluded from the calculation of these statistics.

<sup>b</sup>Due to the reduced availability of products for these variables, the statistics presented here are likely (potentially serious) underestimates of the present uncertainty in  $q2m$  and  $\Delta T2m$  fields associated to surface flux products. The restricted number of products is also the reason for the superiority of the pairwise uncertainty estimates to those based on the maximum departure from the ensemble mean.

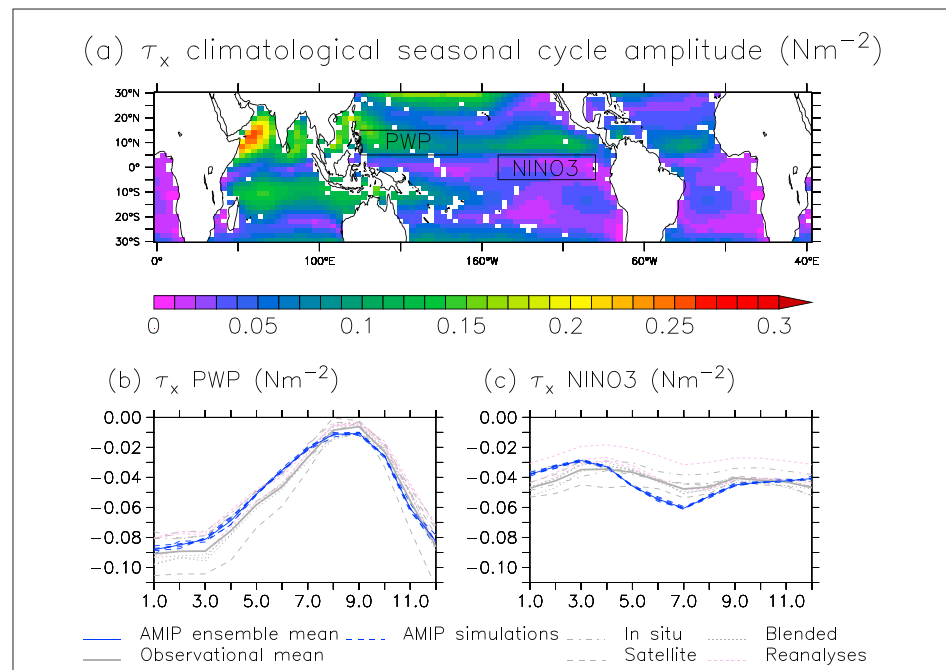
<sup>c</sup>Note that IFREMER  $\tau_x$  could be considered an outlier in this data collection. If it was excluded from the statistics, these values would become, in order, 21%, 17%, 17%, 13%, 29%, and 24%.

The wind stress uncertainty ranges issued from this intercomparison are slightly narrower than the ones reported by *Smith et al.* [2011] for zonal wind stress. This is possibly due to the use of updated data products and our exclusion of the reanalyses from the estimate of observational spread.

The Taylor diagrams in Figures 2a and 2b show that the correlation coefficient between the spatial variations of the individual climatological wind stress products and the ensemble observational mean exceeds 0.98 for both  $\tau_x$  and  $\tau_y$ . The different observational annual mean wind stress data sets thus agree reasonably well in terms of their spatial patterns. They also agree well on the  $\tau_x$ ,  $\tau_y$  gradients, as their respective standard deviations do not differ by more than 10% from  $\sigma_{s, OBS}$  for both wind stress components (except for one  $\tau_x$  product, which we discuss at the end of this subsection).

Figure 4 offers a glimpse of the climatological zonal wind stress seasonality. Intertropical seasonal variations differ spatially not only in terms of amplitude (e.g., Figure 4a) and timing but also in terms of shape, as some regions are characterized by an annual cycle (e.g., Figure 4b), while others show semiannual variations (e.g., Figure 4c). The amplitude of the climatological seasonal cycle is very large compared to the average annual mean magnitudes,  $\bar{OBS}_{SA}$  being on the order of 0.06 N m<sup>-2</sup> for  $\tau_x$  (Figures 4a and 3a) and 0.05 N m<sup>-2</sup> for  $\tau_y$  (Figure 4b). The maximum differences between the different products and the observational mean estimate represent 41% and 33% of the mean values of the climatological seasonal cycle amplitude for the two components of the wind stress, while the pairwise differences represent 30% and 24%, respectively (Table 3).

The agreement between the wind stress products for the climatological seasonal cycle (Figures 3a and 3b) is similar to that found for the annual mean distributions (Figures 2a and 2b). The similar standard deviations and the high correlation coefficients with the common reference for  $\delta\tau_{ijt}$  and  $\delta\tau_{ij}$  highlight robust observational estimates of the climatological spatial patterns and seasonality of wind stress. The only exception is the IFREMER zonal wind stress product (outlier diamond symbol on Taylor diagrams and side bars of Figures 2a and 3a). Despite very similar  $\tau_y$  estimates to the other products, the IFREMER satellite product shows considerably higher climatological  $\tau_x$  magnitudes (side bar, Figure 2a), stronger spatial variability (Figure 2a)



**Figure 4.** (a) Amplitude of the climatological seasonal cycle of the observational mean surface zonal wind stress,  $\tau_x$ ; (b) and (c) observational and simulated climatological seasonal cycles of  $\tau_x$ , averaged over two regions of the same size in the West Pacific warm pool (PWP: 125°E–185°E, 5°N–15°N) and in the East equatorial Pacific (NINO3: 210°E–270°E, 5°S–5°N), respectively.

and higher climatological seasonal cycle amplitude and variability (Figure 3a). A recent, fully revised, version of the IFREMER product [Bentamy et al., 2013] may offer an improved agreement with the other  $\tau_x$  estimates.

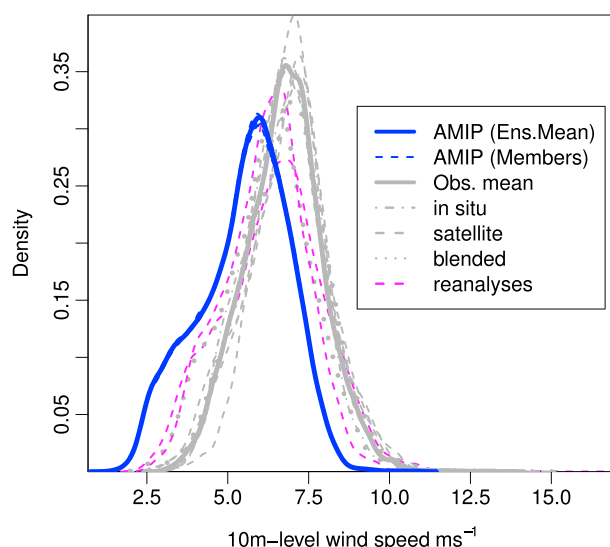
### 3.3. Heat Fluxes

With an order of magnitude of  $10 \text{ W m}^{-2}$  over the intertropical oceans (Figure 1c1), the sensible heat flux is a weak ( $\sim 8\%$ ) component of the total ocean-atmosphere turbulent heat flux in these regions. It is tightly linked with the ocean-lower atmosphere temperature contrast and the wind speed (equation (3)).  $SH$  is thus positive (heat flux from the ocean to the atmosphere) for most of the tropics and presents low or even negative values above oceanic upwelling regions where the shoaling of the thermocline brings cold seawater to the surface (e.g., Pacific equatorial upwelling—Figure 1c1).

The spread between the 13 estimates of  $SH$  considered is comparatively very high. Maximum differences of  $9 \pm 3 \text{ W m}^{-2}$  (Figure 1c2) between individual estimates and the observational mean are almost as large as the climatological mean values of the flux themselves (Figure 1c1) and exceed the amplitude of the mean climatological seasonal cycle (Table 3). IFREMER and JRA25 provide the highest, and Da Silva and J-OFURO2 the lowest estimates of the ensemble (highest and lowest two symbols on the side bar of Figure 2c). Pairwise differences among products show a similar distribution of lower/higher uncertainty regions (contours versus colors in Figure 1c2) and are roughly 30% lower than the  $\Delta_1$ OBS estimates. As shown in Figures 2c and 3c, the seasonality and spatial patterns are also very poorly correlated between the different products, with correlation coefficients ranging from 0.5 to 0.97. The regions of highest uncertainty correspond to a large degree to strong wind stress regions (Figures 1c2, 1a1, and 1b1).

Most of the turbulent heat exchanged between the ocean and the atmosphere in the tropics is in the form of latent heat. The observational annual mean climatological latent heat flux (Figure 1d1) is on the order of  $122 \pm 27 \text{ W m}^{-2}$ . The associated pattern carries the signature of the cold surface waters in upwelling regions, resulting in low latent heat fluxes, as well as that of the near-surface wind speed (equation (2) and Figure 1e1).

The 13 latent heat flux data sets agree somewhat better on spatial patterns and seasonal variability than the corresponding sensible heat flux estimates (Figures 2d and 3d versus Figures 2c and 3c). However, the maximum deviation between individual products and the observational mean is  $27 \pm 6 \text{ W m}^{-2}$  (Figure 1d2),



**Figure 5.** Distributions of the values of climatological monthly 10 m level wind speeds over the intertropical oceans in: AMIP simulations (blue: solid line - ensemble mean; dotted lines - ensemble members) and different observational data sets (other colors).

regions of highest uncertainty on the latent heat flux seasonal cycle amplitude coincide to a large extent to those of the sensible heat flux (not shown). A considerable spread is also found between the spatial distributions of the latent heat flux (Figure 2d), with individual observational  $LH$  patterns showing correlation coefficients with the observational mean over the entire study field from 0.97 down to 0.82, and much less in certain regions.

### 3.4. Uncertainties in Temperature, Wind, and Humidity

Much like the surface wind stress, the 10 m level wind speed patterns and seasonal variations agree reasonably well between the different observational products (Figures 2e and 3e). However, the spread in absolute values ( $\Delta_1\text{OBS}$  on the order of  $0.6 \pm 0.2 \text{ m s}^{-1}$ ) corresponds to approximately 25% of the mean amplitude of the climatological seasonal variations (Table 3). Figure 5 summarizes the distribution of climatological monthly wind speed values in the different data sets when considering all grid points over the intertropical oceans. It is notable that the reanalysis products and TropFlux (pink dashed and similar grey dotted curves in Figure 5) tend to provide lower estimates of the near-surface wind speed for most regions over the intertropical oceans and contain several occurrences of very low climatological monthly wind speed values ( $W_{10m} < 3 \text{ m s}^{-1}$ ). Such low climatological values are absent from the other data sets. Another notable observation is that NCEP contains considerably more spatial variability than all the other data sets (visible as the isolated reanalysis product with the highest normalized standard deviation in Figure 2e). No such signature is found, however, for the NCEP wind stress components (Figures 2a and 2b). Likewise, the outstanding characteristics of the IFREMER  $\tau_x$  climatology noted in section 3.2 are not reflected in the comparison of the corresponding wind speed field with the other  $W_{10m}$  products.

Only five of the data sets we have considered provided both the sea surface temperature and the 2 m level air temperature, allowing the computation of the sea-air temperature contrast (Table 2). Out of these, the non-reanalysis products, TropFlux and DFS4, are in good agreement on the values, mean patterns and seasonal variations of  $\Delta T_{2m}$  (Table 3 and Figures 2f and 3f). On the other hand, the reanalysis products show important departures from the two blended data fields. JRA25 and ERA-Interim present higher mean  $\Delta T_{2m}$  magnitudes (side bar, Figure 2f), and the spatial and seasonal variability in NCEP/NCAR and JRA25 presents much lower correlations with the ensemble mean than the other three products (Figures 2f and 3f). The reduced number of non-reanalysis data sets has the effect of reversing the relationship we have seen so far between the two uncertainty estimate types, with  $\Delta_1\text{OBS} < \Delta_2\text{OBS}$  for  $\Delta T_{2m}$ , as the “typical distance” between the two data sets becomes necessarily superior to their “maximum departure from the mean” (Table 3).

roughly equivalent to 54% of the average climatological seasonal cycle amplitude of sea-air  $LH$  at low latitudes (Table 3). Comparatively, the pairwise differences between the climatological annual mean  $LH$  products are equivalent to roughly 41% of the mean tropical seasonal amplitude  $\overline{\text{OBS}}_{SA}$  (Table 3). JRA25 and DFS4 give the highest estimates of the latent heat flux (side bar, Figure 2d), while the lowest estimates are given by FSU3 and, except for some oceanic upwelling regions, the Da Silva data sets.

Despite a better agreement between the  $LH$  than between the  $SH$  products, there is still a large spread in the amplitude and even shape of the mean seasonal cycle of  $LH$ , with some observational data sets differing by as much as 56% from the ensemble mean in terms of average seasonal cycle amplitude over the intertropical oceans (Table 3). Interestingly, the



Among the variables considered in this study, the sea surface temperature is the best constrained (Table 3), with the 11 observational data sets agreeing on the large-scale patterns (Figure 2g) as well as on the seasonal variability (Figure 3g and Table 3). FSU3 is markedly warmer than all the other data sets, while HOAPS3 and the Da Silva climatology give the coldest estimates of the sea surface temperature.

Although to a very large extent covarying with the sea surface temperature, the 2 m level specific air humidity shows relatively more spread in magnitudes between the three available non-reanalysis observational fields (Table 3). Out of these, HOAPS3 and DFS4 provide consistently lower humidity estimates (not shown) than TropFlux, and also than NCEP/NCAR and JRA25. HOAPS3 also has a different statistical distribution of climatological intertropical  $q2m$  values compared to the reanalysis and blended products, with less high and more intermediate values (not shown). This corresponds to a westward extension of the relatively dry regions over the eastern subequatorial ocean basins in HOAPS3 compared to the other climatologies. Correlations with the observational mean larger than 0.97 indicate that the different  $q2m$  products agree well in terms of spatial and seasonal variation patterns (Figures 2h and 3h). However, a 14% spread in the mean amplitude of the seasonal cycle exists between the three non-reanalysis  $q2m$  products considered (Table 3). The spatial distribution of this uncertainty (not shown) matches that of the latent heat flux.

Note that the data sets associated to the highest (JRA25) and lowest (FSU3) estimates of the latent heat flux do not have extreme estimates of the (available) corresponding surface state variables that would explain the  $LH$  values. Together with the lack of wind speed-wind stress correspondence for the outstanding characteristics of IFREMER  $\tau_x$  and NCEP  $W10m$  within the observational ensemble, this indicates a disjunction of the climatological magnitudes of the turbulent surface fluxes and flux-related state variables among the different products.

### 3.5. Systematic Differences Between Flux Product Types?

We further assess whether we can identify systematic differences between the different flux product types, i.e., in situ, satellite, blended, and reanalyses.

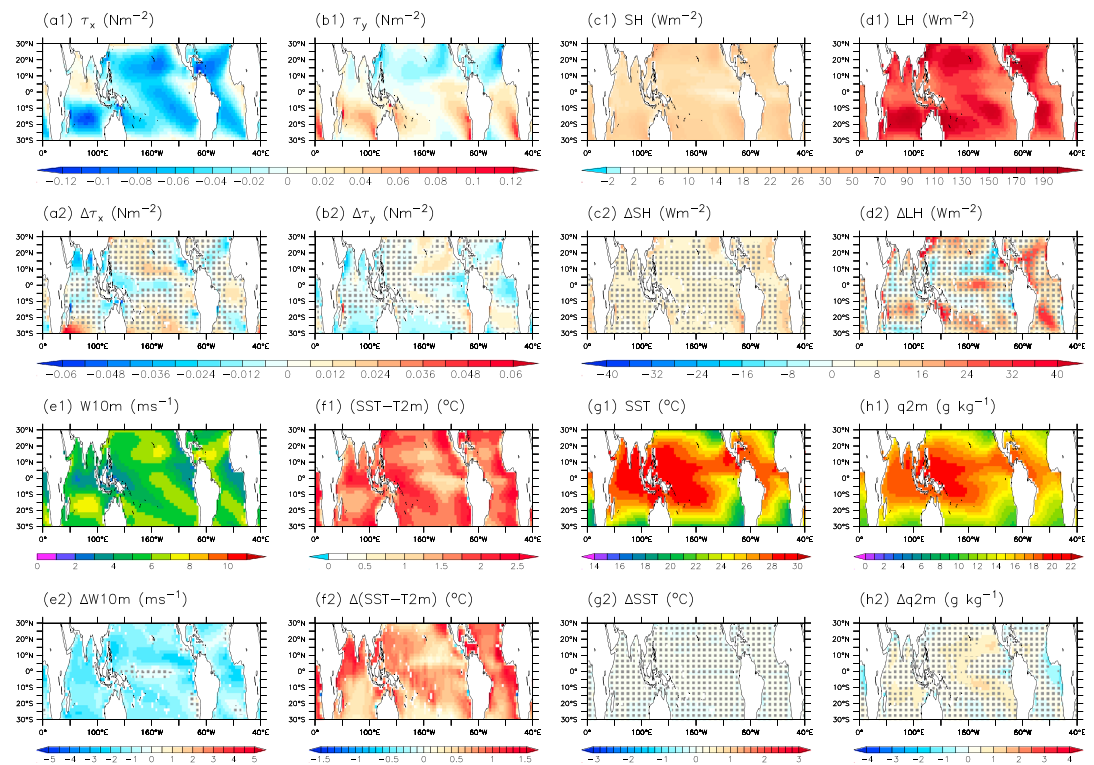
The four satellite-based latent heat flux products carry a particular spatial pattern signature (not shown). This consists of amplified fluxes (compared to the observational mean distribution) along tropical east to equatorial west diagonals over the Atlantic and Pacific basins, around 20–30°S in the Indian Ocean and over the Arabian Sea. These regions roughly coincide with the regions of strong zonal wind stress visible in Figure 1a1. This pattern is responsible for the high spatial variability of the satellite-based  $LH$  products visible in Figure 2d. Only two of the satellite products (IFREMER and HOAPS3) show the same signature in their  $SH$  fields.

The reanalyses also carry a common specificity in their  $LH$  and  $SH$  spatial patterns (not shown), consisting of exaggerated heat flux values off the subequatorial and tropical west coasts of Africa, North America, and South America. This feature is more pronounced for JRA25 than for NCEP/NCAR. Ignoring a mean shift toward low flux magnitudes for  $LH$ , this pattern is also somewhat present in the TropFlux heat fluxes and might be linked to the atmospheric stability formulation in the flux bulk formula. In turn, this could respond to a weak representation of stratocumulus clouds, a classical problem of many current atmospheric models.

With the few exceptions noted above, which are not strictly category specific, we did not detect any category-based clustering of the different types of observational products, either in terms of magnitudes or large-scale statistics describing the spatial patterns and seasonal variations. Furthermore, even though they tend to provide somewhat lower  $W10m$  and higher  $\Delta T2m$  estimates than most observational products and while often commented out of the surface flux observational product suite, atmospheric reanalyses seem to provide intertropical turbulent air-sea fluxes comparable to the other products we have analyzed—except for the aforementioned coastal bias.

## 4. Air-Sea Flux Representation in the IPSL CMIP5 AMIP Simulations

In this section we use the previously described observational ensemble to analyze how the sea-air turbulent fluxes and associated state variables (equations (1)–(3)) are represented in LMDZ5A, in light of the observational spread. Our objective is to identify and draw attention to the robust, important large-scale flux-related errors in the model climatology. Thus, rather than looking for a skill score, we assess if the model results are within or outside the range covered by the different observational products.



**Figure 6.** AMIP simulated climatological annual mean (a1) zonal wind stress, (b1) meridional wind stress, (c1) sensible heat flux, (d1) latent heat flux, (e1) near-surface wind speed, (f1) sea-air temperature contrast, (g1) sea surface temperature, (h1) near-surface specific air humidity over the intertropical oceans, and (a2–h2) the associated model-mean observations differences; the dotted regions correspond to areas where the model bias is not considered significant relative to the maximum observational spread—see text for details.

#### 4.1. Model Evaluation Strategy

We build on the analyses used to describe the observational ensemble (section 3.1) to show how the model compares to it. Following this, we first look at the climatological annual mean fields of the model ensemble mean  $\overline{\text{AMIP}}$ , in Figures 6a1–6h1. The mean model biases with respect to the different observational data sets  $\Delta_{\text{AMIP}}$  are shown in colors in Figures 6a2–6h2. For a “maximum model bias” estimate  $\Delta_1_{\text{AMIP}}$  comparable to  $\Delta_1_{\text{OBS}}$  (equation (4)), we consider the maximum absolute value of the difference between each individual AMIP simulation and the mean observational reference. The regions not dotted in Figures 6a2–6h2 indicate “significant” model biases, defined as those grid points where  $\Delta_1_{\text{AMIP}} > \Delta_1_{\text{OBS}}$ . Details on these statistics can be found in Appendix B. Alternative definitions of significant model biases are possible. The  $\Delta_1$  criterion is the most suited to our present goal and is thus used in Figures 6a2–6h2 and Table 4. However, for comparison, we have also devised an alternative criterion using  $\Delta_2_{\text{OBS}}$  (equation (6)) and RMS differences between all AMIP–OBS combinations. We briefly present this parallel assessment in Appendix C. The results of this alternative analysis are very similar to the results presented below, using the  $\Delta_1$  criterion, except for the sensible heat flux, where the  $\Delta_2$  criterion results in a much larger area where the model bias is considered significant.

As for the description of the observational ensemble, we use the Taylor diagrams and their side bars (Figure 2) to show how the model compares to the observational ensemble mean for the mean magnitudes over the tropics and the spatial variability of the annual mean fields (equation (7)). The same is done for the characterization of the seasonal variability (equation (8); results in Figure 3).

Taking advantage of the availability of the ensemble of AMIP historical simulations, we have also analyzed the differences introduced by the internal variability, regarded as noise, in the atmospheric model. We have found them to be invariably negligible compared to the differences between the model results and the observational mean, for all of the variables considered. Results for these individual model realizations are plotted in Figures 2–5.

**Table 4.** Summary of Mean Statistics Over the Intertropical Oceans Comparing the AMIP Simulation Results to the Observational Mean<sup>a</sup>

Variable $X$	Climatological annual mean values		Seasonal cycle amplitude
	$\frac{ \Delta X }{ \bar{X} }$	$\frac{ \Delta X }{\bar{X}_{SA}}$	$\frac{ \Delta X_{SA} }{\bar{X}_{SA}}$
SST	0.2%	1.6%	4%
q2m	2%	10%	10%
$\Delta T_{2m}$	66%	79%	36%
W10m	18%	54%	31%
$\tau_y$	28%	14%	32%
$\tau_x$	21%	16%	30%
LH	9%	22%	29%
SH	62%	82%	33%

<sup>a</sup> $|\Delta X|$  is the spatial average of the model bias magnitudes  $|\Delta AMIP|$  (equation (B2) in Appendix B) for the annual mean values of variable  $X$  over the intertropical oceans;  $|\bar{X}|$  and  $\bar{X}_{SA}$  are the spatial averages of the annual observational ensemble mean magnitudes  $|\overline{OBS}|$  and of the observational climatological seasonal cycle amplitude  $OBS_{SA}$ , respectively (see text in section 3.1 for details);  $|\Delta X_{SA}|$  is the spatially averaged magnitude of the climatological seasonal cycle amplitude bias  $\Delta AMIP_{SA}$  (equation (B4) in Appendix B). Note that reanalysis products were excluded from these statistics.

6a2). The decomposition in Figures 2a and 2b shows that wind stress component average magnitudes over the intertropical oceans are within the estimates from the different wind stress products, whereas the bias patterns are significant. The model shows a larger RMS difference from the mean observational reference than any of the observational products. This reflects small pattern shifts (low correlation) rather than a difference of spatial variability amplitude (Figures 2a and 2b)

These patterns are consistent with the 10 m level wind speed bias patterns (Figure 6e2), as can be expected from equation (1). However, the simulated wind speed values are systematically lower than observed over most of the tropics (Figure 6e2). The mean wind speed bias of the model over the intertropical oceans is  $1.2 \pm 0.5 \text{ m s}^{-1}$ , equivalent to 18% of the observed annual mean W10m (Table 4). Furthermore, the probability distribution functions of climatological monthly W10m in Figure 5 shows that the model simulates consistently lower wind speeds than observed. Some climatological values are lower than the minimum observational climatological values (down to  $2 \text{ m s}^{-1}$ ), and the model presents a much higher occurrence of wind speeds between 2 and  $4 \text{ m s}^{-1}$  than in most observations. Interestingly, the only observational data sets that contain such low climatological wind speed values are the reanalyses and the hybrid product TropFlux. These are the same products that have a similar statistical distribution of climatological W10m with the model (Figure 5) and a remarkably similar spatial pattern when compared to the observational mean reference (not shown, but similar to Figure 6e2). This suggests potentially similar behaviors of the LMDZ5A model and of the atmospheric models used in the generation of these products.

At the seasonal time scale, the simulated climatological seasonal cycle amplitude over the intertropical oceans is close to the mean observational value for  $\tau_x$  (side bar, Figure 3a;  $0.061 \pm 0.036 \text{ N m}^{-2}$  simulated versus  $0.059 \pm 0.040 \text{ N m}^{-2}$  observed), whereas it is underestimated for  $\tau_y$  (Figure 3b;  $0.042 \pm 0.034 \text{ N m}^{-2}$  simulated versus  $0.048 \pm 0.037 \text{ N m}^{-2}$  observed  $\tau_y$  amplitude). Despite the good large-scale agreement (Figure 3a), significant regional biases in terms of seasonal cycle amplitude are found for  $\tau_x$ . For example, it exhibits an amplified seasonal cycle in the NINO3 region ( $210^\circ\text{E}$ – $270^\circ\text{E}$ ,  $5^\circ\text{S}$ – $5^\circ\text{N}$ , Figure 4c), which, in a coupled model, could have repercussions on the representation of the Pacific equatorial upwelling and the El Niño–Southern Oscillation (ENSO) variability. In the West Pacific warm pool (e.g.,  $125^\circ\text{E}$ – $185^\circ\text{E}$ ,  $5^\circ\text{N}$ – $15^\circ\text{N}$ ), the seasonality of

It should be noted that the Program for Climate Model Diagnosis and Intercomparison SST data set [Taylor *et al.*, 2000] used to force the AMIP simulations is well within the observational envelope, close to the observational mean in terms of magnitude and spatial variability (blue symbol in Figure 2g). We thus argue that the choice of this SST data set for the AMIP forcing should not be responsible for biases on other variables (following our evaluation approach), at least on the magnitude and spatial variability of the climatological annual mean.

#### 4.2. Wind Stress

Using prescribed observational SSTs, the atmospheric model produces a generally good representation of the climatological wind stress (Figures 2a and 2b). However, some significant biases do emerge in the annual mean pattern over certain regions and can take both positive and negative values. The wind stress is underestimated, for example, over the Arabian Sea and the Bay of Bengal, and overestimated along tropical east to equatorial west diagonals in the South Pacific and Atlantic Oceans (Figures 6a2 and 6b2). Additionally, compared to the mean observations, the simulated  $\tau_y$  exhibits a somewhat weaker and wider convergence region in the Pacific and Atlantic Oceans around  $5$ – $7^\circ\text{N}$  (Figures 6b1 and 6b2), while the simulated  $\tau_x$  results show a narrower band of trade winds located too far north in the equatorial Pacific (Figures 6a1 and

$\tau_x$  is slightly underestimated, which is mostly due to relatively low  $\tau_x$  values during boreal summer (Figure 4b). These biases represent about 30% of the seasonal amplitude at the tropical scale (Table 4). Differences in local timing and shape of the seasonal cycle translate into degraded correlations and RMS differences in Figure 3a and 3b compared to similar estimates for the annual mean patterns in Figure 2. These seasonal differences in wind stress have their counterpart in the seasonal variability of the wind speed (Figure 3e).

### 4.3. Sensible Heat Flux

The AMIP *SH* values fall on the high end of, but within the observational range (Figure 2c). Notable exceptions where the sensible heat flux is overestimated even compared to the highest observational data are the eastern boundaries of the Atlantic and Pacific basins between approximately 10° and 30°S and 10° and 30°N (Figures 6c2 and 6c1). These bias regions are very similar to those found in reanalyses, indicating a potentially similar source of error in different atmospheric models.

Despite the model underestimated near-surface wind speed, the simulated annual mean *SH* exceeds by  $7 \pm 3 \text{ W m}^{-2}$ , or roughly 62% (Table 4), the annual observational average. This is due to a consistent, significant overestimation (on the order of 66%, Table 4) of the sea-air temperature contrast over the entire intertropical oceans domain (Figure 6f2). Both the climatological value and bias *SH* patterns match those of  $\Delta T2m$  (Figures 6c1, 6f1, 6c2, and 6f2). A modulation of the sensible heat flux by the near-surface wind speed is, however, present, for example, in the east equatorial Atlantic and Pacific, and the Indo-Pacific warm pool region, where the accentuated negative wind speed bias partially compensates the overestimated  $\Delta T2m$  (Figures 6c2, 6e2, and 6f2).

### 4.4. Latent Heat Flux

Similarly, despite a generally good representation of *q2m* (as a result of its tight coupling with the imposed SST—Figures 6h2, 6h1, and 6g1) and the underestimated *W10m*, the climatological latent heat flux values are in many regions in the upper quartile of the observational products. The differences between the simulated and the observational mean latent heat fluxes are physically important ( $11 \pm 8 \text{ W m}^{-2}$ ), but they are not significant for most regions considering the large spread of observational values (Figure 6d2).

Large-scale statistics of *LH* over the intertropical oceans (Figures 2d and 3d) only distinguish the model from the observations for the seasonality patterns (Figure 3d). In particular, the slightly lower correlation with the mean observational reference ( $r_{t,AMIP} = 0.86$ ) than the individual observational climatologies ( $r_{t,OBS}$  between 0.88 and 0.97) indicates slightly less agreement with the observations on the timing and/or collocation of the seasonal variations.

The pattern of the difference between AMIP and the  $\overline{OBS} LH$  data set is dominated by the wind speed bias pattern. Overestimated latent heat fluxes primarily correspond to regions of relatively overestimated wind speed (with respect to the mean, negative, *W10m* bias) and vice-versa (Figures 6d2 and 6e2). However, some effects of the weak humidity bias are visible in the latent heat flux representation (Figures 6d2 and 6h2). One such effect consists of horseshoe patterns of slightly overestimated humidity in the Indian and intertropical Pacific Oceans leading to relatively lower latent heat fluxes. Likewise, drier than observed conditions in the eastern equatorial Atlantic and, to a smaller extent, Pacific regions drive relatively high latent heat fluxes in these areas, despite a severe underestimation of the collocated wind speeds (Figures 6d2, 6e2, and 6h2).

## 5. Summary and Discussion

The present study serves the evaluation of turbulent fluxes in the atmospheric component of the IPSL-CM5A model. In this context, we have assembled a database of 14 climatological surface flux products (Table 1) to be used for large-scale model evaluation of turbulent ocean-atmosphere interactions in the intertropical zone (30°S–30°N). This database spans four different categories of flux products—in situ, satellite based, atmospheric reanalysis, and blended—and includes climatological monthly fields of the turbulent momentum and heat fluxes, as well as of the associated surface state variables (Table 2).

We treated this database as an observational ensemble and estimated the observational spread in two different ways, considering both the state variables and turbulent fluxes over a climatological annual cycle computed from nearly three decades with maximum data coverage (1979–2005).

Overall, the spatial distributions of the two observational spread estimates are very similar (colors versus contours in Figures 1a2 to 1h2), while their magnitudes can differ substantially. The typical distance between

two products ( $\Delta_2$ OBS) is generally smaller than the maximum departure from the observational mean ( $\Delta_1$ OBS), except for situations where they are calculated from very few observational products (e.g., two products for  $\Delta T2m$  and three products for  $q2m$  in this study). In the latter case, the different products tend to be all similarly far from the ensemble mean and the typical pairwise distance between them becomes superior to the maximum distance to the ensemble mean. The two measures of observational spread offer complementary information and the choice of any one observational uncertainty estimate depends on the application.

As expected, SST was found to be the most reliable variable, with the spread of the mean values approximately equivalent to 1.5% of the average absolute values and 12% of the average climatological seasonal cycle amplitude. Only moderately higher relative uncertainties have been found among these products for  $q2m$  and  $\Delta T2m$ . However, since we excluded reanalyses, these uncertainties have been estimated based on three data sets for  $q2m$  and only two data sets for  $\Delta T2m$ . They could thus be underestimates of the current uncertainty in these fields. Wind speed and wind stress uncertainties have been found to be close to the target accuracies suggested by Gulev *et al.* [2010] for many climatological applications: below  $1 \text{ m s}^{-1}$  for wind speed and below  $0.01 \text{ N m}^{-2}$  for the meridional and for the zonal wind stress, if the IFREMER  $\tau_x$  product is excluded.

Large uncertainties that hinder the full assessment of climate model results have been found for the turbulent surface heat fluxes both for the annual mean and the seasonal cycle, a problem already highlighted by other studies [Wittenberg *et al.*, 2006; Reichler and Kim, 2008]. While a target accuracy in the individual heat flux components is on the order of  $2\text{--}3 \text{ W m}^{-2}$  [Gulev *et al.*, 2010], and biases larger than  $10 \text{ W m}^{-2}$  are considered serious problems in climate studies [WCRP, 1989; Bourassa *et al.*, 2008] average uncertainties on the order of  $10 \text{ W m}^{-2}$  and  $30 \text{ W m}^{-2}$  have been found for the sensible and latent heat fluxes, respectively. The uncertainties are of the same order of magnitude as the climatological  $SH$  values and represent 20% and 50% of the average  $LH$  annual mean value and climatological seasonal cycle amplitude, respectively. These results show that important efforts are still needed to progress toward a consensus on low-latitude heat flux estimates.

The same ranking among the variables in terms of spread between products has been found for the spatial variations around the mean. The exception was a surprisingly good agreement between the different products in terms of the spatial variations of the wind stress components, given a lower agreement in wind speed patterns (Figures 2a, 2b, and 2e).

A larger spread has been found among products when considering the amplitudes of the climatological seasonal cycles compared to that between the annual mean values, with a common ranking of the variables according to their respective observational uncertainty. Meanwhile, the shapes of the seasonal cycles were as well or even better correlated between the different observational products than the respective annual mean spatial patterns (Figure 3 versus Figure 2). In addition, the seasonal cycle amplitude uncertainties of the sensible heat flux, latent heat flux, and air humidity share a common spatial distribution.

When comparing in situ-based, satellite-based, hybrid, and reanalysis products, the only category-specific characteristic detected was a common tendency of satellite products to show particularly enhanced latent heat fluxes (and sensible heat fluxes, for two of the products) in regions roughly coinciding with the strong  $\tau_x$  regions in Figure 1a1, compared to the  $LH$  patterns of the other observational data sets.

At the large scale, only two of the analyzed observational climatologies (Table 2) may be qualified as outliers: the IFREMER zonal wind stress, for its high spatiotemporal variability and absolute annual mean values, and the NCEP/NCAR 10 m level wind speed, for its large spatial gradients. None of these outstanding characteristics is reflected in the other associated variables. Similar findings for other products indicate that no flux product can be excluded on the basis of a systematic bias in its ancillary data. Note that, despite having been proven outdated in what concerns data coverage and energy balance, for the large-scale criteria chosen in this analysis the Da Silva climatology fits well within the described observational ensemble.

In showing that there are no clear separations between the different flux product types in terms of their features and that there are no systematic outliers, this comparison supports the use of this wide array of observational products as an ensemble for large-scale model validation. Using this approach, we were able to analyze the LMDZ5A atmospheric simulation results in light of the observational spread and to identify its significant biases.

We show that LMDZ5A develops two important biases related to turbulent ocean-atmosphere exchanges across the intertropical region. First, 10 m level wind speeds are consistently underestimated, on average by



approximately  $1.3 \text{ m s}^{-1}$  (Figure 6e2). This is equivalent to 20% of the annual mean observed signal, and more than 50% of the mean amplitude of the  $W10m$  seasonal cycle at low latitudes.

A counterintuitive result is that, despite the weak surface winds, LMDZ5A simulations with prescribed SST offer a generally good representation of climatological surface wind stresses in terms of climatological magnitudes, and spatial and seasonal variability amplitudes. However, the analysis of spatial and temporal correlations with the observational mean revealed shifts in the annual mean patterns of  $\tau_x$  and  $\tau_y$  (Figures 2a and 2b) and biases in the shape of the seasonal cycles (Figures 3a and 3b). As expected, the  $\tau_x$ ,  $\tau_y$  spatial pattern biases match those of  $W10m$ .

We find the  $LH$  biases and variability to be to a large extent driven by  $W10m$ . However, despite the AMIP  $W10m$  spatial pattern falling outside the observational range (Figure 2e), the model  $LH$  pattern statistics are similar to the observational ones (Figure 2d). This contrast is primarily due to the better observational constraints on the wind speed than on the latent heat flux patterns and further emphasizes the large uncertainties in the observational  $LH$  products. On average over the intertropical oceans the LMDZ5A AMIP latent heat fluxes are  $11 \text{ W m}^{-2}$  higher than the mean observational field, even though the model-simulated near-surface air humidity is very similar to the observed data, and the simulated wind speed is significantly lower than observed. This model-observational mean  $LH$  difference is large considering typical accuracy targets [Gulev et al., 2010; WCRP, 1989; Bourassa et al., 2008] but it is within observational bounds as a result of even larger observational uncertainties. Further investigation is needed to explain the absent effect of the  $W10m$  underestimation on the LMDZ5A wind stress and latent heat fluxes, effect that could be expected from the respective bulk formulae (equations (1) and (2)).

Finally, LMDZ produces relatively strong ocean-atmosphere sensible heat fluxes that reflect the significant, systematic exaggeration of the sea-air temperature contrast in the intertropical region (Figure 6f2), the second main bias of the LMDZ model. The mean  $\Delta T2m$  bias is found to be on the order of  $0.7^\circ\text{C}$ , approximately 70% of the observed climatological values (Table 4), much larger than the estimated observational spread of roughly 6% (Table 3). While additional observational data might modify these statistics, the large model departure from the considered observations inspires good confidence that  $\Delta T2m$  overestimation would remain a significant bias of the LMDZ5A model at low latitudes.

This evaluation of the LMDZ5A model suggests several targets for model improvement. First to be addressed should be the two identified robust model biases in the 10 m level wind speed and in the 2 m level air temperature, followed by a check of consistency between the surface fluxes and the meteorological state variables. Ongoing work concerns the investigation of the coupling between convection, the boundary layer and the formulation of the momentum and heat transfer coefficients.

We highlight that, like LMDZ5A, reanalyses tend to provide higher  $\Delta T2m$  values than other observational data sets—even though not as high as the model (Figure 2f). Furthermore, while not as low as the model  $W10m$  values, the reanalyses and the hybrid TropFlux product also provide lower climatological wind speeds than the other observational products (Figure 2e) and show similar  $W10m$  patterns to the LMDZ5A model when compared to the mean observational reference. These products also exhibit exaggerated sensible heat fluxes off the subequatorial and tropical west coasts of Africa, North America, and South America, a bias equally found in the LMDZ5A simulations (Figure 6c2). All these results indicate a potentially common source of error in the atmospheric models used in the generation of Tropflux, the reanalyses and the LMDZ5A model analyzed in this study.

This paper shows that the observational ensemble is well suited to highlight systematic model biases when the flux-related state variables and their relationships to the fluxes are also considered. It illustrates the importance of considering observational uncertainty in model evaluation, and calls for efforts in reducing these uncertainties, especially for the turbulent surface heat fluxes. Our analyses and the use of the observational ensemble offer a framework for systematic evaluation of turbulent surface fluxes and associated atmospheric and oceanic variables in model simulations.

## Appendix A: Taylor Diagrams

The Taylor diagrams in Figure 2 summarize, for each climatological annual mean data set  $x$ , three statistics of its spatial variability relative to that of the observational reference (the observational ensemble mean



data set) over the intertropical oceans. The correlation coefficient  $r_{s,x}$ , is a measure of spatial pattern shape resemblance. With the same notation used in section 3.1, it is expressed as

$$r_{s,x} = \frac{\sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \delta x_{ij} \cdot \overline{\delta OBS}_{ij}}{\sqrt{\sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \delta x_{ij}^2} \cdot \sqrt{\sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \overline{\delta OBS}_{ij}^2}} \quad (\text{A1})$$

where  $N_i, N_j$  represent the numbers of grid cells along the longitude and latitude axes, respectively,  $\delta x_{ij}$  is the spatial anomaly of the annual mean data set  $x$  at position  $(i, j)$  with respect to its spatial average over the domain, and  $\overline{\delta OBS}_{ij}$  is the spatial anomaly of the mean observational reference annual mean field.

The normalized standard deviation  $\sigma_{s,x}/\sigma_{s,\overline{OBS}}$ , is a measure of spatial pattern amplitude relative to that of the reference

$$\sigma_{s,x}/\sigma_{s,\overline{OBS}} = \frac{\sqrt{\frac{1}{N_i N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \delta x_{ij}^2}}{\sqrt{\frac{1}{N_i N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \overline{\delta OBS}_{ij}^2}} \quad (\text{A2})$$

Finally, the normalized centered root-mean-square difference  $E'_{s,x}/\sigma_{s,\overline{OBS}}$ , a total measure of “pattern difference,” is calculated as

$$E'_{s,x}/\sigma_{s,\overline{OBS}} = \frac{\sqrt{\frac{1}{N_i N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} (\delta x_{ij} - \overline{\delta OBS}_{ij})^2}}{\sqrt{\frac{1}{N_i N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \overline{\delta OBS}_{ij}^2}} \quad (\text{A3})$$

The equations used to produce the Taylor diagrams contain weighting by grid cell area. For simplicity, this is not indicated in equations (A1)–(A3).

Similarly, the Taylor diagrams in Figure 3 represent the  $r_{t,x}$ ,  $\sigma_{t,x}/\sigma_{t,\overline{OBS}}$ , and  $E'_{t,x}/\sigma_{t,\overline{OBS}}$  statistics, defined in the same way as the corresponding spatial variations statistics in equations (A1)–(A3), with  $\delta x_{ij}$ ,  $\overline{\delta OBS}_{ij}$ , and  $N_i N_j$  being replaced by  $\delta x_{ijt}$  (the climatological seasonal anomaly around the local climatological annual mean value of  $x$  at month  $t$ ),  $\overline{\delta OBS}_{ijt}$  (the local climatological seasonal anomaly of the ensemble observational mean data set), and  $N_i N_j N_t$  (where  $N_t = 12$  is the number of climatological months), respectively, and the summations being performed on all three dimensions  $i, j$ , and  $t$ .

## Appendix B: Model Evaluation Statistics

The AMIP ensemble mean, used as main model reference, is defined as

$$\overline{AMIP} = \frac{1}{n_{AMIP}} \sum_{i=1}^{n_{AMIP}} AMIP_i \quad (\text{B1})$$

where  $n_{AMIP}$  is the number of simulations ( $n_{AMIP} = 5$ ),  $AMIP_i$  represents individual AMIP runs and  $\overline{AMIP}$  is the simulation ensemble mean, at every grid point. These reference model annual mean fields are shown in Figures 6a1–6h1.

For an estimate of the model bias with information on where the model results are higher/lower than the observational estimates, we use:

$$\begin{aligned} \Delta_{AMIP} &= \frac{1}{n_{OBS}} \sum_{j=1}^{n_{OBS}} \overline{(AMIP - OBS_j)} \\ &= \overline{AMIP} - \overline{OBS} \end{aligned} \tag{B2}$$

where  $n_{OBS}$  is the number of non-reanalysis observational products. These biases for the eight variables analyzed are shown in the color maps in Figures 6a2–6h2.

We make use of the estimated observational uncertainty to map where these biases are significant relative to the observational spread. Using the maximum observational uncertainty estimate  $\Delta_1 OBS$  (equation (4)), we define a similar measure for the model:

$$\Delta_1 AMIP = \max_{i=1..n_{AMIP}} \left( |AMIP_i - \overline{OBS}| \right) \tag{B3}$$

and define the model bias to be "significant" where  $\Delta_1 AMIP > \Delta_1 OBS$ . This criterion ensures that the model results are outside the observational spread where a "significant bias" has been identified. It is thus suited to our present goal and is used in Figures 6a2–6h2 and Table 4.

$\Delta_{AMIP}$  is a local measure of bias for the annual mean model data. We complement this local assessment with information on the spatial and temporal coherence of the different variables between the model results and the different observational products, by comparing large-scale statistics of their respective spatial patterns and seasonal variations in Figures 2 and 3. The statistics in the Taylor diagrams are calculated against the observational mean (see Appendix A) as if the model was a given observational product.

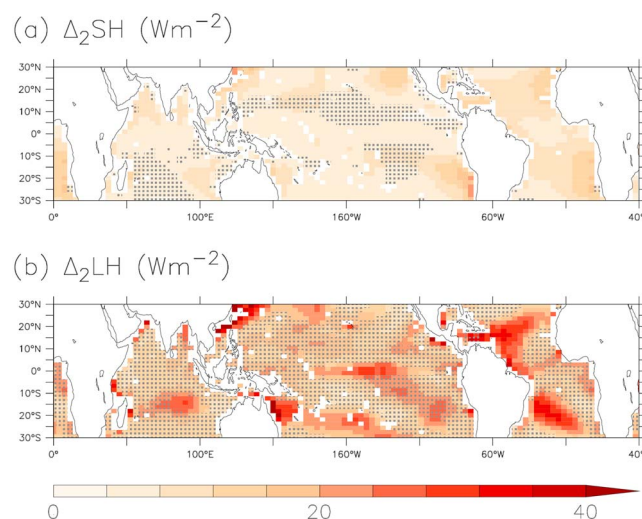
Finally, for the assessment of the representation of seasonality in the model, we also employ an estimate of the model bias for the amplitude of the climatological seasonal cycle marked with the underscore  $_{SA}$ :

$$\Delta_{AMIP}_{SA} = \overline{AMIP}_{SA} - \overline{OBS}_{SA} \tag{B4}$$

### Appendix C: RMS Criterion for Significant Model Biases

For an alternative criterion for the definition of significant model biases, we can use the typical distance between the observational products  $\Delta_2 OBS$  (equation (6)) and a similarly defined typical model-observational product distance:

$$\Delta_2 AMIP = \left[ \frac{\sum_{i=1}^{n_{AMIP}} \sum_{j=1}^{n_{OBS}} (AMIP_i - OBS_j)^2}{n_{AMIP} \cdot n_{OBS}} \right]^{0.5} \tag{C1}$$



**Figure C1.** RMS pairwise differences between the individual AMIP simulations and the non-reanalysis observational products (equation (C1)) for the climatological annual mean (a) sensible heat flux and (b) latent heat flux. Dotted regions correspond to regions where these differences are smaller than the pairwise RMS interobservation differences (equation (6)).

where  $AMIP_i$  is an individual AMIP simulation,  $OBS_j$  is an individual, non-reanalysis observational product,  $n_{AMIP}$  is the number of simulations in the AMIP ensemble, and  $n_{OBS}$  is the number of non-reanalysis products in the observational ensemble.

The model bias can thus be defined as significant where  $\Delta_2 AMIP > \Delta_2 OBS$ . Figure C1 shows two examples of  $\Delta_2 AMIP$  maps, for the two turbulent heat fluxes, with their associated regions of significant model bias left undotted.

The AMIP  $SH$  values fall on the high end of, but within the observational range (Figure 2c).

This is why  $SH$  is the only variable for which using the pairwise RMS difference criterion for identifying significant model biases results in a markedly different assessment than the maximum difference criterion  $\Delta_1$  (see Appendix B). The  $\Delta_2$  AMIP map for  $SH$  in Figure C1a shows the sensible heat flux bias to be significant over most of the tropics, while Figure 6c2, where the  $\Delta_1$  criterion is used, indicates that this bias is not significant over most tropical regions.

For all other variables analyzed, where, at most individual grid points, the model results fall either well inside or well outside the observational spread, the two criteria result in very similar regions of significant versus nonsignificant AMIP biases. This is illustrated by the comparison of the free versus dotted regions in Figure 6d2 (using the  $\Delta_1$  criterion) and Figure C1b (using  $\Delta_2$ ).

### Acknowledgments

We would like to thank the flux product developers, the Computational & Information Systems Laboratory (CISL)/National Center for Atmospheric Research (NCAR) and the International Research Institute for Climate and Society (IRI) for providing the observational data sets used in this study, as well as A. Bentamy, B. Barnier, R. Dussin, Z. Liu, A. Savtchenko, and S. Smith for technical help obtaining specific data. The simulations analyzed in this study were provided by the IPSL pole de modelisation and supported by Labex L-IPSL, which is funded by the ANR (grant ANR-10-LABX-0018) and by the European FP7 IS-ENES2 project (grant 312979). The observational data base, model results, and analysis routines used for this study can be made available upon request, pending verification with original data developers. We are grateful to S. Somot, B. Barnier, F. Hourdin, S. Smith, S. Gulev, A. Andersson, and C. Rio for useful discussions around the topics of model evaluation, observational and reanalysis flux products, and tropical atmospheric convection. Graphical representations in this manuscript have benefited from the FAST Ferret package developed by P. Brockmann. This project has been supported by a Commissariat à l'énergie atomique et aux énergies alternatives Contrat de formation par la recherche (CEA/CFR) grant awarded to A. Găinușă-Bogdan.

### References

- Andersson, A., K. Fenning, C. Klepp, S. Bakan, H. Graßl, and J. Schulz (2010), The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data—HOAPS-3, *Earth Syst. Sci. Data*, 2, 215–234.
- Barnier, B. (2001), The flux problem seen from below the ocean surface, in *Proceedings of the WCRP/SCOR Workshop on Intercomparison and Validation of Ocean-Atmosphere Flux Fields*, pp. 35–40, World Meteorol. Organ., Geneva, Switzerland.
- Bates, S. C., B. Fox-Kemper, S. R. Jayne, W. G. Large, S. Stevenson, and S. G. Yeager (2012), Mean biases, variability, and trends in air-sea fluxes and sea surface temperature in the CCSM4, *J. Clim.*, 25, 7781–7801.
- Bentamy, A., K. B. Katsaros, A. M. M.-N. nez, W. M. Drennan, E. B. Forde, and H. Roquet (2003), Satellite estimates of wind speed and latent heat flux over the global oceans, *J. Clim.*, 16, 637–656.
- Bentamy, A., S. A. Grodsky, K. Katsaros, A. M. M.-N. nez, B. Blanke, and F. Desbiolles (2013), Improvement in air-sea flux estimates derived from satellite observations, *Int. J. Remote Sens.*, 34(14), 5243–5261.
- Berry, D. I., and E. C. Kent (2009), A new air-sea interaction gridded dataset from ICOADS with uncertainty, *Bull. Am. Meteorol. Soc.*, 90, 645–656.
- Blanc, T. V. (1985), Variation of bulk-derived surface flux, stability, and roughness results due to the use of different transfer coefficient schemes, *J. Phys. Oceanogr.*, 15, 650–669.
- Bourras, D. (2006), Comparison of five satellite-derived latent heat flux products to moored buoy data, *J. Clim.*, 19, 6291–6313.
- Bourassa, M. A., P. J. Hughes, and S. R. Smith (2008), Surface turbulent flux product comparison, *Flux News*, 5, 22–24.
- Braconnot, P., and C. Frankignoul (1993), Testing model simulations of the thermocline depth variability in the tropical Atlantic from 1982 through 1984, *J. Phys. Oceanogr.*, 23(4), 626–647.
- Brodeau, L., B. Barnier, A.-M. Treguier, T. Penduff, and S. Gulev (2010), An ERA40-based atmospheric forcing for global ocean circulation models, *Ocean Model.*, 31(3–4), 88–104, doi:10.1016/j.ocemod.2009.10.005.
- Brunke, M. A., Z. Wang, X. Zeng, M. Bosilovich, and C.-L. Shie (2011), An assessment of the uncertainties in ocean surface turbulent fluxes in 11 reanalysis, satellite-derived, and combined global datasets, *J. Clim.*, 24, 5469–5493.
- Businger, J. A. (1966), Transfer of momentum and heat in the planetary boundary layer, in *Proceedings of the Symposium on Arctic Heat Budget and Atmospheric Circulation*, pp. 305–332, The Rand Corporation, Santa Monica, Calif.
- Businger, J. A. (1988), A note on the Businger-Dyer profiles, *Boundary Layer Meteorol.*, 42, 145–151.
- Chaudhuri, A. H., R. M. Ponte, G. Forget, and P. Heimbach (2013), A comparison of atmospheric reanalysis surface products over the ocean and implications for uncertainties in air-sea boundary forcing, *J. Clim.*, 26(1), 153–170.
- Chou, S.-H., E. Nelkin, J. Ardizzone, R. M. Atlas, and C.-L. Shie (2003), Surface turbulent heat and momentum fluxes over global oceans based on the Goddard satellite retrievals, version 2 (GSSTF2), *J. Clim.*, 16, 3256–3273.
- da Silva, A., A. C. Young, and S. Levitus (1994), Atlas of surface marine data 1994, volume 1: Algorithms and procedures, *Tech. Rep. 6*, U.S. Department of Commerce, NOAA, NESDIS, Washington, D. C.
- Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137, 553–597.
- Dufresne, J.-L., et al. (2013), Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5, *Clim. Dyn.*, 40(9–10), 2123–2165.
- Dyer, A. J. (1974), A review of flux-profile relationships, *Boundary Layer Meteorol.*, 7(3), 363–372.
- Fairall, C. W., et al. (2010), Observations to quantify air-sea fluxes and their role in climate variability and predictability, in *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society. OceanObs'09: Sustained Ocean Observations and Information for Society*, vol. 2, edited by J. Hall, D. E. Harrison, and D. Stammer, pp. 299–313, European Space Agency, Noordwijk, Netherlands.
- Gleckler, P. J. (2005), Surface energy balance errors in AGCMs: Implications for ocean-atmosphere model coupling, *Geophys. Res. Lett.*, 32, L15708, doi:10.1029/2005GL023061.
- Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008), Performance metrics for climate models, *J. Geophys. Res.*, 113, D06104, doi:10.1029/2007JD008972.
- Gulev, S., T. Jung, and E. Ruprecht (2007a), Estimation of the impact of sampling errors in the VOS observations on air-sea fluxes. Part I: Uncertainties in climate means, *Am. Meteorol. Soc.*, 20(2), 279–301.
- Gulev, S., T. Jung, and E. Ruprecht (2007b), Estimation of the impact of sampling errors in the VOS observations on air-sea fluxes. Part II: Impact on trends and interannual variability, *J. Clim.*, 20(2), 302–315.
- Gulev, S., C. Fairall, and V. Ryabinin (2008), Surface Fluxes and WCRP Science, *Flux News*, 5, 2–4.
- Gulev, S. K., et al. (2010), Surface energy, CO<sub>2</sub> fluxes and sea ice, in *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*, edited by J. Hall, D. E. Harrison, and D. Stammer, pp. 1–26, ESA Publication WPP-306, Venice, Italy.
- Hourdin, F., et al. (2006), The LMDZ4 general circulation model: Climate performance and sensitivity to parametrized physics with emphasis on tropical convection, *Clim. Dyn.*, 27(7–8), 787–813.
- Hourdin, F., et al. (2013), Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model, *Clim. Dyn.*, 40, 2167–2192.
- Hughes, P. J., M. A. Bourassa, J. J. Rolph, and S. R. Smith (2012), Averaging-related biases in monthly latent heat fluxes, *J. Atmos. Oceanic Technol.*, 29, 974–986.
- Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski (2008), A new sea surface temperature and sea ice boundary dataset for the community atmosphere model, *J. Clim.*, 21, 5145–5153.

- Jones, C., P. Peterson, and C. Gautier (1999), A new method for deriving ocean surface specific humidity and air temperature: An artificial neural network approach, *J. Appl. Meteorol.*, *38*(8), 1229–1245.
- Jones, S. F., and Y. Toba (eds.) (2001), *Wind Stress Over the Ocean*, Cambridge Univ. Press, Cambridge, U. K.
- Josey, S. A., and D. I. Berry (2010), Air-sea fluxes of heat and freshwater, in *MCCIP Annual Report Card 2010-11*, pp. 1–9, MCCIP Science Review, Lowestoft, U. K.
- Josey, S. A., E. C. Kent, and P. K. Taylor (1999), New insights into the ocean heat budget closure problem from analysis of the SOC air-sea flux climatology, *J. Clim.*, *12*, 2856–2880.
- Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*(3), 437–471.
- Kent, E. C., and D. I. Berry (2005), Quantifying random measurement errors in Voluntary Observing Ships' meteorological observations, *Int. J. Climatol.*, *25*, 843–856.
- Kent, E. C., S. D. Woodruff, and D. I. Berry (2007), Metadata from WMO publication no. 47 and an assessment of Voluntary Observing Ship Observation Heights in ICOADS, *J. Atmos. Oceanic Technol.*, *24*, 214–234.
- Kubota, M., N. Iwasaka, S. Kizu, M. Konda, and K. Kutsuwada (2002), Japanese Ocean Flux Data Sets with Use of Remote Sensing Observations (J-OFURO), *J. Oceanogr.*, *58*, 213–225.
- Kubota, M., A. Kano, H. Muramatsu, and H. Tomita (2003), Intercomparison of various surface latent heat flux fields, *J. Clim.*, *16*, 670–678.
- Kubota, M., N. Iwabata, M. F. Cronin, and H. Tomita (2008), Surface heat fluxes from the NCEP/NCAR and NCEP/DOE reanalyses at the Kuroshio Extension Observatory buoy site, *J. Geophys. Res.*, *113*, C02009, doi:10.1029/2007JC004338.
- Kumar, B. P., J. Vialard, M. Lengaigne, V. S. N. Murty, and M. J. McPhaden (2012), TropFlux: Air-sea fluxes for the global tropical oceans—description and evaluation, *Clim. Dyn.*, *38*(7–8), 1521–1543.
- Large, W. G., and S. G. Yeager (2009), The global climatology of an interannually varying air-sea flux data set, *Clim. Dyn.*, *33*(2–3), 341–364.
- Louis, J. F. (1979), A parametric model of vertical eddy fluxes in the atmosphere, *Boundary Layer Meteorol.*, *17*, 187–202.
- Moore, G. W. K., and I. A. Renfrew (2002), An assessment of the surface turbulent heat fluxes from the NCEP-NCAR reanalysis over the western boundary currents, *J. Clim.*, *15*, 2020–2037.
- National Center for Atmospheric Research (2013), *Atmospheric Reanalysis: Overview and Comparison Tables*, Univ. Corp. for Atmos. Res., Boulder, Colo. [Available at <http://climatedataguide.ucar.edu/reanalysis/atmospheric-reanalysis-overview-comparison-tables>.]
- Onogi, K., et al. (2007), The JRA-25 Reanalysis, *J. Meteorol. Soc. Jpn.*, *85*(3), 369–432.
- Reanalyses.org (2013), *Advancing Reanalysis*. [Available at <https://reanalyses.org/>]
- Reichler, T., and J. Kim (2008), Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model, *J. Geophys. Res.*, *113*, D05106, doi:10.1029/2007JD009278.
- Reynolds, R. W. (1988), A real-time global sea surface temperature analysis, *J. Clim.*, *1*, 75–86.
- Rio, C., et al. (2013), Control of deep convection by sub-cloud lifting processes: The ALP closure in the LMDZ5B general circulation model, *Clim. Dyn.*, *40*, 2271–2292.
- Röske, F. (2006), A global heat and freshwater forcing dataset for ocean models, *Ocean Model.*, *11*, 235–297.
- Shie, C.-L., L. S. Chiu, R. Adler, S. Gao, R. Chokngamwong, I.-I. Lin, E. Nelkin, J. Ardizzone, P. Xie, and F.-C. Wang (2010), A recently revived and produced global air-sea surface turbulent fluxes dataset—GSSTF2b: Validations and findings, in *Proceedings of the Joint 2010 CWB Weather Analysis and Forecasting and COA 5th International Ocean-Atmosphere Conference*, pp. 307–312, Chinese-American Oceanic and Atmospheric Association, Greenbelt, Maryland.
- Smith, S. D. (1988), Coefficients for sea surface wind stress, heat flux, and wind profiles as a function of wind speed and temperature, *J. Geophys. Res.*, *93*, 15,467–15,472.
- Smith, S. R., P. J. Hughes, and M. A. Bourassa (2011), A comparison of nine monthly air-sea flux products, *Int. J. Climatol.*, *31*(7), 1002–1027.
- Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*(D7), 7183–7192.
- Taylor, K. E., D. Williamson, and F. Zwiers (2000), The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations, *PCMDI Rep. 60*, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore Natl. Lab., Livermore, Calif.
- Tomita, H., M. Kubota, M. F. Cronin, S. Iwasaki, M. Konda, and H. Ichikawa (2010), An assessment of surface heat fluxes from J-OFURO2 at the KEO and JKEO sites, *J. Geophys. Res.*, *115*, C03018, doi:10.1029/2009JC005545.
- World Climate Research Programme (WCRP) (1989), WOCE surface flux determinations—A strategy for in situ measurements, *Working Group on in Situ Measurements for Fluxes WCRP-23 WMO/TD No.304*, World Meteorol. Organ., World Clim. Res. Programme, Geneva, Switzerland.
- World Climate Research Programme (WCRP) (2000), Intercomparison and validation of ocean-atmosphere energy flux fields. Final report of the Joint WCRP/SCOR Working Group on Air-Sea Fluxes (SCOR Working Group 110), *Tech. Rep. WCRP-112 (WMO/TD-1036)*, World Meteorol. Organ., World Clim. Res. Programme, Geneva, Switzerland.
- World Climate Research Programme (WCRP) (2012), Action plan for WCRP research activities on surface fluxes, *WCRP Informal 01/2012*, World Meteorol. Organ., World Clim. Res. Programme, Geneva, Switzerland.
- Wittenberg, A. T., A. Rosati, N.-C. Lau, and J. J. Plushay (2006), GFDL's CM2 Global Coupled Climate Models. Part III: Tropical Pacific Climate and ENSO, *J. Clim.*, *19*, 698–722.
- Yu, B., G. J. Boer, F. W. Zwiers, and W. J. Merryfield (2011), Covariability of SST and surface heat fluxes in reanalyses and CMIP3 climate models, *Clim. Dyn.*, *36*(3–4), 589–605.
- Yu, L., and R. A. Weller (2007), Objectively analyzed air-sea heat fluxes for the global ice-free oceans (1981–2005), *Bull. Am. Meteorol. Soc.*, *88*(4), 527–539.