The DOI for this manuscript is doi: 10.1175/BAMS-D-17-0218.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Tapiador, F., R. Roca, A. Genio, B. Dewitte, W. Petersen, and F. Zhang, 2018: Is precipitation a good metric for model performance? Bull. Amer. Meteor. Soc. doi:10.1175/BAMS-D-17-0218.1, in press.

# Is precipitation a good metric for model performance?

Francisco J. Tapiador[1,*], Rémy Roca[2], Anthony Del Genio[3], Boris Dewitte[2,4,5], Walt Petersen[6] and Fuqing Zhang[7]

[1,*] University of Castilla-La Mancha. Institute of Environmental Sciences. Department of Environmental Sciences. Earth and Space Sciences Group. Av. Carlos III s/n, 45071, Toledo, Spain. Email: francisco.tapiador@uclm.es

[2] Observatoire Midi-Pyrénées, Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS) 14, Av. Edouard Belin, 31400, Toulouse, France. Email: remy.roca@legos.obs-mip.fr, bxd@legos.obs-mip.fr

[3] NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, Email: anthony.d.delgenio@nasa.gov

[4] Centro de Estudios Avanzado en Zonas Áridas (CEAZA), Coquimbo, Chile. Email: boris.dewitte@ceaza.cl

[5] Departamento de Biología, Facultad de Ciencias del Mar, Universidad Católica del Norte, Coquimbo, Chile.

[6] NASA Marshall Space Flight Center, Earth Science Office, ST-11, National Space Science and Technology Center, 320 Sparkman Drive, Huntsville, AL, 35805. Email: walt.petersen@nasa.gov

[7] Department of Meteorology and Atmospheric Science, and Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, PA 16802. Email: fzhang@psu.edu

[*] Corresponding author

The suitability of precipitation as a metric for model performance and as a tool for model improvement is explored

**ABSTRACT**. Precipitation has often been used to gauge the performances of numerical weather and climate models, sometimes together with other variables such as temperature, humidity, geopotential, and clouds. Precipitation, however, is singular in that it can present a high spatial variability and probably the sharpest gradients amongst all meteorological fields. Moreover, its quantitative measurement is plagued with difficulties and there are even notable differences among different reference datasets. Several additional issues have yield to sometimes question its usefulness in model validation. This

essay discusses the use of precipitation for model verification and validation, and the crucial role of highly precise and reliable satellite estimates, such as those from the core observatory of NASA's Global Precipitation Mission (GPM).

Precipitation is essential for the existence of life and for human activity, too much of which can lead to flooding, a major hazard whose accurate forecast is always in demand. Too little precipitation on the other hand will incur drought, leading to crop failures, death of livestock and other potential hazards such as increased fire risk. For this reason, precipitation is one of the primary outputs of weather and climate models. Despite its significance, precipitation is an atmospheric variable that is notoriously difficult to predict in numerical weather models. It is not uncommon that models fail to pinpoint the exact location and timing of precipitation at the surface, along with its intensity and total accumulation, as well as the phase of hydrometeors.

In the climate realm, the ability of models to simulate precipitation has been described as 'dreadful' (Stephens et al. 2010). As Figure 1 shows, the dispersion in the mean precipitation pattern among 31 CMIP5 models can be large, with discrepancies of the order to the magnitude of the signal. This is not surprising as precipitation results from complex processes that are mostly parametrized in atmospheric models owing to their nonlinear nature and multiscale aspects that are still not well known and far from being sufficiently resolved. There remain significant sensitivities of the models to the use of different mixing and cloud parameterizations independent of whether the numerical core can correctly simulate the dynamics of the atmosphere (Tan et al. 2016, Cesana et al. 2017).

As a 'final' product of the modeling, precipitation suffers the multiplicative effect of errors in both thermodynamics and dynamics. In order to correctly simulate precipitation, one first has to be able to successfully model (up to some precision) longwave and shortwave radiation, the onset and strength of convection, humidity, the microphysics of liquid, solid and mixed phases. One also has indeed to model well the dynamics of the atmosphere so the air density, pressure, wind and the temperature are in the right place at the precise moment.

A key property of precipitation is that it can be spatially patchy, in contrast with the variables such as temperature, water vapor content, and wind speed that are either more smoothly-varying or the fields with more clear-cut gradients such as those featuring in temperature near the boundary between two different air masses. Considered as a scalar field, the spatial variability of rain can differ sharply from other meteorological fields (Figure 2). For solid precipitation, discrepancies are even larger. A major difference in terms of the spatiotemporal structure of instantaneous precipitation is the likely presence of large areas with a constant, zero, values with scattered, greater than zero and exponentially growing values. Such a feature is uncommon in other meteorological fields, which tend to be smoother and more consistent over time. This makes prediction extremely difficult as minor mismatches in either time or space can yield drastically different scores with errors larger than 100% not uncommon. For example, in a summer shower, one can easily transition from 50 mm/h to 0 mm/h rain rates in a few tens of meters. A behavior having such a high level of non-linearity is certainly difficult to model.

***Cons.*** There are several reasons not to privilege precipitation as a metric. Errors in modeled precipitation come from uncertainties and model shortcomings in both clouds and convection and error propagation is multiplicative. To be specific, an error of just one degree in the Sea Surface Temperature (SST) estimation around Palmén's 26 ºC threshold

(Palmén 1948) can result in convection been or not been triggered, and that can dramatically impact the mean large circulation and potentially shift precipitation regimes. Simply put, the precipitation field is a much more complicated field to interpret (and correct) although simple models taking into account this convective threshold effect show some skill for instance in the tropical zone (cf. Jauregui and Takahashi, 2017).

The complexity of the processes behind rainfall and snow is also a curse for model improvement (the major drive of model validation) since it neither eases the interpretation of the biases in models nor identifies the specific sources of the bias. This is because latent heating balances radiative cooling in the atmosphere (or alternatively, because evaporation, which balances precipitation, must balance radiative and sensible heating at the surface) in the climatological mean. Thus, by itself, precipitation biases cannot guide model improvement.

In spite of all efforts and huge advances over the last decades, precipitation is not well modeled yet, and that is a valid point against promoting its use in validation. The number of free parameters and empirical choices in microphysics modules is large, and includes intricate details such as the efficiency of drop coalescence, aerosol activation threshold, fall velocities, cloud fraction parameters, assumed droplet number concentration and entrainment rate. As an example, precipitation rate in single-moment microphysics schemes [those most commonly used in Global Circulation/Climate Models (GCMs) which advect hydrometeor mass only] varies significantly among schemes. Double-moment schemes (which also advect hydrometeor number concentration) fair better, but still show discrepancies between the methods (Shipway and Hill 2012). Aerosols and the chemistry of clouds and precipitation are key to further advancing modeling, as recognized by the recent Decadal Survey for Earth Science and Applications from Space (National Academies 2018), and the same applies to convection. However, in spite of the advances, there are still critical processes that are not modeled in detail, notably the aqueous chemistry, which is practically absent in today's GCMs.

There are many aspects of cloud physics where the exact mechanisms that produce precipitation are unknown. The same applies for the exact value of empirical parameters embedded into various parameterizations (Tapiador et al. 2018). For example, in warm clouds, collision-coalescence theories suggest that precipitation should take hours to form, yet rain often is produced within timescales of minutes. While there are many theories (e.g. specific aerosol initiating precipitation, turbulence), this and other microphysical problems remain an active area of research where more understanding is required to produce more accurate precipitation forecasts and climate projections.

Furthermore, precipitation from convective clouds depends on dynamics that is either unresolved at the global model grid scale (as for isolated cells) or is comparable to or larger than the grid size and thus in the "grey zone" where processes are partly resolved and partly parameterized (as in organized mesoscale systems). Therefore, it can be argued that modeled precipitation is still fairly incomplete and too dependent on empirical values obtained in a few field campaigns carried out over small regions of the planet.

Another fact that would favor alternatives to precipitation such as humidity, geopotential, or cloud properties as a metric is that the large sensitivity of atmospheric models to cloud and mixing parameterizations precludes validating aspects of the dynamical response to Sea Surface Temperature (SST) from precipitation observations. It should be noted that the best estimates of global precipitation continue to be inconsistent with the best estimates of the Earth's surface and atmospheric energy balance (Stephens et al., 2012). Until these are reconciled, models cannot be overly influenced by mean state biases relative to these estimates. It is important to note here that mean state biases in global models are vastly overrated as a basis for deciding which models have the best predictive power, because of the variety of tuning approaches and metrics chosen for analysis (Schmidt et al., 2017).

Regarding the potential role of precipitation in trend detection, compensating effects among different possible processes associated with climate change and precipitation ('wet get wetter' versus 'dry get drier' mechanism; i.e. areas with large precipitation amounts are expected to get even more in models and the inverse for arid zones) make the detection of trends difficult. The trends in tropical precipitation associated to anthropogenic forcing are less significant than those in SST (Cai et al. 2014), as Figure 1 illustrates. The figure shows that the dispersion of the trend in precipitation is larger than for SST relatively to the ensemble mean value, illustrating the different pattern of mean trend and dispersion of the mean trend for precipitation compared to SST. Note that the models seem to agree in the amplification of the southern branch of the double ITCZ, since the dispersion among the models is weaker there. It has also been shown that changes in the precipitation cycles in the historical period are minute (Tapiador et al. 2016). In fact, it is even doubtful that models can simulate precipitation cycles with the required accuracy and precision.

Another well-known issue in the validation of precipitation estimates are the large uncertainties in the reference data (IPCC 2013). Gauge-only, gauge plus satellite, and satellite-only datasets usually disagree in the location and quantity of precipitation (Adler et al. 2017). Gauges have known issues such as in-splash, out-splash or difficulties measuring in windy conditions; they suffer increased uncertainties and errors when the precipitation is solid rather than liquid; performing very localized measurements might not be representative of the area around; gauges have an extremely low spatial coverage (cf. Kidd et al. 2017) and usually under-sample the range of amounts which occurred in any precipitation event. Ground-based radars, which are also used to evaluate models, present large uncertainties such as the use of standardized power-law Z-R relationships which are often inaccurate for some regimes. In addition, radar often misses light precipitation (due to reflectivity being proportional to the sixth power of hydrometeor diameter and drizzle drops being small). Furthermore, while the spatial coverage of radar is quite good in developed countries such as the U.S.A., it is often very poor in the tropics, and in developing countries.

The urge for consistency in reference data has prompted initiatives such as the European Global Precipitation Climate Record that aims to build a dataset suitable for climate model validation, including the best-available data and an objective estimate of the uncertainties (Roca et al. 2014). In the near future, measurements from the Global Precipitation Measuring (GPM) mission will certainly help thanks to the dual-frequency precipitation radar and multi-frequency/polarization microwave radiometer (GMI) capabilities resident on the GPM core observatory (Skofronick-Jackson 2014). However, the GPM satellite datasets have not been collected for a long enough period (the satellite was launch in 2014) to derive the more than 20-yr long series required for validating climate models, albeit it is vital to validate hypotheses on tropical storms and hurricanes (Figure 3) and to verify the solid precipitation estimates of weather models (Bytheway and Kummerow, 2018). Moreover, there are also inherent limitations and uncertainties in the GPM derived precipitation estimates as well.

In addition to those observational issues, not all models automatically conserve moisture, which is essential for precipitation. This is especially true for semi-Lagrangian advection approaches (which are computationally less expensive than the Eulerian advection used by some models such as the Weather Research and Forecast model or WRF). In such cases, mass-conservation methods have to be applied in order to correct the issue (Zerroukat and Shipway, 2017) which represents a serious issue for validating the physics.

Precipitation is also one of the more computationally expensive parameterizations of any weather and climate model (around 10% of the total cost). Other precipitation-related processes (e.g. aerosol-cloud interactions) can also be quite expensive. Therefore, even if we get everything else correct in the model, our ability to accurately forecast precipitation will be a complex trade-off between how much computational power can be afforded to run the models quickly enough to produce operational weather forecasts and how much improvement can be gained from increasing the microphysical complexity of the model.

Such a state of affairs might suggest that precipitation is not a good metric to gauge model performance, i.e. to decide if a model is suited to the purpose it was built. In the case of weather forecasting, one primary use of a forecast model is for determining when, where, and how much in raining, but given the chaotic nature of the moist atmosphere, predictability of precipitative processes will intrinsically have decreasingly smaller predictive lead times at finer scales (Zhang et al. 2003, 2007) which means that it is next to impossible for a forecast model to precisely pinpoint precipitation in both space and time (right time, right place) given strong spectral power and variabilities of precipitation at smaller scales (Guilloteau et al. 2017; Bei and Zhang 2013).

In the case of climate, models are intended to check whether or not embedded hypotheses yield a climate consistent with observations, the consistency of which is often measured in terms of biases and correlations against instrumental records of temperature and precipitation; a recent study of Zhang et al. (2016) showed that very limited skill for either the CMIP3 or CMIP5 ensemble of models in their predictive capability for simulating regional precipitation at scales below 2000 km.

***Pros.*** There are, however, good arguments to favor precipitation as a good metric of model ability and thus favor its use for model improvement. The other side of the 'too stringent test' argument is that it has been so difficult to get it right, that precipitation should actually be considered as the ultimate test for model performance. It is hard to conceive that it would be possible to get instantaneous precipitation right for the wrong reasons at a spatial resolution of kilometers. Even if the temporal aggregation smooths the field when climatologies are built, deficiencies in models quickly reveal themselves in the precipitation field, with the double-ICTZ rain bands being a classical example (Li and Xie, 2014; Popp and Lutsko, 2017).

Disparities amongst reference precipitation data can also be a strength rather than a weakness in terms of achieving a faithful representation of nature in climate models. When different satellite estimates of rain rate disagree, important information is revealed that can help to fine tune models (Hourdin et al. 2017). For example, the considerable discrepancy between passive microwave and radar estimates of rain rate in the eastern Pacific ITCZ (Liu and Zipser 2013) revealed that assumptions about the depth or microphysical properties of rain-producing clouds that work well in some regions are not universally valid. While the issue has been known for a long time, the specific details, and crucially the mechanistic description, are better expressed in terms of precipitation.

Precipitation estimates have already proven their usefulness for model improvement. Almost half (48%) of modelers consider important or very important the use of global precipitation as a metric, and almost two thirds (65%) the same for regional patterns of precipitation (Hourdin et al. 2017). Examples of success include the use to better constrain model simulations of aerosol direct and indirect forcing (Chung and Soden 2017); the phase, amplitude and propagation of diurnal precipitation cycles (Dai et al. 1999); determination of the sensitivity of extreme precipitation to changes in temperature (Allan and Soden 2008); and critical insight into the 'dry get drier and wet get wetter' mechanism of global warming (Allan et al. 2010).

The usefulness of precipitation is also apparent when it is compared with its alternatives. For example, precipitation was instrumental in documenting the existence and propagation of the Madden–Julian Oscillation (MJO) anomalies (Madden and Julian 1994; Del Genio et al. 2015; Wang et al. 2015). Here the advantage of precipitation over the more commonly used Outgoing Longwave Radiation (OLR) is that OLR anomalies over the maritime continent can be affected by the fairly ubiquitous high cloud cover. Instead, the rain anomalies are proved to be very helpful to isolate the onset phase of the MJO, when shallow and congestus rain dominate as the biggest source of error in GCM cumulus parameterizations and prevent the development of a robust MJO. This particular case illustrates that it is precisely because of its complexity that precipitation can be superior to other variables: OLR-based indices of convection greatly overestimate surface rain over Africa, because they sense only the high cold cloud and cannot tell that rain is evaporating more strongly into the relatively dry lower troposphere there and not reaching the ground to the extent that it does in humid regions such as the Amazon (Liu and Zipser 2005, Ling and Zhang 2011).

The diurnal cycle is another good example of the relevance of precipitation as a metric. The phase of the diurnal cycle of precipitation over land is thought to be incorrect in most GCMs (e.g., Dai 1999, Yin and Porporato, 2017). However, there are some differences in the phase of the diurnal cycle depending on the dataset used. For example, rain climatologies that rely on IR measurements (e.g., TRMM 3B42) tend to peak ~3 hr earlier in the afternoon than climatologies that are based on radar data (e.g., TRMM 3B68) (Kikuchi and Wang, 2008), telling us that the former is likely biased by high clouds that are not producing rain or not producing heavy rain.

There are many other examples to favor precipitation. In tropical cyclone (TC) research, the magnitude of precipitation by itself is a key measure of the severity of the hazard (while on the other hand the evolution, structure and intensity of severe convective storms and TCs can be critically dependent on the type and amount of precipitation). Here, better estimates and better observations of precipitation physics offered by GPM (Figure 3) and other microwave satellite sensors permit the testing of assumptions with unprecedented capabilities (e.g., Sieron et al. 2017, 2018), providing new analytical capabilities to investigate emerging phenomena such as TCs landing in Europe (cf. Tapiador et al. 2007). Amongst other findings it appears that for TCs the amount of surface precipitation is dominantly controlled by dynamics (water lifting) while the role of microphysical processes is secondary (but still important).

***Fundamental reasons to favor precipitation.*** There are also fundamental physical reasons to favor precipitation as a metric to elucidate processes still poorly represented in models. One is the connection between precipitation and the atmospheric energy budget (L'Ecuyer et al., 2015). Changes in global mean precipitation are determined by changes in radiative cooling of the atmosphere (Stephens and Ellis, 2008), so it is extremely important to be as precise as possible in determining such changes if the model is intended to understand changes in the radiative forcing, either by natural or anthropogenic causes. In the tropics, mean precipitation and the extreme of the distribution is largely dominated by organized mesoscale convective systems (Roca et al., 2014, Rossow et al., 2015) and the trends in precipitation are also related to the fate of organized convection (Tan et al. 2015). Representation of organized mesoscale systems in GCMs is still in its infancy (Del Genio et al., 2012) while grand-domain CRM simulations become more and more available. Both contribute to making precipitation in the tropics important for gauging new generation model performances, and therein comparison with observations is critical. The partitioning of rain into convective and stratiform components is crucial to the latent heating profile of convective systems, because the former peaks in the lower/mid-troposphere while the latter peaks in the upper troposphere. This affects the tropical general circulation (Schumacher et al. 2004). GCMs have so far been able to capture the major features of the climate without representing organized mesoscale systems, which show a transition from bottom to top-heavy heating over the life cycle (by underestimating convective entrainment and over-producing deep penetrative convection that penetrates too deeply, and thus capturing some of the upper-level heating as an artifact of this error). Getting the right answer for the wrong reason in a climatological mean field in this way is one example of the limitations of using mean fields as metrics. The model parameterization errors only become obvious when higher-order variability metrics such as the MJO or the continental diurnal cycle, which depend on the timing of the transition from bottom-heavy to top-heavy latent heating profiles, are used for evaluation. The latent heating algorithms that have been developed for satellite rain data diagnose this partitioning from characteristics of the rain and reflectivity fields to produce realistic heating profiles and thus to improve representation of this heating in GCMs (Tao and Shi 2016). The same arguments can also be applied to high-resolution, limited area models, which are commonly used for weather forecasting but in the last few years have been extended to climate predictions as well (Kendon et al. 2017).

Processes of SST/wind/precipitation interaction are also likely to be involved in long-term trends and variability in the surface circulation in the tropics. For instance, while in the sub-tropical eastern boundary upwelling regions, an increase of the equatorward winds is expected (and observed in some regions) owing to the poleward displacement and intensification of the anticyclone/Hadley cells, in the tropical Pacific region, the trends in upwelling-favorable winds are more ambiguous and are sensitive to concurrent changes in sea surface temperature and rainfall, as observed off Peru from coupled model experiments (Belmadani et al., 2014). Therefore, processes associated with moist convection and subsidence in the far eastern Pacific are likely important to understand trends in upwelling systems and their investigation will benefit from precipitation observations and will require model evaluations based on those.

Another fundamental reason for using precipitation as a model-comparison metric is that precipitation is often considered as a proxy for inferring change statistics in extreme events. To name but one example, the precipitation response to SST during strong El Nino events encapsulates the process associated with the nonlinear amplification of the Bjerknes feedback (Takahashi and Dewitte 2016) and therein can be considered a better metric of ENSO extremes than SST anomalies alone. Thus, the relationship between precipitation in the eastern equatorial Pacific (NINO3 region) and the SST gradient near the equatorial region during El Nino exhibit a marked nonlinear pattern that enhances/eases the detection of extreme events. In fact, a precipitation-based definition of an extreme El Niño event (those El Niño for which the NINO3 rainfall index is above 5 mm/day) has been proposed recently which is based on the precipitation anomalies averaged over the NINO3 (150°W-90°W; 5°S-5°N) region (Cai et al. 2014, 2017). Based on this precipitation-based index, Cai et al. (2014) analyzed CMIP3 and CMIP5 models and found a doubling in the occurrence of extreme El Niño events in the future in response to greenhouse warming, while no significant change in statistics in extreme El Nino events is found based on the "classical" NINO34 SST index. Power et al. (2013) also shows that ENSO-driven precipitation exhibits a clearer longer-term change than SST anomalies. Thus, precipitation may be seen as a “better” field to reveal/diagnose/quantify the non-linear relationship between the variability in the climate system and changes in mean state.

***Summary.*** This essay discusses the usefulness of precipitation for model verification and validation, and the crucial role of highly precise and reliable satellite estimates, such as those from the GPM core observatory, to test model hypotheses and assumptions. It is widely acknowledged that good climate models are those capable of correctly simulating the MJO, ENSO or the mean ITCZ, but it should be noted that those processes are also precisely identified as a fingerprint in the precipitation field (Figure 4), a fact that reinforces use of precipitation for model verification.

Yet however there are several other compelling reasons to favor precipitation as a metric of model performance, not the least of which is assuring a tough test of model performance. At the end, it can be said that the ultimate test of a fully-fledged coupled model is to get precipitation right, a demand that is also spurred by the societal demand for more reliable forecasts of extreme rainfall events, and that includes weather and climate models. As noted, models still have a limited ability to simulate precipitation at adequate temporal and spatial resolution. Such shortcomings demonstrate not only the need to continue devoting resources to improving models, but also suggest that precipitation can be used as a stringent quantitative criterion to evaluate model advances.

Concomitantly, the evaluations of models based on precipitation reinforce the need to continually improve the precipitation estimates themselves. Developments in the observation network should follow the path imposed by progresses in modeling that continue to reveal the importance of scale interactions in convective activity and its upscaling effect on climate. The more we will get to the higher-resolution and more complex models, the more pressing the need to validate aspects of the circulation that had been disregarded or poorly modeled so far, and this includes precipitation physics at the first place.

Finally, it is worth remembering that some of the processes ultimately producing precipitation occur at planetary scales but that some others develop at very small scales (microns). We are unlikely to ever be able to resolve the smallest scales in a weather or climate model. Precipitation will continue to require parameterizations and therefore the resulting precipitation will be highly dependent on the empirical choices and assumptions embedded into these. Therein the likely continuing suitability of this crucial element for life to gauge model performance.


**Acknowledgements**

F. J. Tapiador acknowledges projects CGL2013-48367-P and CGL2016-80609-R. B. Dewitte acknowledges supports from FONDECYT (projects 1171861). We thank the three anonymous reviewers for their constructive comments. Katerina Goubanova (CEAZA) is thanked for providing WRF model simulations performed using HPC resources from CALMIP (Grant 2014-1044).


**References:**


Adler, R., G. Gu, M. Sapiano, J. Wang, G. Huffman, 2017: Global Precipitation: Means, Variations and Trends During the Satellite Era (1979-2014). Surveys in Geophysics 38: 679-699, doi:10.1007/s10712-017-9416-4

Allan, R. P., B. J. Soden, V. O. John, W. Ingram, and P. Good, 2010: Current changes in tropical precipitation. *Environ. Res. Lett.*, **5**, 25205, doi:10.1088/1748-9326/5/2/025205.

Allan R P, Soden B J, John V O, Ingram W and Good P, 2010: Current changes in tropical precipitation Environ. Res. Lett. 5 025205

Bei, N., and F. Zhang, 2014: Mesoscale Predictability of Moist Baroclinic Waves:Variable and Scale Dependent Error Growth. Advances in Atmospheric Sciences, 31, 995–1008.

Belmadani, A. , V. Echevin, F. Codron, K. Takahashi, and C. Junquas, 2014: What dynamics drive future wind scenarios for coastal upwelling off Peru and Chile? Clim. Dyn. 43, 1893-1914, doi:10.1007/s00382-013-2015-2

Bytheway, J. L., C. D. Kummerow, and F. Collins, 2018: Consistency Between Convection Allowing Model Output and Passive Microwave Satellite Observations. Journal of Geophysical Research, doi:10.1002/2017JD027527.

Cai, W., S. Borlace, M. Lengaigne, P. van Rensch, M. Collins, G. Vecchi, A. Timmermann, A. Santoso, M. McPhaden, L. Wu, M. H. England, G. Wang, E. Guilyardi, and F.-F. Jin, 2014: Increasing Frequency of Extreme El Niño Events due to Greenhouse Warming. Nature Climate Change, 4, 111-116, DOI: 10.1038/nclimate2100

Cai W. G. Wang A. Santoso, X. Lin, and L.Wu, 2017: Definition of Extreme El Niño and Its Impact on Projected Increase in Extreme El Niño Frequency, Geophys. Res. Lett. https://doi.org/10.1002/2017GL075635

Cesana, G., Suselj, K., & Brient, F., 2017: On the dependence of cloud feedbacks on physical parameterizations in WRF aquaplanet simulations. Geophysical Research Letters, 44, 10,762–10,771. https://doi.org/10.1002/2017GL074820

Dai, A., F. Giorgi, and K. E. Trenberth, 1999: Observed and model simulated precipitation diurnal cycle over the contiguous United States. *J. Geophys. Res.*, **104**, 6377–6402, doi:https://doi.org/10.1029/98JD02720.

Del Genio, A. D., 2012: Representing the Sensitivity of Convective Cloud Systems to Tropospheric Humidity in General Circulation Models. *Surv. Geophys.*, **33**, 637–656, doi:10.1007/s10712-011-9148-9. http://dx.doi.org/10.1007/s10712-011-9148-9.

Del Genio, A.D., J. Wu, A.B. Wolf, Y. Chen, M. Yao, and D. Kim, 2015: Constraints on Cumulus Parameterization from Simulations of Observed MJO Events. J. Climate, 28, 6419–6442, https://doi.org/10.1175/JCLI-D-14-00832.1

Guilloteau, C., E. Foufoula-Georgiou, and C.D. Kummerow, 2017: Global multiscale evaluation of satellite passive microwave retrieval of precipitation during the TRMM and GPM eras: effective resolution and regional diagnostics for future algorithm development, Journal of Hydrometeorology, 18(11), 3051-3070, doi:10.1175/JHM-D-17-0087.1.

Hourdin, F., T. Mauritsen, A. Gettelman, J. Golaz, V. Balaji, Q. Duan, D. Folini, D. Ji, D. Klocke, Y. Qian, F. Rauser, C. Rio, L. Tomassini, M. Watanabe, and D. Williamson, 2017: The Art and Science of Climate Model Tuning. Bull. Amer. Meteor. Soc., 98, 589–602, https://doi.org/10.1175/BAMS-D-15-00135.1

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Jauregui, Y.R. y Takahashi, K., 2017: Simple physical-empirical model of the precipitation distribution based on a tropical sea surface temperature threshold and the effects of climate change. Climate Dynamics. doi:10.1007/s00382-017-3745-3

Kidd, C., A. Becker, G.J. Huffman, C.L. Muller, P. Joe, G. Skofronick-Jackson, and D.B. Kirschbaum, 2017: So, How Much of the Earth's Surface Is Covered by Rain Gauges?. Bull. Amer. Meteor. Soc., 98, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1

Kikuchi, K. and B. Wang, 2008: Diurnal Precipitation Regimes in the Global Tropics. J. Climate, 21, 2680–2696, https://doi.org/10.1175/2007JCLI2051.1

Kendon, E.J., N. Ban, N.M. Roberts, H.J. Fowler, M.J. Roberts, S.C. Chan, J.P. Evans, G. Fosser, and J.M. Wilkinson, 2017: Do Convection-Permitting Regional Climate Models Improve Projections of Future Precipitation Change?. Bull. Amer. Meteor. Soc., 98, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1

L'Ecuyer, T. S., and Coauthors, 2015: The observed state of the energy budget in the early twenty-first century. *J. Clim.*, **28**, 8319–8346, doi:10.1175/JCLI-D-14-00556.1.

Li, G. and S.-P. Xie, 2014: Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial pacific cold tongue and double ITCZ problems. J Clim, 27 (4), 1765,1780.

Ling, J., and C. Zhang, 2011: Structural evolution in heating profiles of the MJO in global reanalyses and TRMM retrievals. J. Climate, 24, 825–842, doi:10.1175/2010JCLI3826.1.

Liu, C., and E. J. Zipser, 2005: Global distribution of convection penetrating the tropical tropopause, J. Geophys. Res., 110, D23104, doi:10.1029/2005JD006063.

Liu C., and E. Zipser, 2013: Regional variation of morphology of the organized convection in the tropics and subtropics, Part I: regional variation, J. Geophy. Res., 118, 453–466, doi:10.1029/2012JD018409.

Madden R.E. and Julian P.R., 1994: Observations of the 40-50 day tropical oscillation-A review. Mon Wea Rev 122: 814–837

National Academies of Sciences, Engineering, and Medicine, 2018: Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space. Washington, DC: The National Academies Press. https://doi.org.10.17226/24938.

Palmén, E. H., 1948: On the formation and structure of tropical hurricanes. Geophysica, 3, 26–39.

Popp, M. and N. J. Lutsko, 2017: Quantifying the zonal-mean structure of tropical precipitation. Geophysical Research Letters, 44, DOI: 10.1002/2017GL075235.

Power, S., Delage, F., Chung, C., Kociuba, G., Keay, K., 2013: Robust twenty-first-century projections of El Niño and related precipitation variability. Nature. 502, 541-545.

Power, S.B., F.P.D. Delage, C.T.Y. Chung, H. Ye, B.F. Murphy, 2017: Humans have already increased the risk of major disruptions to Pacific rainfall. Nature Communications, doi:10.1038/ncomms14368.

Roca, R., J. Aublanc, P. Chambon, T. Fiolleau, and N. Viltard, 2014: Robust observational quantification of the contribution of mesoscale convective systems to rainfall in the tropics. *J. Clim.*, **27**, 4952–4958, doi:10.1175/JCLI-D-13-00628.1.

Rossow, W. B., A. Mekonnen, C. Pearl, and W. Goncalves, 2013: Tropical precipitation extremes. *J. Clim.*, **26**, 1457–1466, doi:10.1175/JCLI-D-11-00725.1.

Schmidt, G.A., D. Bader, L.J. Donner, G.S. Elsaesser, J.-C. Golaz, C. Hannay, A. Molod, R. Neale, and S. Saha, 2017: Practice and philosophy of climate model tuning across six U.S. modeling centers. *Geosci. Model Dev.*, **10**, 3207-3223, doi:10.5194/gmd-10-3207-2017.

Schumacher, C., R.A. Houze, and I. Kraucunas, 2004: The Tropical Dynamical Response to Latent Heating Estimates Derived from the TRMM Precipitation Radar. J. Atmos. Sci., 61, 1341–1358, https://doi.org/10.1175/1520-0469(2004)061<1341:TTDRTL>2.0.CO;2

Sieron S. B., E.E. Clothiaux, F. Zhang, Y. Lu and J. Otkin, 2017: Fast Radiative Transfer Modeling for All-Sky Microwave Satellite Radiances: Modifying CRTM with Microphysics-Consistent Cloud Optical Properties. Journal of Geophysical Research - Atmosphere, 122, DOI: 10.1002/2017JD026494.

Sieron, S. B., F. Zhang, E. E. Clothiaux, L. N. Zhang, and Y. Lu, 2018: Representing Precipitation Ice Species with Both Spherical and Non-Spherical Particles for Microphysics-Consistent Cloud Microwave Scattering Properties. *Journal of Advances in Modeling Earth Systems*, **10**, 1011–1028.

Shipway, B. J. and Hill, A. A. (2012), Diagnosis of systematic differences between multiple parametrizations of warm rain microphysics using a kinematic framework. Q.J.R. Meteorol. Soc., 138: 2196-2211. doi:10.1002/qj.1913

Skofronick-Jackson, G., W.A. Petersen, W. Berg, C. Kidd, E.F. Stocker, D.B. Kirschbaum, R. Kakar, S.A. Braun, G.J. Huffman, T. Iguchi, P.E. Kirstetter, C. Kummerow, R. Meneghini, R. Oki, W.S. Olson, Y.N. Takayabu, K. Furukawa, and T. Wilheit, 2017: The Global Precipitation Measurement (GPM) Mission for Science and Society. Bull. Amer. Meteor. Soc., 98, 1679–1695, https://doi.org/10.1175/BAMS-D-15-00306.1

Stephens, G. L., and T. D. Ellis, 2008: Controls of Global-Mean Precipitation Increases in Global Warming GCM Experiments. *J. Clim.*, **21**, 6141–6155, doi:10.1175/2008jcli2144.1. http://dx.doi.org/10.1175/2008jcli2144.1.

Stephens, G.L., J. Li, M. Wild, C.A. Clayson, N. Loeb, S. Kato, T. L'Ecuyer, P.W. Stackhouse, M. Lebsock and T. Andrews, 2012: An update on Earth's energy balance in light of the latest global observations. *Nature Geosci.* **5**, 691-696.

Stephens, G. L., T. L'Ecuyer, R. Forbes, A. Gettlemen, J.-C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes, 2010: Dreary state of precipitation in global models, J. Geophys. Res., 115, D24211, doi:10.1029/2010JD014532.

Takahashi K. and B. Dewitte, 2016: Strong and Moderate nonlinear El Niño regimes. Climate dynamics, doi: 10.1007/s00382-015-2665-3.

Tan, J., C. Jakob, W. B. Rossow, and G. Tselioudis, 2015: Increases in tropical rainfall driven by changes in frequency of organized deep convection. *Nature*, **519**, 451–

454, doi:10.1038/nature14339. http://dx.doi.org/10.1038/nature14339.

Tan I., T. Storelvmo, M. D. Zelinka, 2016: Observational constraints on mixed-phase clouds imply higher climate sensitivity. Science, DOI: 10.1126/science.aad5300

Tao, W.-K., and X. Li, 2016: The relationship between latent heating, vertical velocity, and precipitation processes: The impact of aerosols on precipitation in organized deep convective systems, J. Geophys. Res. Atmos., 121, 6299–6320, doi:10.1002/2015JD024267.

Tapiador, F.J., M.A. Gaertner, R. Romera, and M. Castro, 2007: A Multisource Analysis of Hurricane Vince. Bull. Amer. Meteor. Soc., 88, 1027–1032, https://doi.org/10.1175/BAMS-88-7-1027

Tapiador, F. J., A. Behrangi, Z. S. Haddad, D. Katsanos, and M. de Castro, 2016: Disruptions in precipitation cycles: Attribution to anthropogenic forcing, J. Geophys. Res. Atmos., 121, 2161–2177, doi:10.1002/2015JD023406.

Tapiador, F.J., Sánchez, J.L. and García-Ortega, E., 2018: Empirical Values and Assumptions in the Microphysics of Numerical Models, Atmospheric Research, submitted.

Wang, S., A. H. Sobel, F. Zhang, Y. Qiang Sun, Y. Yue, L. Zhou, 2015: Regional Simulation of the October and November MJO Events Observed during the

CINDY/DYNAMO Field Campaign at Gray Zone Resolution. Journal of Climate, 28, 2097-2119.

Yin J, Porporato A., 2017: Diurnal cloud cycle biases in climate models. Nat Commun 8:2269. doi: 10.1038/s41467-017-02369-4

Zerroukat, M. and Shipway, B. J. (2017), ZLF (Zero Lateral Flux): a simple mass conservation method for semi-Lagrangian-based limited-area models. Q.J.R. Meteorol. Soc, 143: 2578-2584. doi:10.1002/qj.3108

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. Journal of the Atmospheric Sciences, 60, 1173-1185.

Zhang, F., N. Bei, R. Rotunno, C. Snyder and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Cloud-resolving experiments and multistage error growth dynamics. Journal of the Atmospheric Sciences, 64, 3579-3594.

Zhang, F., W. Li and M. E. Mann, 2016: Scale-dependent Regional Climate Predictability over North America Inferred from CMIP3 and CMIP5 Ensemble Simulations. Advances in Atmospheric Sciences, 33, 905-918.

## Figure captions

**Figure 1**: Mean and trend in SST and Precipitation: (a) Ensemble mean of the mean precipitation and (d) dispersion (root mean square) amongst the ensemble for the historical runs (1920-2005) of 31 CMIP5 models (unit=mm/day). The thick red line indicates the 27°C isotherm for the ensemble mean. Ensemble mean of the long-term trend in (c) SST and (d) precipitation for the RCP8.5 scenario simulations (2006-2095) (unit is °C/decade for SST and mm/day/decade for precipitation). Dispersion of the trend in (e) SST and (f) precipitation amongst the ensemble for the for the RCP8.5 scenario simulations (2006-2095).

**Figure 2:** Fields of several meteorological variables from the UCLM-WRF model over Korea. The spatial variability is measured using the semivariance (normalized so fields can be compared). As the lag distance varies, the variables become more and more decorrelated. Note the peculiar spatial decorrelation of precipitation. The grid size of the simulations is 300m, and the fields are instantaneous estimates. Data from the ICE-POP2018 Winter Olympics Campaign.

**Figure 3**: GPM core observatory dissection of hurricane Maria on 18 September 2017. The figure illustrates the ability of the GPM-Core Observatory satellite to map combined GMI radiometer-estimated precipitation rates in a broad 2-D swath with a coincident narrower-swath of 3-D storm structure and hydrometeor phase profiled using GPM DPR. Here warm colors indicate liquid precipitation rates and cool colors indicate precipitation rates in the ice-phase. Credit: NASA Goddard Space Flight Center, Science Visualization Studio.

**Figure 4**: Sensitivity of precipitation within the ITCZ in the eastern tropical Pacific to cumulus (CU) and planetary boundary layer (PBL) parametrizations in WRF (horizontal resolution= 30km): (a) mean precipitation for March 2007 from TRMM,

(b) ensemble mean and (c) dispersion (i.e. standard deviation) for precipitation in 25 simulations of March 2007 using different combinations of 5 CU and 5 PBL parametrizations, and (d) Characteristics of the ITCZ over the two regions (0°-15°N,130°W-100°W) and (0°-15°S, 130°W-100°W) in observations (grey bars) and the 25 simulations (color bars): Bars indicate the latitudinal extension of the branches of the ITCZ. The thick black line indicates the latitude of the relative maximum precipitation during this particular month. The number near each bar provide the value of total precipitation and the bar thickness is proportional to this value.

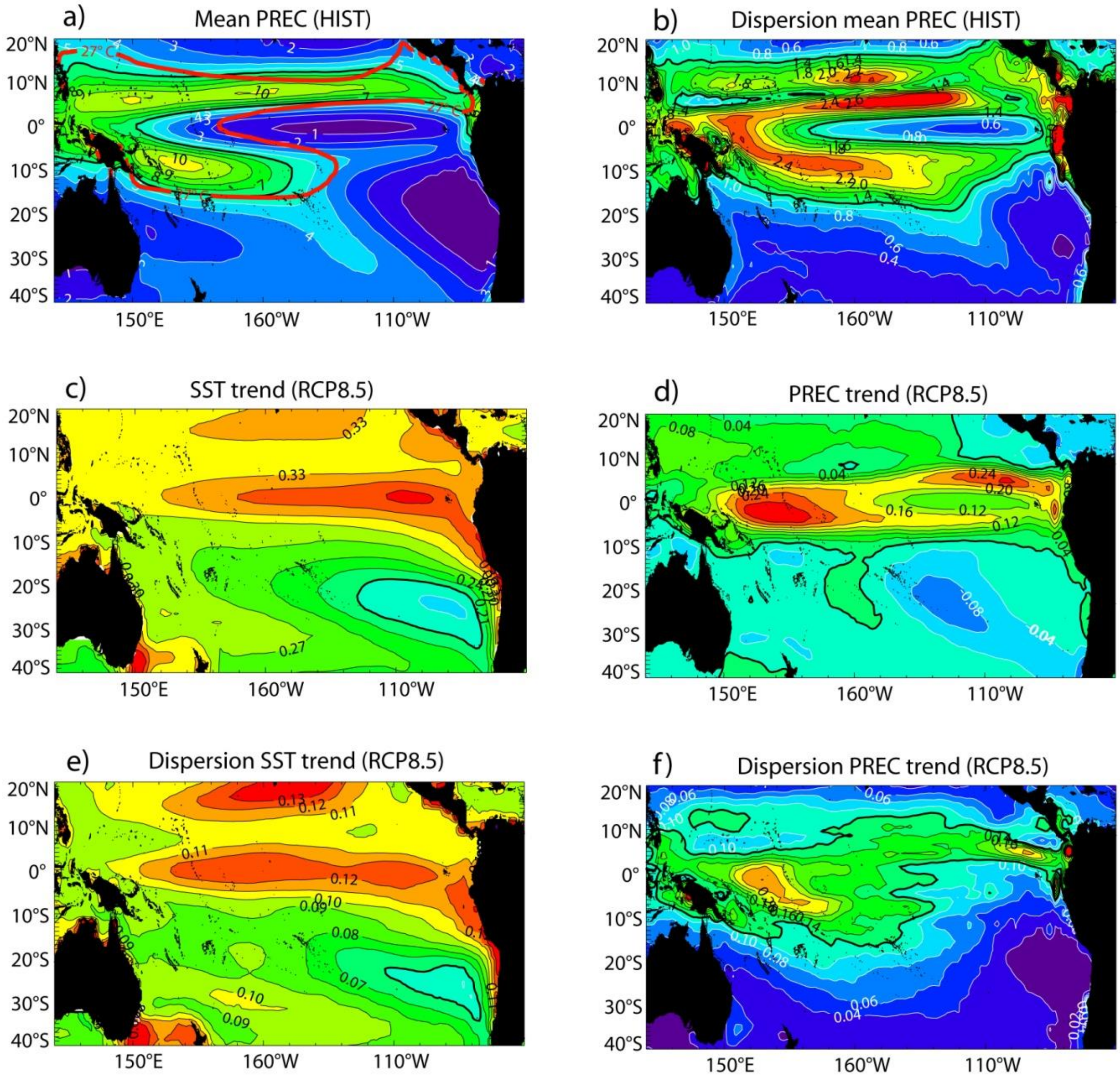


**Figure 1**: Mean and trend in SST and Precipitation: (a) Ensemble mean of the mean precipitation and (d) dispersion (root mean square) amongst the ensemble for the historical runs (1920-2005) of 31 CMIP5 models (unit=mm/day). The thick red line indicates the 27°C isotherm for the ensemble mean. Ensemble mean of the long-term trend in (c) SST and (d) precipitation for the RCP8.5 scenario simulations (2006-2095) (unit is °C/decade for SST and mm/day/decade for precipitation). Dispersion of the trend in (e) SST and (f) precipitation amongst the ensemble for the for the RCP8.5 scenario simulations (2006-2095).

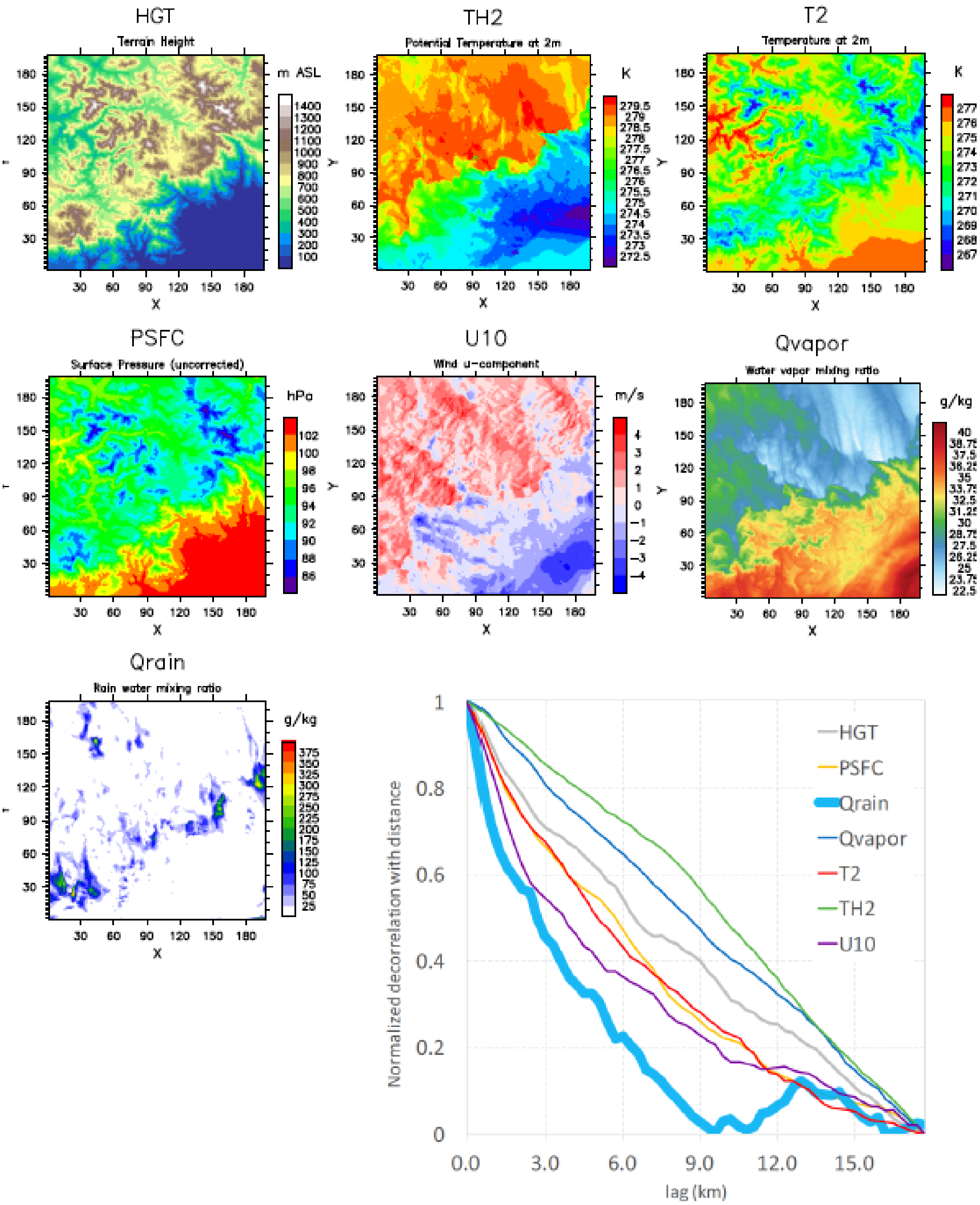


**Figure 2:** Fields of several meteorological variables from the UCLM-WRF model over Korea. The spatial variability is measured using the semivariance (normalized so fields can be compared). As the lag distance varies, the variables become more and more decorrelated. Note the peculiar spatial decorrelation of precipitation. The grid size of the simulations is 300m, and the fields are instantaneous estimates. Data from the ICE-POP2018 Winter Olympics Campaign.

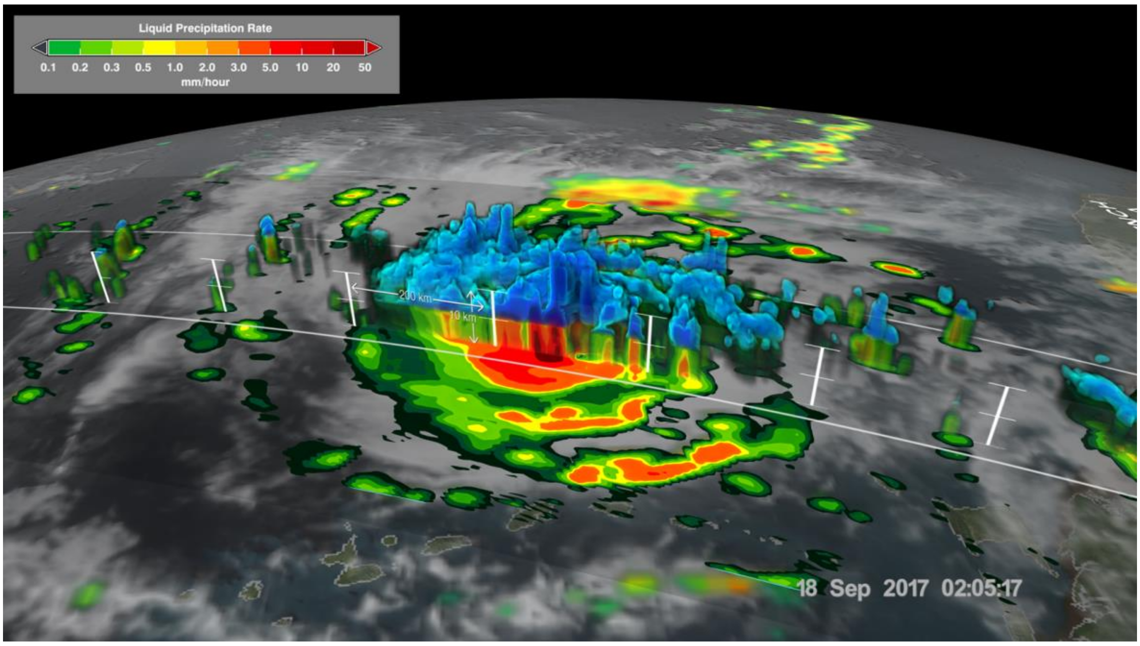


**Figure 3**: GPM core observatory dissection of hurricane Maria on 18 September 2017. The figure illustrates the ability of the GPM-Core Observatory satellite to map combined GMI radiometer-estimated precipitation rates in a broad 2-D swath with a coincident narrower-swath of 3-D storm structure and hydrometeor phase profiled using GPM DPR. Here warm colors indicate liquid precipitation rates and cool colors indicate precipitation rates in the ice-phase. Credit: NASA Goddard Space Flight Center, Science Visualization Studio.

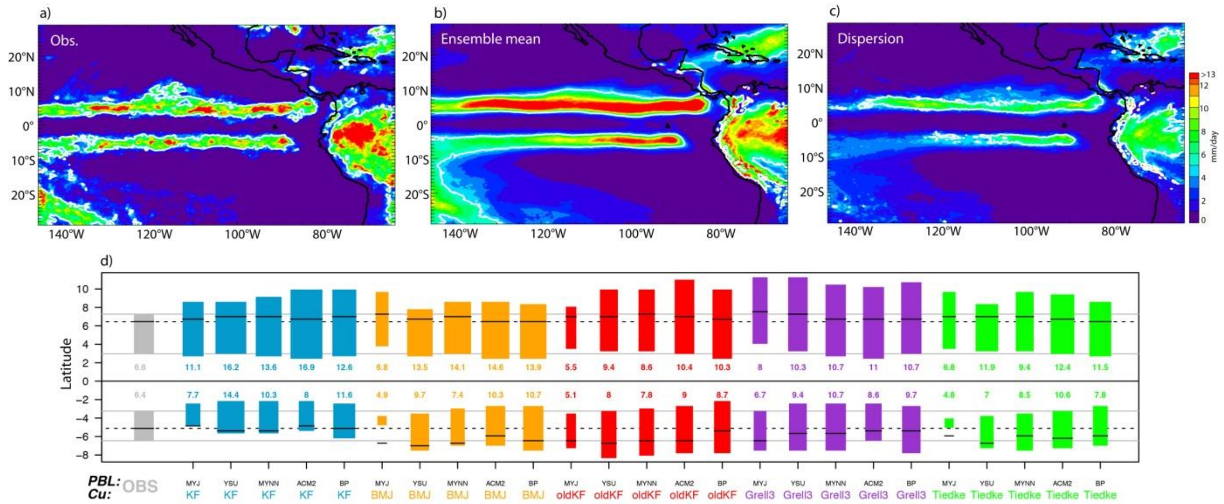


**Figure 4**: Sensitivity of precipitation within the ITCZ in the eastern tropical Pacific to cumulus (CU) and planetary boundary layer (PBL) parametrizations in WRF (horizontal resolution= 30km): (a) mean precipitation for March 2007 from TRMM, (b) ensemble mean and (c) dispersion (i.e. standard deviation) for precipitation in 25 simulations of March 2007 using different combinations of 5 CU and 5 PBL parametrizations, and (d) Characteristics of the ITCZ over the two regions (0°-15°N,130°W-100°W) and (0°-15°S, 130°W-100°W) in observations (grey bars) and the 25 simulations (color bars): Bars indicate the latitudinal extension of the branches of the ITCZ. The thick black line indicates the latitude of the relative maximum precipitation during this particular month. The number near each bar provide the value of total precipitation and the bar thickness is proportional to this value.