High sensitivity of tropical precipitation to local sea-surface temperature

Article

Accepted Version

High sensitivity of tropical precipitation to local sea-surface temperature

Peter Good[1*], Robin Chadwick[1,2], Christopher E. Holloway[3], John Kennedy[1], Jason A. Lowe[1,4], Romain Roehrig[5], Stephanie S. Rushley[6]

[1]MetOffice Hadley Centre, Exeter, United Kingdom.

[2]Global Systems Institute, University of Exeter.

[3]Department of Meteorology, University of Reading, Reading, United Kingdom.

[4]Priestley International Centre for Climate, University of Leeds, United Kingdom.

[5]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

[6]Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195, USA.

Abstract

Precipitation and atmospheric circulation are the coupled processes through which tropical ocean surface temperatures drive global weather and climate[1–5]. Local ocean surface warming tends to increase precipitation, but this local control is hard to disentangle from

remote effects of conditions elsewhere. Such remote effects occur, for example, from El Niño Southern Oscillation (ENSO) events in the equatorial Pacific, which alter precipitation across the tropics. Atmospheric circulations associated with tropical precipitation are predominantly deep, extending up to the tropopause. Shallow atmospheric circulations[6–8], impacting the lower troposphere, also occur, but the importance of their interaction with precipitation is unclear. Uncertainty in precipitation observations[9,10], and limited observations of shallow circulations[11], further obstruct understanding of the ocean's influence on weather and climate. Despite decades of research, persistent biases remain in many numerical model simulations[12–18], including excessively-wide tropical rainbands[14,18], the 'double-intertropical convergence zone (ITCZ) problem'[12,16,17] and too-weak responses to ENSO[15]. These demonstrate stubborn gaps in our understanding, reducing confidence in forecasts and projections. Here we show that the real world has a high sensitivity of seasonal tropical precipitation to local sea-surface temperature. Our best observational estimate is 80% precipitation change per g/kg change in the saturation specific humidity (itself a function of the ocean surface temperature). This observed sensitivity is higher than in 43 of the 47 climate models studied, and is associated with strong shallow circulations. Models with more realistic sensitivity have smaller biases across a wide range of metrics. Our results apply to both temporal and spatial variation, over regions where climatological precipitation is around 1 mm/day or greater. Novel analysis of multiple independent observations, physical constraints and model data, underpin these findings. The spread in model behaviour is further linked to differences in shallow convection, providing a focus for accelerated research, to improve seasonal forecasts through multidecadal climate projections.

Main paper

We first define a measure (k_{qsat}) of the sensitivity of seasonal mean precipitation to variation in local sea surface temperature (SST). We will show that k_{qsat} is a key property of the atmosphere, using it to link diverse gaps in understanding to a limited subset of physical mechanisms. Precipitation increases non-linearly with SST[19]. Since tropical precipitation increases roughly exponentially with column atmospheric water vapour[20,21]; and over seasonal or longer timescales, SST variation forces variation in column water vapour[22], via differences in saturation specific humidity of the ocean surface (q_{sat}, Methods), we define k_{qsat} as follows:

$$\log_e(P_1/P_0) \approx k_{qsat} * (q_{sat,1} - q_{sat,0}) + \langle\text{other processes}\rangle \qquad \text{Equation 1.}$$

This describes the variation in precipitation (from P_0 to P_1) driven by local variation in q_{sat} (from $q_{sat,0}$ to $q_{sat,1}$). This approximation is validated within the calculation of k_{qsat} (Methods). Moist static energy arguments[19] also predict a roughly exponential relationship between q_{sat} and P. k_{qsat} quantifies the combined effect of the physical processes by which local SST anomalies affect precipitation at the same location. The `other processes' term includes the effects of internal atmospheric variability independent of SST, and of remote forcing from land or SST elsewhere, which can be large at individual locations or times. In order to estimate k_{qsat}, we filter out these other processes, by combining information from multiple locations and times (see Methods). We evaluate k_{qsat} from interannual variability, with P_0 and $q_{sat,0}$ taken as seasonal climatological means at each location for each season. However, our estimates of k_{qsat} are shown to be also informative about spatial variations in precipitation.

k_{qsat} relates most directly to the strength of percentage variations in precipitation. Writing Equation 1 in exponential form, percentage precipitation differences are a function of k_{qsat} and q_{sat}:

$$(P_1 – P_0)/P_0 * 100 \approx 100 [\exp(k_{qsat} * (q_{sat} - q_{sat,0})) – 1]. \qquad \text{Equation 2}$$

Absolute differences depend also on the reference precipitation P_0:

$$P_1 – P_0 \approx P_0[(\exp(k_{qsat} * (q_{sat} - q_{sat,0})) – 1], \qquad \text{Equation 3}$$

(in absolute terms, precipitation variations are largest in regions of large mean precipitation[19]. However, we will show that spatial variation in P_0 itself also depends partly on k_{qsat}.

Validating satellite observations

Given uncertainty in precipitation observations[9,10], we perform a high-precision evaluation of log(precipitation) (as Equation 1) from two satellite datasets: TRMM[23,24] (3B43, v7) and GPCP[25] (v2.3), both for 1998-2015. We do this (Methods) using in-situ raingauge data from 89 buoys of the Global Tropical Moored Buoy Array (GTMBA)[26–28]. Satellite-GTMBA validation is challenging: on top of satellite error[9,10], the GTMBA point observations include noise from small-scale variability unresolved by satellite data, missing data, error in individual raingauges and wind undercatch[29]. Our method reduces this noise considerably, giving a tight relationship between GTMBA and TRMM data (Figure 1a). Critically, the best fit gradient ≈1, so TRMM accurately retrieves differences in log precipitation. On the other hand, GPCP underestimates differences in log precipitation (Figure 1b, gradient > 1; differences between TRMM and GPCP emerge primarily at low precipitation[9]), although GPCP is more suitable over larger spatial scales (Extended data Figure 1). Since TRMM captures differences in log(precipitation) more accurately than GPCP, TRMM is used below.

Model precipitation simulations

We first highlight precipitation biases in 28 atmospheric models from the fifth Coupled Model Intercomparison Project (Methods), each forced by observed SST (CMIP5 AMIP experiment; Figure 2a-f). We quantify temporal, seasonal and spatial variation in precipitation: the 1997-98 El Niño divided by the mean of the 1998-2000 La Niñas; Aug-Oct divided by Feb-Mar seasons, and precipitation scaled by its latitudinal maximum. Our metrics coincide with significant spatial or temporal differences in SST (Methods). Spatial variation across the west Pacific is excluded, for example, because spatial gradients in SST are weak there, so model differences in k_{qsat} will be less important for spatial variation there. Given the form of Equation 1, precipitation is shown on log scales as ratios. Although some models are close to the observations, in others, biases exceed a factor of five in the El Niño/La Niña ratio, seasonal cycles over the Atlantic and West Pacific, and in the Atlantic spatial pattern for the Aug-Oct season (Figure 2a-d). Biases over a factor of two occur in the spatial patterns of the East Pacific annual mean (the long-standing 'double-ITCZ' problem[12,17]) and the Indian Ocean for November-April (Figure 2e-f). Such biases are known, but their causes are not well understood.

These biases (Figure 2a-f) all correspond to excessively weak spatial/temporal variations in precipitation (precipitation ratios too close to 1; including excessively-wide inter-tropical convergence zones[14,18]). This suggests a hypothesis (H_0), that the sensitivity of seasonal precipitation to local SST (k_{qsat}) may be too weak in many models.

To test H_0 objectively, we use a method independent of Figure 2 (Figure 2 was used to propose H_0). This involves estimating k_{qsat} for each model using different data.

Evaluating k_{qsat} in models

We evaluate k_{qsat} using interannual variability in seasonal mean precipitation and SST (the AMIP SST dataset[30] used to drive the model experiments; using years 1980-2005). k_{qsat} is calculated using gridpoint values of seasonal precipitation and q_{sat}, from each location in the study region, and for each year. With these data, Equation 1 becomes a model of the effect of local interannual SST variability on precipitation:

$$\log_e(P(x,t)/P_0(x)) \approx k_{qsat} * (q_{sat}(x,t) - q_{sat,0}(x)) + \langle\text{other processes}\rangle,$$

where (x,t) indicates values for each gridpoint and year; and here, P_0 and $q_{sat,0}$ are the corresponding climatological means for each gridpoint. We estimate k_{qsat} from these data using a modified regression approach (detailed in Methods), minimising the influence of other processes in Equation 1.

To minimise observational error, we exclude the 30% of the tropical oceans with the lowest climatological mean SST (Figure 3b-e, area outside white contour).

Taking logarithms means that all areas of our study region contribute relatively equally to our k_{qsat} estimate (Methods). Consequently, k_{qsat} is relevant over most of the tropical oceans (Figure 3a, correlations are high except for the left bar). k_{qsat} is inapplicable over the coolest, driest ocean regions (Figure 3a, left bar; area masked in Figure 3b-e). The applicable region corresponds to climatological precipitation > ~1 mm/day (Figure 3b-e, orange contour).

k_{qsat} is intended to be independent of large-scale SST spatial patterns. To avoid bias from the large, recurrent ENSO pattern, our 'sortav' regression method first processes the data so all years contribute equally. Linear regression is then applied to obtain k_{qsat}. Rankings of CMIP5 models by k_{qsat} are robust: insensitive to season, to using fewer years of data, or to excluding ENSO years - Extended data Figure 2e-g. Calculated this way, k_{qsat} is less sensitive to the time period used than with simple least squares regression (Extended Data Figures 2h, 7).

We find that the sensitivity of precipitation to local SST variability is much stronger in some models than in others: k_{qsat} varies across CMIP5 models by a factor of 2.5 (0.26-0.66 kg/g; median = 0.46). We group the models into 'high-k_{qsat}' (the 6 models with the largest k_{qsat} values), 'low-k_{qsat}' (the lowest 6 k_{qsat} values) and 'mid-range' subsets.

In Equation 2, setting (q_{sat} - $q_{sat,0}$) to 1, expresses k_{qsat} as the percentage precipitation change per g/kg change in saturation specific humidity (q_{sat}):

$$(P_1 - P_0)/P_0 * 100 \approx 100\,[\exp(k_{qsat} * 1) - 1] = 100\,[\exp(k_{qsat}) - 1]. \qquad \text{Equation 4}$$

Expressed this way, the precipitation sensitivity in CMIP5 models spans 30-93% per g/kg (median = 58%). For context, q_{sat} can vary by a few g/kg, 10° either side of the East Pacific ITCZ during Aug-Oct, and anomalies during ENSO events have a similar magnitude.

High sensitivity of precipitation to SST

We hypothesised above (H_0) that k_{qsat} may be too low in most models. To begin testing this, the results in Figure 2a-f are replotted, but with the 'high-k_{qsat}' subset of models highlighted in magenta (Figure 2g-l). The 'high-k_{qsat}' subset shows much better agreement with TRMM than the full ensemble, in all six panels. Conversely, the 'low-k_{qsat}' subset performs much worse (Extended data Figure 3). Next, we calculate $k_{qsat}^{(spatial)}$ (Methods): as k_{qsat}, but using spatial patterns in climate means, rather than internal variability (Figure 4b). Again, models closest to the observations (Figure 4b, horizontal line) tend to be those with high k_{qsat}. These results all imply that k_{qsat} should be high in the real world (H_0).

These results also show that k_{qsat} is relevant to both spatial and temporal variations in precipitation. We emphasise this by quantifying the overall sensitivity of precipitation to local SST ($k_{qsat}^{spattemp}$), including both spatial and temporal variations (including spatial variation in P_0, Methods). $k_{qsat}^{spattemp}$ is well correlated with k_{qsat} (Extended data Figure 2i). This confirms that k_{qsat} is a useful measure of the underlying sensitivity of precipitation to local SST, relevant to spatial and temporal variations. k_{qsat} does not give information about tropical mean precipitation, which is governed by different processes[31]. k_{qsat} remains our primary measure of precipitation sensitivity to local SST, because it is insensitive to details of SST patterns. In contrast, $k_{qsat}^{spattemp}$ and $k_{qsat}^{(spatial)}$ may be sensitive to the specific spatial patterns in climatological SST (Methods), explaining some of the noise in Figure 4b and Extended data Figure 2i.

We estimate a lower bound for k_{qsat}, using observed interannual variability (independent of Figure 2; Methods). Three values of k_{qsat} are estimated, exactly as for the models, but using TRMM precipitation, and q_{sat} from each of three different SST datasets (HadISST[32] version 1.1, ERSST[33] version 4 and COBE[34] version 2). Uncertainties are estimated, from SST error (including regression dilution bias) and internal variability (the TRMM observational period only partly overlaps the model simulation period). A lower observational bound (95% confidence) of 0.51 kg/g for k_{qsat} is obtained.

For a central observational estimate of k_{qsat} (details in Methods), we return to Figure 2. We ask: if all CMIP5 models had the same value of k_{qsat}, with what value would they best reproduce the observations in Figure 2? We first find where, geographically, the models are most sensitive to k_{qsat}. This reveals seven intervals (shaded in Figure 2g-l). For each interval, model errors relative to TRMM are regressed against modelled k_{qsat} (Extended data Figure 6). For each interval, k_{qsat} is estimated as where the regression line intercepts the x-axis (the value for a theoretical model with zero precipitation error). These seven estimates of k_{qsat} range from 0.56 to 0.68 kg/g (Figure 4a, white dashed lines), all larger than our lower bound estimate. The spread of estimates comes from uncertainty in processes not quantified by k_{qsat}. The range of conditions used, covering spatial, seasonal and temporal variability across different locations, helps to quantify and mitigate this uncertainty. Robustness is tested by plotting results from the sixth model intercomparison project (CMIP6, not used to select the seven intervals) on Extended data Figure 6. The mean of the seven k_{qsat} values (0.6 kg/g; or, using Equation 4, 80% per g/kg) is our central estimate (for 1980-2005; other periods would give slightly different values, from internal SST variability - Extended data Figure 2h).

These results, from two independent methods, suggest that most models underestimate k_{qsat}. Our central estimate (0.6 kg/g, Figure 4a, horizontal black line; Figure 4b, vertical line) is greater than 43 of the 47 model values from CMIP5 and CMIP6. This implies that models underestimating the sensitivity of precipitation to local SST underlies a range of model biases over tropical oceans. CMIP5 and CMIP6 have similar ranges of k_{qsat} values (Figure 4a), highlighting the need for accelerated model development.

Other studies[35] have found biases in a different aspect of the SST-precipitation relationship: model precipitation often tracks SST maxima more closely than in observations. We quantify this in each CMIP5 model as the correlation coefficient between climatological spatial patterns of precipitation and SST, for each season, then average the four seasonal values. This 'spatial-correlation index' is uncorrelated with k_{qsat} ($r = 0.01$; i.e. models with a high spatial-correlation index can have high, low or intermediate k_{qsat}), so it involves different processes.

Processes behind uncertainty in k_{qsat}

To guide model improvements, we explore what causes model differences in k_{qsat}, revealing links to shallow atmospheric circulations. We first note that k_{qsat} involves processes unrelated to tropical mean precipitation: the correlation across CMIP5 models between the two measures is 0.03. Energy budgets constrain tropical mean precipitation[31], while the value of k_{qsat} affects precipitation variation in both time and space.

Beginning with interannual variability, we define $k_{qsat}^{wap}(p)$: the sensitivity of the vertical pressure velocity (wap) to local SST change, at each pressure level (p). This is evaluated like k_{qsat}, using data from all seasons across the tropical oceans, but using wap(p) instead of log(precipitation). A deep mode dominates tropical variability[36], so the CMIP5 mean profile of k_{qsat}^{wap} peaks around 450 hPa (Figure 5a). In contrast, model spread in k_{qsat} is linked to shallow circulations: correlations between $k_{qsat}^{wap}(p)$ and k_{qsat} (Figure 5b) peak near 700hPa ($r = -0.9$; correlations are small at 1000hPa, as wap approaches zero near the surface). That is, in models with high k_{qsat} (as in the real world), shallow circulations respond strongly to SST anomalies. Although deep circulations are important[36] in all models, and shallow circulations have been linked to mean precipitation, especially over the Eastern Pacific[7], our novel finding is that shallow circulations are central to model uncertainty in SST-driven precipitation variability, across the tropics.

Shallow circulations are further linked to model differences in climate means. We study zonal, ocean-only means over 180W-10E (most of the Pacific, entire Atlantic), in Aug-Oct, when meridional SST gradients are strong. CMIP5 results are used to define key regions (black and orange lines in Figure 5c-e), and CMIP6 used to test the conclusions. As expected[4,7], CMIP5 mean vertical velocity profiles are bottom heavy, but extend throughout the troposphere (Figure 5c, colours). Meridional wind (white contours) peaks near the tropopause, with a weaker shallow flow between 500-700hPa[37]. Again, however, inter-model spread in k_{qsat} is associated with shallow circulations: models with high k_{qsat} have stronger shallow descent south of the ITCZ between 600-850hPa (Figure 5d, yellow; Extended Data Figure 8a, magenta line; Extended Data Figure 8d). They also tend to have stronger shallow ascent in the ITCZ (Figure 5d, blue region; Extended Data Figure 8c, magenta line), stronger trade winds and stronger return flow between 500-700hPa (Figure 5d, white contours; Extended Data Figure 8b, magenta line). This shallow circulation is weak in the low-k_{qsat} mean (Extended Data Figure 8a-c, blue lines).

The weak link between model differences in k_{qsat} and deep circulations may arise partly from physical constraints. In descending air, differences in vertical velocity are largest below about 600 hPa (Extended Data Figure 8a,i). This is partly explained by the vertically-integrated dry static energy (DSE) budget. This budget constrains vertical velocities, requiring balance between radiative, sensible and latent heating, and advection of DSE (Methods). In descending air above 600 hPa, there are fewer uncertain processes affecting this budget, with negligible energy input from cloud and precipitation. Here, therefore, downward advection of DSE is balanced mostly by dry clear-sky radiative cooling (Extended Data Table 1). Further, vertical temperature profiles are similar across CMIP5 models. This constrains both radiative cooling and vertical gradients of DSE, limiting model differences in vertical velocity. Model temperatures are constrained, near the surface by SST, and near the tropopause as modelling groups aim to reproduce observed outgoing longwave radiation. Below 600hPa, there are additional sources of uncertainty, from cloud and precipitation[38], leading to larger model differences in shallow descent, and so stronger links to differences in k_{qsat}.

The depth of the meridional return flow (500-700hPa, Figure 5e, white contours) suggests a circulation driven by precipitating shallow convection[37,39]. The alternative, sea-breeze mechanism reaches lower levels[37]. Both circulation types may exist over the Galapagos (Extended Data Figure 8g): here, model meridional winds between 600-700hPa (shallow-convection-type) are uncorrelated with those between 700-850hPa (sea-breeze-type), indicating different physical processes at these two levels. The column dry static energy budget (Methods) also implicates precipitating shallow convection. Model differences in this budget are predominantly a balance between vertical advection integrated over 600-1000 hPa, and precipitation latent heating (Extended Data Table 1, final column): in descending air below 600 hPa, high-k_{qsat} models have stronger shallow advective warming from stronger shallow descent, balancing weaker warming from weaker precipitation. As expected, the ensemble mean in descending air is mostly a balance between radiative cooling and advective warming.

Causality is hard to fully establish, but the most likely explanation of our results is that model differences in k_{qsat} mostly originate from model differences in the behaviour of shallow precipitating convection. Such differences would affect the sensitivity of precipitation to SST directly. The consequent differences in shallow latent heating also appear to lead to differences in shallow circulations. This couples shallow vertical velocities in descending and ascending air (Extended Data Figure 8e), further modifying the sensitivity of precipitation to SST.

Significant model differences in the physical representation of convection[40,41], including its coupling to circulation[5,42], are well established. Improvements in k_{qsat} in the CNRM model from CMIP5 (0.43 kg/g) to CMIP6 (0.54 kg/g) are largely associated with convection scheme changes. Running CNRM-CM6 with the CNRM-CM5 convection scheme gives k_{qsat}=0.36 kg/g, even smaller than that of CNRM-CM5. That is, the effect of changing the convection scheme is partly offset by changes in other schemes. Changes from CM5 to CM6 include the shallow convection scheme and the transition from shallow to deep convection[43]. These, and other physics schemes, including boundary layer, deep or mid-level convection, or microphysics, could all affect how shallow precipitating convection responds to SST.

Other processes may have smaller contributions. Dry shallow circulations[44] are more important over hot, dry land. A limited role for differences in radiation (Extended Data Table 1, final column), suggests cloud parameterizations are not dominant. Differences in dynamical schemes are thought to be less likely to be important for tropical precipitation biases, although coupling of dynamics to physics is important[45–47].

Models with stronger shallow circulations can also import more moist static energy in ascending air, driving enhanced deep convection[48]. Our results support this (Extended Data Figure 8h), showing that model differences in shallow ascent are strongly positively correlated with differences in deep ascent. This is an indirect link to k_{qsat}, which is more weakly associated with deep ascent rates (Figure 5d).

Strong real-world shallow circulations

Shallow circulations are challenging to observe[11], but our results suggest they are stronger in the real world than in most models. Models with high k_{qsat} (as in the real world) tend to have strong shallow circulations (in both climate means and internal variability). We test this further with two independent observations, in Aug-Oct. In models, there is a strong (r=0.86) relationship between shallow descent, and northward trade winds (Figure 4c). If the shallow descent is strong in the real world, we would expect strong northward trades – and this is confirmed by QuikSCAT satellite observations[49] (SeaWinds scatterometer, Level 3 product, years 1999-2009 - horizontal line in Figure 4c). There is, similarly, a strong (r=0.81) relationship (Figure 4d) between 500-700hPa meridional winds over our study region, and the mean 600-700hPa meridional wind over Galapagos and Christmas Islands (915 MHZ wind profiler[8,50] observation sites, years 1994-2005 at Galapagos, 1990-2002 at Christmas Island; few observations reach above 600hPa). If the large-scale 500-700hPa wind is strong in the real world, we would expect strong southward winds in the observations, and this is what the wind profilers show (horizontal line in Figure 4d). These results, using multiple observations, confirm a previous suggestion based on a single reanalysis product[5].

Conclusions

Our results show that k_{qsat} is linked to a range of model biases in precipitation and atmospheric circulation. Improving k_{qsat} should reduce those biases, giving greater confidence in seasonal through multi-decadal model projections. k_{qsat} affects the strength of precipitation variation in both time and space. Other biases, in large-scale energy budgets, and teleconnections, also affect precipitation. We show that k_{qsat} can be constrained by observations, and give evidence that improving the representation of shallow tropical precipitating convection, and its coupling to SST and circulation, could improve k_{qsat}. This identifies specific model development goals and gives new ways of linking these to observable physical processes.

References

1. Philander, S. G. H. El-Nino Southern Oscillation Phenomena. *Nature* **302**, 295–301 (1983).
2. Pierrehumbert, R. T. Thermostats, Radiator Fins, and the Local Runaway Greenhouse. *J. Atmos. Sci.* **52**, 1784–1806 (1995).
3. Tian, B. J. Spread of model climate sensitivity linked to double-Intertropical Convergence Zone bias. *Geophys. Res. Lett.* **42**, 4133–4141 (2015).
4. Back, L. E. & Bretherton, C. S. Geographic variability in the export of moist static energy and vertical motion profiles in the tropical Pacific. *Geophys. Res. Lett.* **33**, (2006).
5. Sherwood, S. C., Bony, S. & Dufresne, J. L. Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature* **505**, 37-+ (2014).
6. Trenberth, K. E., Stepaniak, D. P. & Caron, J. M. The global monsoon as seen through the divergent atmospheric circulation. *J. Clim.* **13**, 3969–3993 (2000).
7. Back, L. E. & Bretherton, C. S. A Simple Model of Climatological Rainfall and Vertical Motion Patterns over the Tropical Oceans. *J. Clim.* **22**, 6477–6497 (2009).
8. Zhang, C. D., Nolan, D. S., Thorncroft, C. D. & Nguyen, H. Shallow meridional circulations in the tropical atmosphere. *J. Clim.* **21**, 3453–3470 (2008).
9. Li, R. & Fu, Y. F. Tropical precipitation estimated by GPCP and TRMM PR observations. *Adv. Atmos. Sci.* **22**, 852–864 (2005).
10. Prakash, S., Mahesh, C. & Gairola, R. M. Comparison of TRMM Multi-satellite Precipitation Analysis (TMPA)-3B43 version 6 and 7 products with rain gauge data from ocean buoys. *Remote Sens. Lett.* **4**, 677–685 (2013).
11. Bellon, G., Reitebuch, O. & Naumann, A. K. Shallow Circulations: Relevance and Strategies for Satellite Observation. *Surv. Geophys.* **38**, 1509–1528 (2017).
12. Li, G. & Xie, S. P. Tropical Biases in CMIP5 Multimodel Ensemble: The Excessive Equatorial Pacific Cold Tongue and Double ITCZ Problems. *J. Clim.* **27**, 1765–1780 (2014).
13. Xie, S. P. *et al.* Towards predictive understanding of regional climate change. *Nat. Clim. Chang.* **5**, 921–930 (2015).
14. Byrne, M. P. & Schneider, T. Energetic Constraints on the Width of the Intertropical Convergence Zone. *J. Clim.* **29**, 4709–4721 (2016).
15. Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M. & Vialard, J. ENSO representation in climate models: from CMIP3 to CMIP5. *Clim. Dyn.* **42**, 1999–2018 (2014).
16. Zhang, X. X., Liu, H. L. & Zhang, M. H. Double ITCZ in Coupled Ocean-Atmosphere Models: From CMIP3 to CMIP5. *Geophys. Res. Lett.* **42**, 8651–8659 (2015).
17. Hirota, N. & Takayabu, Y. N. Reproducibility of precipitation distribution over the tropical oceans in CMIP5 multi-climate models compared to CMIP3. *Clim. Dyn.* **41**, 2909–2920 (2013).
18. Dixit, V., Geoffroy, O. & Sherwood, S. C. Control of ITCZ Width by Low-Level Radiative Heating From Upper-Level Clouds in Aquaplanet Simulations. *Geophys. Res. Lett.* **45**, 5788-5797. https://doi.org/10.1029/2018GL078292 (2018).

19. He, J. *et al.* Precipitation Sensitivity to Local Variations in Tropical Sea Surface Temperature. *J. Clim.* **31**, 9225–9238 (2018).

20. Rushley, S. S., Kim, D., Bretherton, C. S. & Ahn, M. S. Reexamining the Nonlinear Moisture-Precipitation Relationship Over the Tropical Oceans. *Geophys. Res. Lett.* **45**, 1133–1140 (2018).

21. Bretherton, C. S., Peters, M. E. & Back, L. E. Relationships between water vapor path and precipitation over the tropical oceans. *J. Clim.* **17**, 1517–1528 (2004).

22. Kanemaru, K. & Masunaga, H. A Satellite Study of the Relationship between Sea Surface Temperature and Column Water Vapor over Tropical and Subtropical Oceans. *J. Clim.* **26**, 4204–4218 (2013).

23. TRMM. Tropical Rainfall Measuring Mission, TRMM (TMPA/3B43) Rainfall Estimate L3 1 month 0.25 degree x 0.25 degree V7, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed 2017. 10.5067/TRMM/TMPA/MONTH/7 (2011).

24. Huffman, G. J. *et al.* The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **8**, 38–55 (2007).

25. Adler, R. F. *et al.* The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation. *Atmosphere (Basel).* **9**, (2018).

26. McPhaden, M. J. *et al.* The tropical ocean global atmosphere observing system: A decade of progress. *J. Geophys. Res.* **103**, 14169–14240 (1998).

27. McPhaden, M. J. *et al.* RAMA The Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction. *Bull. Am. Meteorol. Soc.* **90**, 459-+ (2009).

28. Bourles, B. *et al.* The PIRATA program: History, accomplishments, and future directions. *Bull. Am. Meteorol. Soc.* **89**, 1111-+ (2008).

29. Serra, Y. L., A'Hearn, P., Freitag, H. P. & McPhaden, M. J. ATLAS self-siphoning rain gauge error estimates. *J. Atmos. Ocean. Technol.* **18**, 1989–2002 (2001).

30. Taylor, K. E., Williamson, D. & Zwiers, F. W. The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations. *PCMDI Rep. No. 60, Progr. Clim. Model Diagnosis Intercomparison, Lawrence Livermore Natl. Lab. Livermore, California, 25 pp* (2000).

31. Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224-+ (2002).

32. Rayner, N. A. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, (2003).

33. Huang, B. Y. *et al.* Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons. *J. Clim.* **28**, 911–930 (2015).

34. Hirahara, S., Ishii, M. & Fukuda, Y. Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty. *J. Clim.* **27**, 57–75 (2014).

35. Biasutti, M., Sobel, A. H. & Kushnir, Y. AGCM precipitation biases in the tropical Atlantic. *J.*

Clim. **19**, 935–958 (2006).

36. Holloway, C. E. & Neelin, J. D. The convective cold top and quasi equilibrium. *J. Atmos. Sci.* (2007) doi:10.1175/JAS3907.1.

37. Nolan, D. S., Powell, S. W., Zhang, C. D. & Mapes, B. E. Idealized Simulations of the Intertropical Convergence Zone and Its Multilevel Flows. *J. Atmos. Sci.* **67**, 4028–4053 (2010).

38. Knutson, T. R. & Manabe, S. Time-mean response over the tropical Pacific to increased CO2 in a coupled ocean-atmosphere model. *J. Clim.* (1995) doi:10.1175/1520-0442(1995)008<2181:TMROTT>2.0.CO;2.

39. Wu, Z. H. A shallow CISK, deep equilibrium mechanism for the interaction between large-scale convection and large-scale circulations in the tropics. *J. Atmos. Sci.* **60**, 377–392 (2003).

40. Lin, Y. *et al.* TWP-ICE global atmospheric model intercomparison: Convection responsiveness and resolution impact. *J. Geophys. Res. Atmos.* (2012) doi:10.1029/2011JD017018.

41. Randall, D. A. Beyond deadlock. *Geophys. Res. Lett.* (2013) doi:10.1002/2013GL057998.

42. Bony, S. & Dufresne, J. L. Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.* (2005) doi:10.1029/2005GL023851.

43. Voldoire, A. *et al.* Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *J. Adv. Model. Earth Syst.* (2019) doi:10.1029/2019MS001683.

44. Nie, J., Boos, W. R. & Kuang, Z. Observational evaluation of a convective quasi-equilibrium view of monsoons. *J. Clim.* (2010) doi:10.1175/2010JCLI3505.1.

45. Stevens, B. & Bony, S. What are climate models missing? *Science* (2013) doi:10.1126/science.1237554.

46. Oueslati, B. & Bellon, G. Convective Entrainment and Large-Scale Organization of Tropical Precipitation: Sensitivity of the CNRM-CM5 Hierarchy of Models. *J. Clim.* **26**, 2931–2946 (2013).

47. Vial, J., Bony, S., Stevens, B. & Vogel, R. Mechanisms and Model Diversity of Trade-Wind Shallow Cumulus Cloud Feedbacks: A Review. *Surveys in Geophysics* (2017) doi:10.1007/s10712-017-9418-2.

48. Fläschner, D., Mauritsen, T., Stevens, B. & Bony, S. The signature of shallow circulations, not cloud radiative effects, in the spatial distribution of tropical precipitation. *J. Clim.* (2018) doi:10.1175/JCLI-D-18-0230.1.

49. NASA. *SeaWinds on QuikSCAT Level 3 Surface Northward Wind for Climate Model Co mparison. Ver. 1. PO.DAAC, CA, USA. Dataset accessed [2019-09-06] at https://doi.org/10.5067/QS SNW-CMIP1.* (2012).

50. Gage, K. S., Williams, C. R. & Ecklund, W. L. UHF wind profilers: a new tool for diagnosing tropical convective cloud systems. *Bull. - Am. Meteorol. Soc.* (1994) doi:10.1175/1520-0477(1994)075<2289:UWPANT>2.0.CO;2.

Figures

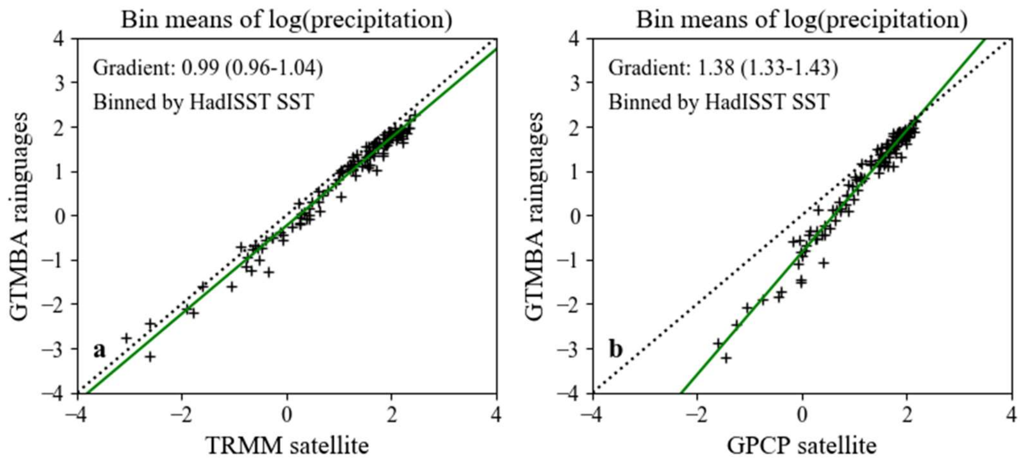


Figure 1.

Validating observations of log precipitation from satellites. GTMBA in-situ raingauge observations versus satellite observations from **a** TRMM and **b** GPCP. Each symbol represents the mean of all seasonal mean data within a given SST bin (Methods). Green line: best fit line (gradient and its 95% confidence interval quoted in each figure); dotted line: 1:1 line.

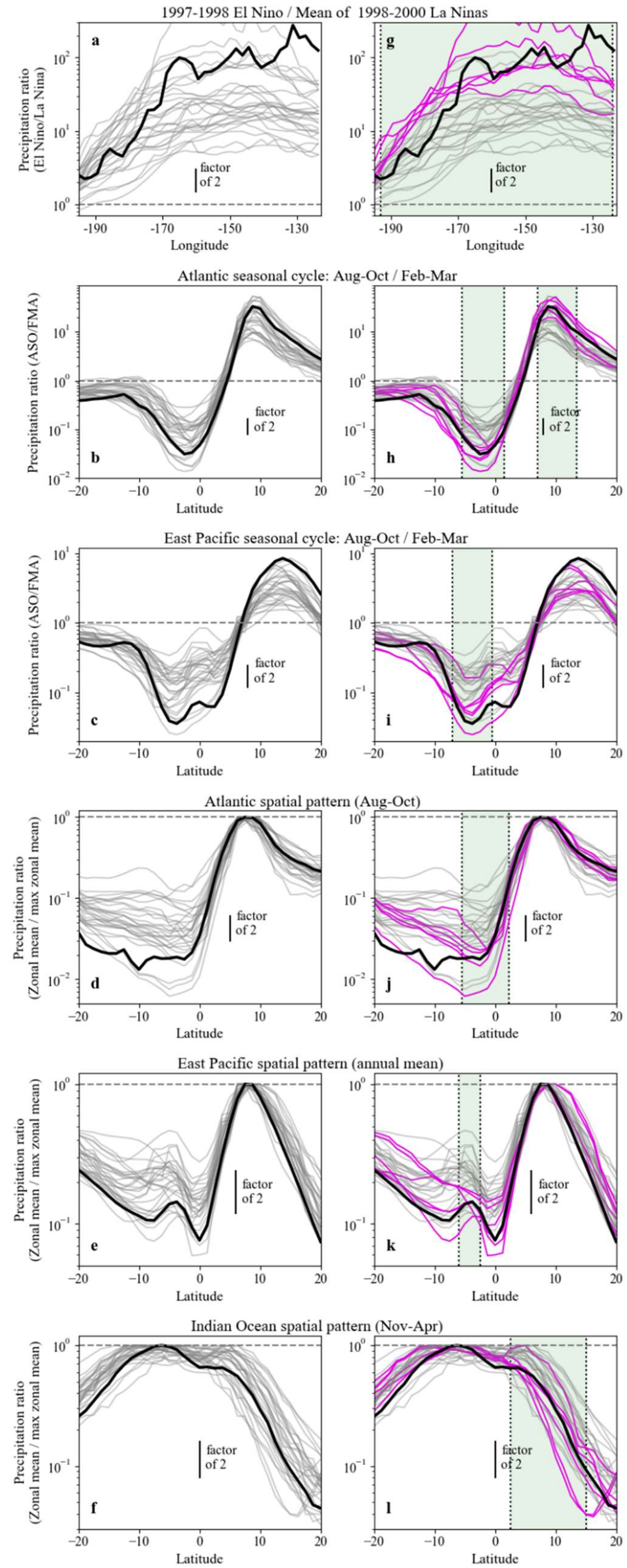


Figure 2. Model precipitation biases. (black) TRMM observations. Horizontal dashed line marks precipitation ratio=1. **a-f** all CMIP5 models are shown in grey lines; **g-i** magenta:

'high-k_{qsat}' subset; grey: other models. Spatial patterns (bottom 3 rows) given by scaling zonal mean precipitation by its latitudinal maximum. Green shading marks the intervals used for the 7 estimates of k_{qsat}. These examples were chosen as they feature large differences/gradients in SST. Precipitation ratios are plotted because of the form of Equation 1.

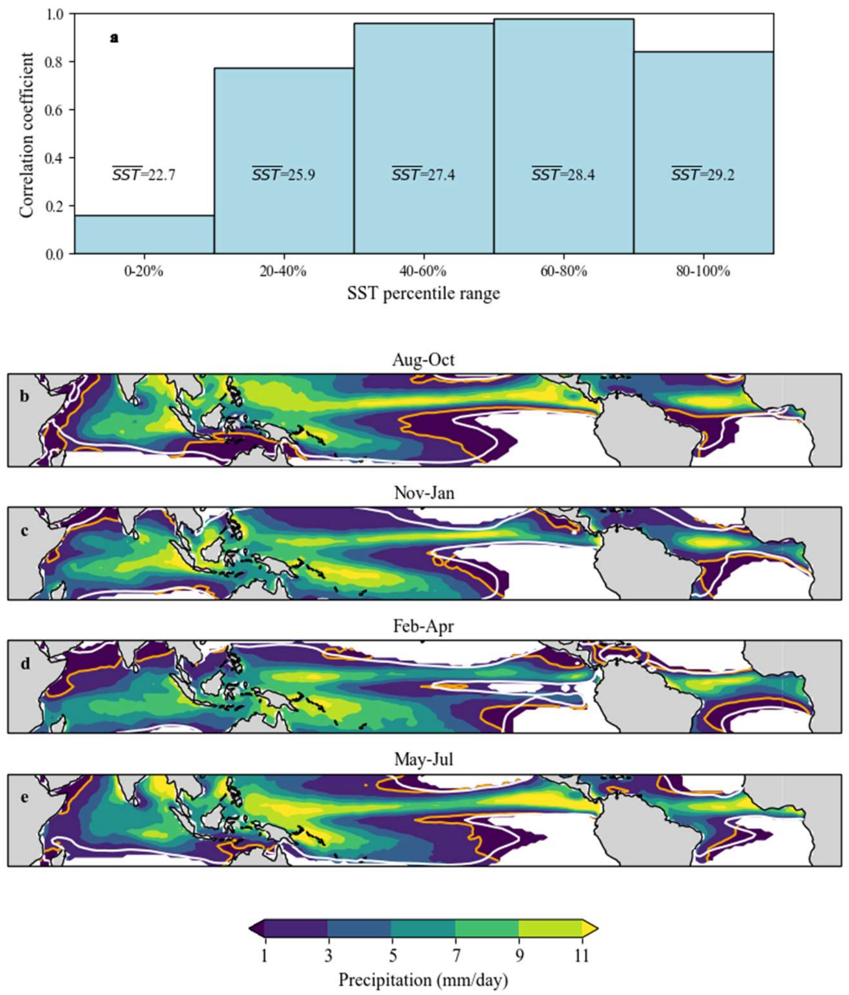


Figure 3. The region of applicability of kqsat. (**a**) each bar represents a climatological zone covering 20% of the tropical oceans, defined by the seasonal climatological SSTs (e.g. the left bar is the zone with the coolest 20% of SSTs – white masked ocean in the maps below). Climatological zones are defined separately for each season. Bar height: the correlation coefficient, across CMIP5 models, between the standard calculation of kqsat, and that calculated only over the selected climatological zone. Mean SST (°C) for each zone is also shown. (**b-e**) Colours: mean TRMM precipitation; orange line highlights 1 mm/day contour. Data is masked over the 20% of the oceans where kqsat is inapplicable (left-hand bar in panel **a** shows low correlation). White contour shows the 30th percentile of SST: the standard calculation of kqsat uses data inside this contour.

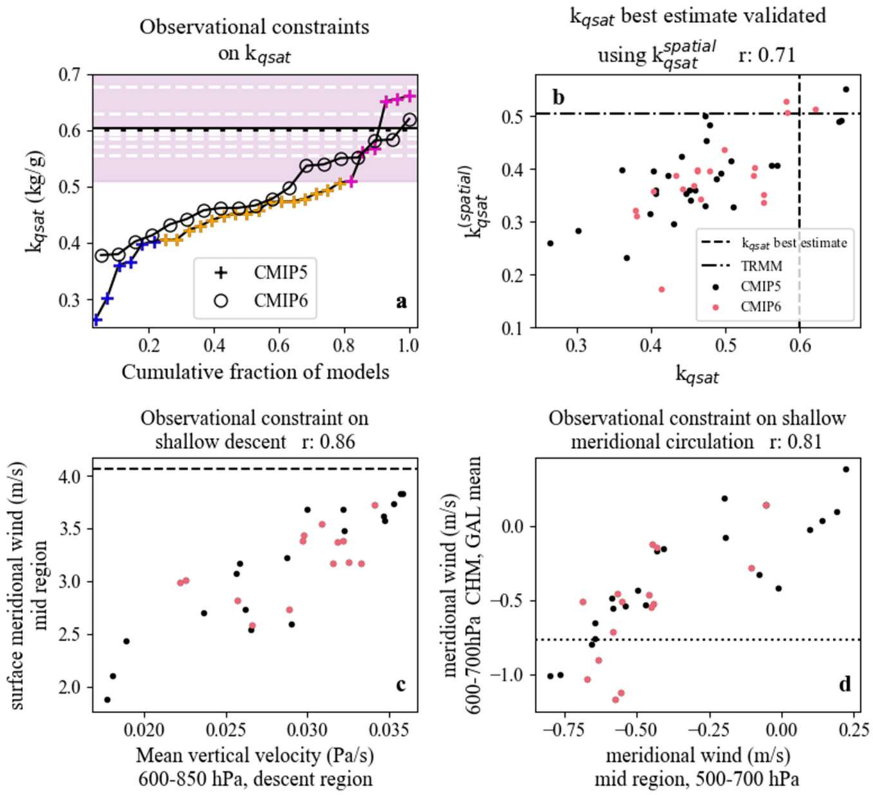


Figure 4. High sensitivity of precipitation to SST, and strong shallow circulations, in the real world. **a** horizontal lines mark (white dashes) the 7 estimates of k_{qsat} (Extended data Figure 6), and the central estimate (black solid); shading marks k_{qsat} values above our lower-bound estimate; symbols mark sorted model k_{qsat} values for (crosses) CMIP5 (blue and magenta denote low-k_{qsat} and high-k_{qsat} model subsets) and (circles) CMIP6. **b** $k_{qsat}^{(spatial)}$ versus k_{qsat}, for each (black) CMIP5 and (red) CMIP6 model; horizontal line: $k_{qsat}^{(spatial)}$ from TRMM observations; vertical line: best estimate of k_{qsat}. **c,d** each symbol represents one CMIP5 (black) or CMIP6 (red) model; title gives Pearson correlation coefficient. **c** surface meridional wind averaged over the mid-region (180W-10E, 1-7N) versus shallow descent index (defined in Figure 5); horizontal line marks QuikSCAT observation. **d** meridional wind averaged over Galapagos & Christmas island, 600-700hPa (few observations above 600hPa) versus meridional wind averaged over the mid region (180W-10E, 1-7N), 500-700hPa; horizontal line marks wind profiler observation.

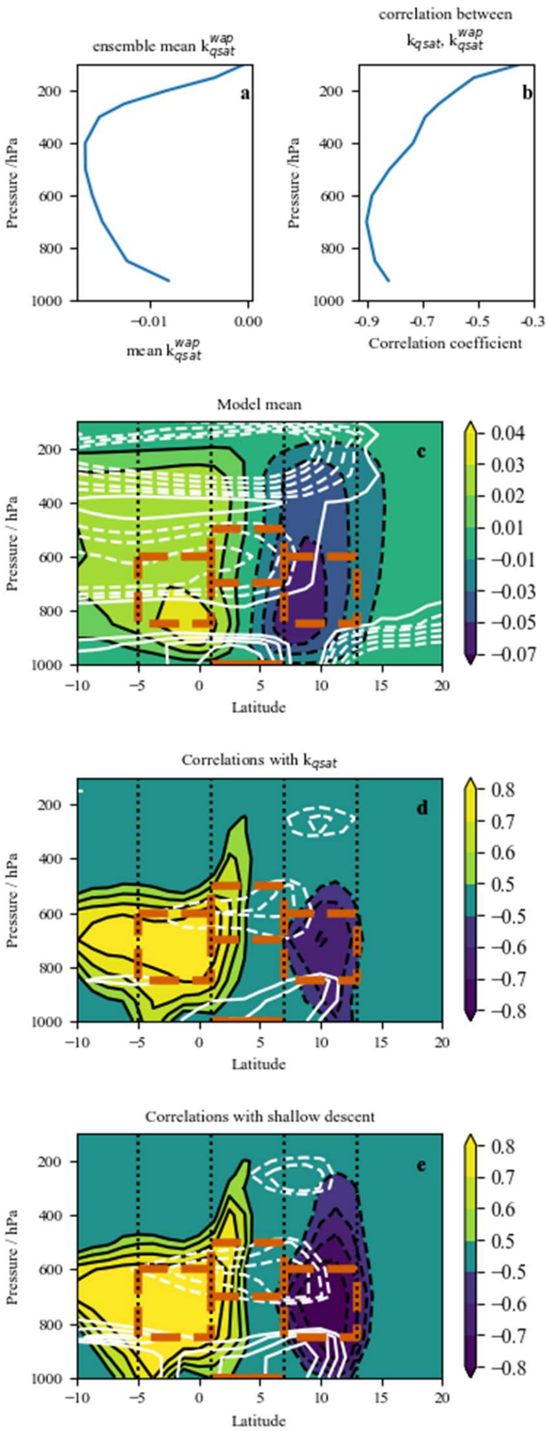


Figure 5. Linking k_{qsat} to shallow circulations. **a,b** quantifies internal variability, and **c-e** climate means. **a** CMIP5 ensemble mean of k_{qsat}^{wap} (Pa kg/g), at each pressure level. **b** inter-model correlations (Pearson r) between k_{qsat}, and k_{qsat}^{wap}, at each pressure level. Correlations are negative because of the definition of wap. **c-e** Aug-Oct, 180W-10E zonal means. **c** CMIP5 ensemble mean of (colours) vertical velocity (Pa/s) and (white contours) meridional

wind. **d** inter-model correlations between k_{qsat} and mean vertical velocity (colours) and between k_{qsat} and mean meridional wind (white contours). **e** as **d**, but for correlations with the shallow descent index instead of k_{qsat} (shallow descent index = vertical velocity averaged over left-hand orange-dashed box: 5S-1N, 850-600 hPa).

Methods

Data domain, SST, precipitation and q_{sat}

Results are based on seasonal means of precipitation and SST, over the tropical oceans (20S-20N). q_{sat} is calculated as the saturation specific humidity at the seasonal mean sea-surface temperature and 1000 hPa air pressure. Use of seasonal mean SST here means that q_{sat} will be lower than the seasonal mean of saturation specific humidity calculated from daily SST (due to sub-seasonal SST variability, and nonlinearity in the humidity calculation). We use seasonal mean SST to minimise observational error: sub-seasonal SST variability is hard to observe accurately.

Observations

Satellite precipitation data are seasonal averages of monthly means from V7 of the 3B43 Tropical Rainfall Measuring Mission (TRMM)[23,24] dataset, covering 1998-2015. Data from V2.3 of the Global Precipitation Climatology Project (GPCP[25]) retrieval are included in Figure 1 only.

In-situ raingauge data from 89 buoys of the Global Tropical Moored Buoy Array (all buoys with more than 1 year of precipitation data), from the Tropical Ocean-Global Atmosphere (TOGA[26]) observing system, the Prediction and Research Moored Array in the Atlantic (PIRATA[28]) and the Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction (RAMA[27]), were retrieved as daily means. Days with lower quality data (quality codes not equal to 1 or 2) were rejected. Monthly means were then calculated only for months with 20 or more days with code 1 or 2 data (other months are marked as missing). Seasonal means were taken only where three consecutive months had non-missing data.

Monthly mean SSTs are taken from four different datasets. The CMIP5 AMIP dataset[30] (the dataset used to drive the AMIP SST-forced model runs) is available only for 1980-2005. This was used to calculate k_{qsat} for each model. For our observational estimate of a lower bound on k_{qsat} (using TRMM precipitation), three SST datasets were used: HadISST[32] version 1.1, ERSST[33] version 4 and COBE[34] version 2 (the AMIP dataset used by the model simulations was not used, due to its limited temporal overlap with the TRMM operational period).

Surface meridional wind observations are from SeaWinds on QuikSCAT Level 3, for the period Aug 1999-Oct 2009. Wind profiler observations at San Cristóbal, Galápagos (0.9°S, 89.7°W, 1994-2005) and Christmas Island (2.8°N, 157.5°W, 1990-2002) used 915 MHz in low mode, as used in other studies of shallow circulation[8].

Model data

All model results are from atmosphere-only AMIP runs (one run per model version) forced by observed SST, corresponding to the period 1980-2005. This includes 28 models from CMIP5 (ACCESS1-0, ACCESS1-3, BNU-ESM, CCSM4, CESM1-CAM5, CNRM-CM5, CSIRO-Mk3-6-0, CanAM4, GISS-E2-R, HadGEM2-A, IPSL-CM5A-LR, IPSL-CM5B-LR, MIROC-ESM, MIROC5, MRI-AGCM3-2H, MRI-AGCM3-2S, MRI-CGCM3, NorESM1-M, inmcm4, bcc-csm1-1-m, bcc-csm1-1, EC-EARTH, MPI-ESM-LR, MPI-ESM-MR, FGOALS-s2, FGOALS-g2, GFDL-CM3, GFDL-HIRAM-C180) and 19 from CMIP6 (BCC-CSM2-MR, BCC-ESM1, CAMS-CSM1-0, CanESM5, CNRM-CM6-1, CNRM-ESM2-1, EC-Earth3-Veg, INM-CM4-8, INM-CM5-0, IPSL-CM6A-LR, MIROC6, HadGEM3-GC31-

LL, UKESM1-0-LL, MRI-ESM2-0, GISS-E2-1-G, GFDL-AM4, GFDL-CM4, NESM3, SAM0-UNICON).

Evaluation of satellite precipitation using GTMBA raingauge data

First, seasonal mean satellite precipitation data (for all seasons) were interpolated linearly to the GTMBA locations. Logarithms of seasonal mean precipitation were then taken, and all datasets masked at times and locations where any data (GTMBA or satellite) were missing. This resulted in 1723 observations of seasonal mean precipitation from each dataset, covering the period 1998-2015 (the overlap between TRMM and GTMBA operational periods). For the remainder of the analysis, the observation location and time are ignored.

To reduce noise effectively, while retaining the signal of interest, we use the fact that precipitation tends to increase with SST, but the noise (as defined here) is largely independent of SST. For each dataset, the 1723 observations were grouped into 120 bins (14 observations per bin). This was done by ranking the observations by seasonal mean SST (the 14 observations corresponding to the 14 lowest SST values were placed in the first bin, and so on). The mean across each bin was then taken, giving 120 bin means of log(seasonal precipitation): giving 120 symbols in Figure 1. 120 bins were chosen, as a mean over 14 observations is sufficient to reduce noise significantly, while retaining a large number of symbols in Figure 1 to assess the method visually. Doubling the number of bins has negligible effect on the gradient in Figure 1. Results are insensitive to which SST dataset is used to bin the data (compare Extended data Figure 4, bottom two rows, with Figure 1).

We demonstrate that regression dilution bias is likely to be small in Figure 1a (method justification in Methods subsection 'Estimating regression dilution bias' below). We do this by repeating the analysis in Figure 1a, but regressing first TRMM against GPCP (Extended data Figure 4, top left), then GPCP against TRMM (Extended data Figure 4, top right). The product of the two regression gradients is 0.98 (close to 1), suggesting that this bias is small.

We also tested sensitivity of the validation to potential undercatch by the GTMBA raingauges in windy conditions[29]. This issue could only bias the gradients in Figures 1a,b if the percentage undercatch varied systematically from low to high precipitation (because Figure 1 shows log precipitation). To test this, we recalculated the gradients in Figures 1a,b, but after masking the data according to the seasonal mean wind speed (also observed by GTMBA buoys). Gradients calculated for low wind (0-4 m/s; 22% of all data) and high wind (6-10 m/s; 27% of all data) show no significant differences from Figure 1a,b (for TRMM, confidence intervals consistently spanning 1 and best estimate within 5% of 1; for GPCP, confidence intervals consistently excluding 1). This suggests that the satellite validation is insensitive to wind undercatch.

Figure 2 data preparation (regions, seasons, time periods)

For Figure 2, all data is first regridded by area-averaging to a common grid (resolution: 1.25º latitude by 1.875º longitude).

The El Niño/La Niña ratio is based on large ENSO episodes in years both simulated by the models and observed by TRMM: the 1997-1998 El Niño divided by the mean of 1998-1999 and 1999-2000 La Niñas, for the El Niño peak season (November-January), averaged over 10S-10N.

Other data are zonal means over the following longitude bands (with land masked out): (Atlantic) 70W-25E; (East Pacific) 150-100W; (Indian Ocean) 50-100E.

Seasonal cycles over Atlantic and Pacific are calculated as the zonal mean for August-October divided by the zonal mean for February-April). These seasons were chosen because they show large differences in SST, but small differences in solar zenith angle (the latter can affect precipitation by altering land temperature).

Spatial patterns are calculated, for each model/observational dataset, as the zonal mean at each latitude, divided by the maximum, zonal mean for the same model/observational dataset. This was calculated for August-October (ASO) for the Atlantic, due to the large meridional SST gradient for this season. November-April was used for the Indian Ocean, as this basin has a significant meridional SST gradient for this period.

Estimating k_{qsat}, part 1: data preparation

k_{qsat} is used here specifically to rank the models and compare with observations. Therefore, the method of calculation needs to be consistent across models and observations, and to minimise the potential for observational error.

For each year, for a given season, the logarithm of seasonal mean precipitation is calculated. The spatial pattern of climatological mean precipitation (Extended data Figure 2j) is dominated by a small area of large precipitation (occupying around 10% of the area). Hence, if we evaluated k_{qsat} without taking the logarithm, our result would be dominated by this small area of the tropics. The spatial distribution of log(precipitation) is much more uniform (Extended data Figure 2k), except for the driest 10% of the tropics (which is eliminated from our analysis as we mask the region of coolest SST). Hence, using log(precipitation) to calculate k_{qsat} ensures that the result is influenced fairly equally by all parts of our analysis region (confirmed in Figure 3a). Our results use seasonal mean precipitation. Use of other timescales would alter k_{qsat}, due to the (nonlinear) logarithm in Equation 1.

For models, k_{qsat} is calculated using data on each model's native grid. For observations, q_{sat} is regridded linearly to the high resolution TRMM horizontal grid.

Before estimating k_{qsat}, to minimise observational error, we exclude the 30% of the tropical oceans with the lowest climatological mean SST. An advantage of using a logarithm in equation 1 is that k_{qsat} estimates are not dominated by the narrow ITCZ region. However, it could mean that error in observing the very lowest rainfall rates could cause large error in our real-world estimate of k_{qsat}. Therefore we mask the regions with coolest SST on average. This masking is only done in calculating k_{qsat}. It is not done in Figure 2, as the climatological means reduce observational error there.

At each location, anomalies relative to climatological means are calculated for each year, for both q_{sat} and log(precipitation). Locations that have missing data in any year are excluded. These data are used in the sortav method, described below.

Estimating k_{qsat}, part 2: sortav regression method

Once the data is prepared as above, our regression method for estimating k_{qsat} (denoted 'sortav') is applied. For python code for this method, and an illustrative example, see Code Availability Statement. The sortav method is designed to prevent dominance from the SST spatial pattern associated with ENSO (an issue because ENSO features large SST anomalies in a consistent pattern). In estimating k_{qsat} from inter-annual variability, our aim is to reduce the 'other-processes' term in Equation 1, by averaging over different SST patterns, with different patterns of large-scale circulation anomalies. If a single SST pattern (ENSO) was allowed to dominate, this would not be effective. If standard linear regression was used, the ENSO pattern would dominate, because ENSO features large SST anomalies. In addition, tropical means of precipitation and SST can vary over time, involving different processes than those represented by k_{qsat}. Our method avoids these issues.

The first step sorts each year of data. Say the data has n locations and y years. For each year, the n anomalies in log(precipitation) are sorted in order of increasing q_{sat} anomaly. This produces, for each year, a vector of length n, with the first element corresponding to the location with the most negative q_{sat} anomaly, and the last element being that with the most positive q_{sat} anomaly. If Equation 1 was exactly true, with zero noise, each vector would be sorted in order of increasing anomaly in log(precipitation). Because of the noise (from large-scale processes), this is not, in general, true. This gives y sorted vectors, each of length n.

We then average over the y years, to produce one mean vector of length n (e.g. the 1st element of this vector is the mean of the y precipitation anomalies found over the most negative q_{sat} anomaly from each year). This averaging removes much of the noise, because the noise is largely independent of q_{sat}. In this mean, years with large SST anomalies have the same weighting as other years, avoiding dominance by ENSO.

The same process is repeated for q_{sat}. This gives two mean vectors, each of length n: for anomalies in log(precipitation) and in q_{sat}.

The relationship between the averaged anomalies in log(precipitation) and q_{sat} is relatively linear in both models and observations (e.g. Extended data Figure 2a-d), suggesting that Equation 1 is a useful approximation in this context. k_{qsat} is then estimated from these vectors, by ordinary least squares linear regression (e.g. the gradients of the best fit lines in Extended data Figure 2a-d).

To compare with our sortav method, alternative estimates of k_{qsat} (marked OLS in Extended Data Figures 2h,7) use standard linear regression between seasonal anomalies in log(precipitation) and q_{sat} (without sort-averaging).

Calculating $k_{qsat}^{(spatial)}$ and $k_{qsat}^{spattemp}$

$k_{qsat}^{(spatial)}$ is calculated using the same method as k_{qsat}, except that it quantifies seasonal and spatial variation in time-mean climate (in contrast with k_{qsat}, which quantifies interannual variability). First, time means are taken for each dataset and season (giving 4 season means per grid point per dataset). For each dataset and each season separately, precipitation is divided by the tropical mean, before taking logarithms. This is done to scale out model variation in the tropical mean (which is controlled by the large-scale energy budget). For q_{sat}, for each dataset and season separately, anomalies are taken with respect to the tropical mean. Masking, to exclude regions with low q_{sat}, is based on annual mean q_{sat}.

$k_{qsat}^{spattemp}$ is defined as follows:

$$P/\bar{P} \approx \exp\left(k_{qsat}^{spattemp} \cdot {q_{sat}}'\right)$$

This approximates variation in precipitation relative to the tropical mean, driven by variation in q_{sat} relative to its tropical mean. The overbar represents the tropical mean for the current season of the current year, and ${q_{sat}}'$ is specifically the q_{sat} anomaly with respect to the tropical mean:

$${q_{sat}}' = q_{sat} - \overline{q_{sat}}$$

Anomalies expressed this way capture temporal and spatial variability associated with variation in local SSTs, but exclude temporal variability in tropical mean precipitation (which is constrained by the large-scale atmospheric energy budget).

$k_{qsat}^{spattemp}$ is estimated using the same approach as kqsat, but anomalies of log(P) and qsat are taken relative to their tropical means for the corresponding season and year (kqsat is evaluated using anomalies with respect to local climatological means for each location).

A disadvantage of both $k_{qsat}^{(spatial)}$ and $k_{qsat}^{spattemp}$ is that they have some sensitivity to other processes (teleconnections) associated with the specific spatial patterns in climatological mean SST (these spatial patterns are filtered out less effectively as there are only 4 seasons, compared to the 25 years of internal variability used to estimate kqsat).

A lower bound for k_{qsat}, using observed interannual variability

This method (estimated lower bound on k_{qsat}) accounts for three forms of observational error, combined using Monte Carlo sampling.

First, systematic error in the observed magnitudes of seasonal mean SST anomalies could bias k_{qsat}. To explore this, we first estimated k_{qsat} (with the sortav method) using TRMM precipitation, and each of the three SST datasets (HadISST, ERSST, COBE) that cover the whole TRMM operational period, giving three direct, unscaled estimates of k_{qsat} (Extended data Table 2a).

Second, error in the SST spatial pattern will cause a low bias in k_{qsat} (regression dilution bias). Typical magnitudes of this bias are estimated (see more detail in Methods section 'Estimating regression dilution bias'). This is done by regressing pairs of SST datasets against each other, with regression coefficients calculated using the sortav method (as used for k_{qsat}). Extended data Table 2b shows the results for each pair of SST datasets. The final column of Extended data Table 2b gives the 6 different estimates of regression dilution bias.

Third, the TRMM operational period does not fully overlap the AMIP SST-forced model simulation period. Thus, k_{qsat} estimated for the TRMM period will be different from that obtained if TRMM were operational throughout the AMIP period (due to a different set of SST patterns during the TRMM and AMIP period). Estimates of typical magnitudes of this error were obtained using samples from coupled ocean-atmosphere simulations (Extended data Figure 7), giving 85 samples of percentage error.

10000 estimates of k_{qsat} were then generated using Monte Carlo sampling. For each estimate, one of the three direct estimates (Extended data Table 2a) was selected at random, 'corrected' by a random selection from the 85 samples of percentage error (Extended data Figure 7), and further corrected using a random selection from the 6 estimates of regression dilution bias (Extended data Table 2b).

The lower bound of the 95% confidence interval of these 10000 estimates is then taken. We only use the lower bound for the following reason: it seems unlikely that the true regression dilution bias is weaker than the minimum value estimated here (around 10%, Extended data Table 2b). This is because we do not expect the SST datasets to be significantly closer to the real SST than they are to each other. However, it is plausible that the regression dilution bias could be larger than estimated, so we do not quote an upper bound for k_{qsat}.

The appropriate lower bound for an atmosphere-only AMIP model may be even higher than our quoted result, as the observed k_{qsat} value may be reduced by the effect of atmospheric internal variability on SST[19]. kqsat values from AMIP models are similar to or higher than from the equivalent coupled models. In contrast, this coupling issue should be small for our central estimate of k_{qsat} (next section), as that is based on metrics where ocean dynamics (for ENSO) or forcing associated with the mean state/seasonal cycle dominate the SST differences/gradients.

Central estimate of k_{qsat} in the real world

We start by finding where, geographically, the models are most sensitive to k_{qsat}. This is done using correlation coefficients (r) between k_{qsat} and the log of precipitation ratios

(Extended data Figure 5) for each latitude (or longitude for ENSO) in each panel in Figure 2. This reveals six discrete intervals (shaded in Figure 2) where $|r| > 0.6$ (about 50% variance explained), and a seventh, showing weaker correlations, over the Indian Ocean. For each model, we average the log of precipitation ratios over each interval, and take the difference from the equivalent value for TRMM. This gives seven error indices for each model (y-axes in Extended data Figure 6). For each interval, the 28 model error indices are regressed against k_{qsat} (x-axes in Extended data Figure 6), and k_{qsat} is estimated from where the line of best fit crosses zero error. This gives seven estimates of k_{qsat}; their mean is our central estimate.

We did not use the coupled ocean-atmosphere models to estimate k_{qsat}, due to the evident residual biases from SST error in these models, and because it isn't possible to use the ENSO response for coupled models, due to model differences in simulations of ENSO SST responses.

Estimating regression dilution bias

Regression dilution bias[51] arises when there is random error in the independent variable (e.g. q_{sat} in estimating k_{qsat}). This causes the regression gradient to be biased low. This bias reduces the gradient by a factor (β) that depends only on the characteristics of the independent variable. We estimate typical magnitudes of this bias using different observations of the independent variable (e.g. different SST datasets), as follows.

Say the vector of true values of the independent variable is **x**, and we have two different observational estimates, **x1** and **x2**. We first regress **x1** against **x2**. This regression uses the same methodology as when regressing the dependent variable against **x**: i.e. for estimating dilution bias in observations of k_{qsat}, the sortav method is applied; and for the satellite precipitation validation, the SST binning is used. The regression gradient (f_{12}) obtained from regressing **x1** against **x2**, will be biased low by a factor β_2 (from error in **x2**). We then regress **x2** against **x1**. This regression gradient (f_{21}) will be biased low by a factor β_1 (from error in **x1**).

We then estimate the dilution bias as:

$$\beta \approx \mathrm{sqrt}(\beta_2 \cdot \beta_1) = \mathrm{sqrt}(f_{12} \cdot f_{21})$$

(with no dilution bias, $f_{12} \cdot f_{21} = 1$, by definition).

This method assumes that the errors in **x1** and **x2** are independent.

Column dry static energy budget (DSE)

Dry static energy (DSE) is given by $s = c_pT + gz$, where c_p is the specific heat at constant pressure, T is the temperature, g is the gravitational acceleration, and z is altitude). DSE is affected by advection, precipitation, radiation and sensible heat:

$$-\langle \omega \frac{\partial s}{\partial p} \rangle - \langle v.\nabla s \rangle + LP + R + Q_{turb} = 0,$$

where ω is the vertical pressure velocity, p is pressure, v is horizontal wind. Angle brackets represent the mass-weighted vertical integral from 1000-100hPa. The first term represents import of DSE via column-integrated vertical advection; the second is horizontal advection. P is the total surface precipitation and L the latent heat of condensation; R is net radiation into the atmospheric column; and Q_{turb} is the surface sensible heat flux.

Methods References

51. Hutcheon, J. A., Chiolero, A. & Hanley, J. A. Random measurement error and regression dilution bias. *Br. Med. J.* **340**, (2010).

Data availability

Datafiles with estimates of k_{qsat} for models and observations, along with sample plotting code, are available from http://doi.org/10.5281/zenodo.3878691. Data from the integration of CNRM-CM6 with the CM5 convection scheme (denoted CNRM-CM6-conv5) are available from https://doi.org/10.5281/zenodo.3875005. Model and observational data is available at the following websites. CMIP5: https://cmip.llnl.gov/cmip5/; CMIP6: https://esgf-node.llnl.gov/projects/cmip6/; GTMBA: https://www.pmel.noaa.gov/gtmba/; TRMM: https://pmm.nasa.gov/data-access/downloads/trmm; GPCP and COBE: https://www.esrl.noaa.gov/psd/; HadISST: https://www.metoffice.gov.uk/hadobs/hadisst; ERSST: http://www1.ncdc.noaa.gov/pub/data/cmb/ersst/v4/netcdf/ .

Code Availability

Python code for calculating k_{qsat}, including the sortav regression routine, is available from http://doi.org/10.5281/zenodo.3878691.

End Notes

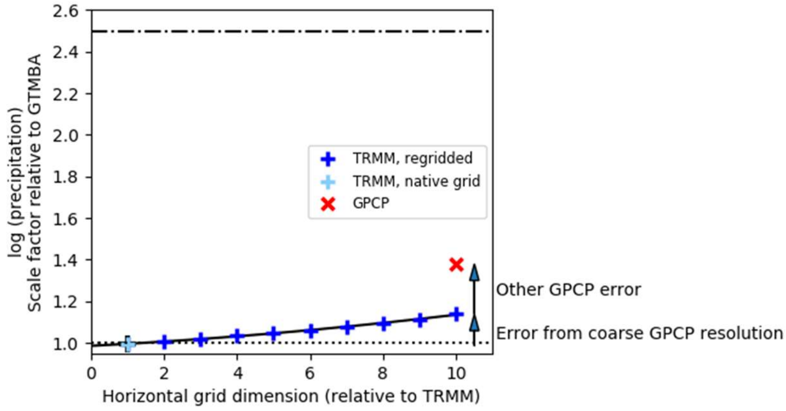
Acknowledgements This work was supported jointly by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra, and by the Newton Fund through the Met Office Climate Science for Service Partnership Brazil (CSSP Brazil). SSR was supported by the National Aeronautics and Space Administration Grant 80NSSC17K0227 and the Korean Meteorological Administration Research and Development Program under grant KMI2018-03110. We acknowledge the GTMBA Project Office of NOAA/PMEL for making the GTMBA data available. The QuikSCAT data were obtained from the NASA EOSDIS Physical Oceanography Distributed Active Archive Center (PO.DAAC) at the Jet Propulsion Laboratory, Pasadena, CA (http://dx.doi.org/10.5067/GHGMR-4FJ01). We acknowledge NOAA/ESRL PSD for the wind profiler data. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups (listed in Methods) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. GPCP data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. COBE-SST2 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/

Author Contributions P.G. conceived and designed the study and performed the analysis. All authors contributed to scientific interpretation and wrote the manuscript. R.R. performed the

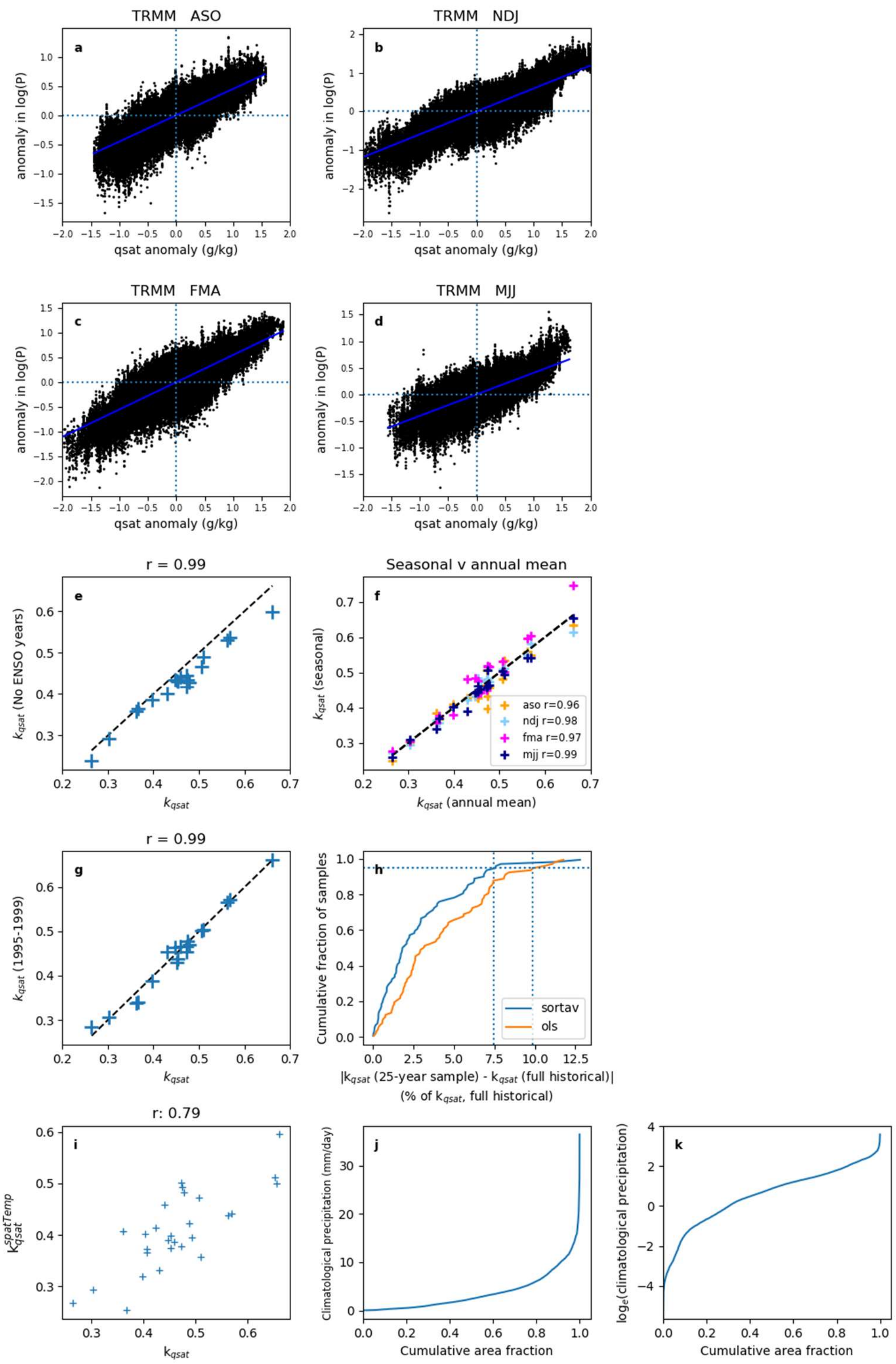
CNRM model simulations. P.G., R.C., C.E.H. and R.R. contributed understanding on physical processes. J.K. provided knowledge of SST observational uncertainty and datasets.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors have no competing financial interests. Correspondence and requests for materials should be addressed to peter.good@metoffice.gov.uk.

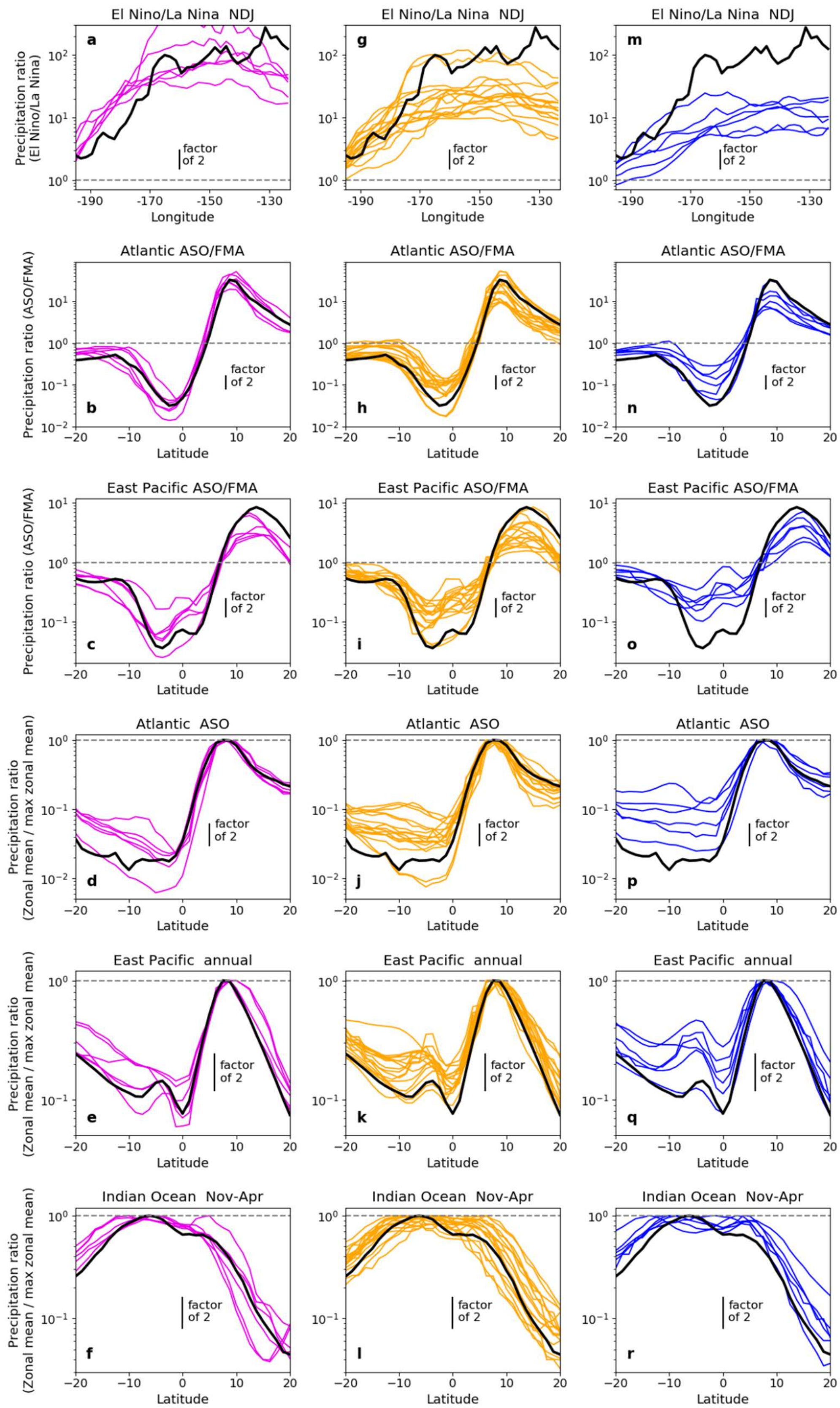
Extended Data Figures



Extended data Figure 1. Effect of low spatial resolution in GPCP satellite observations of log(seasonal precipitation). y-axis: regression gradient in validation against GTMBA raingauge data (i.e. gradients in Figure 1 for light blue and red symbols). x-axis: horizontal grid dimension relative to TRMM (e.g. the TRMM resolution is 0.25°, ten times smaller than the GPCP resolution of 2.5°, so the red symbol is placed at x=10). Dark blue symbols: results when TRMM data is regridded (by area averaging) to coarser grids. The coarser grids are chosen so the grid box edges overlap edges of the native TRMM grid. To give the errors context, the dash-dot line marks the ratio between the largest and smallest model values of k_{qsat} (2.5). Solid black line is a quadratic least-squares best fit line through the TRMM-based data. The intercept of the TRMM best-fit curve at x=0 (i.e. infinitely fine grid) is very close to the value estimated on the TRMM native grid (light blue symbol), indicating that the TRMM grid is sufficiently fine for comparison with the rangauge data on seasonal timescales.

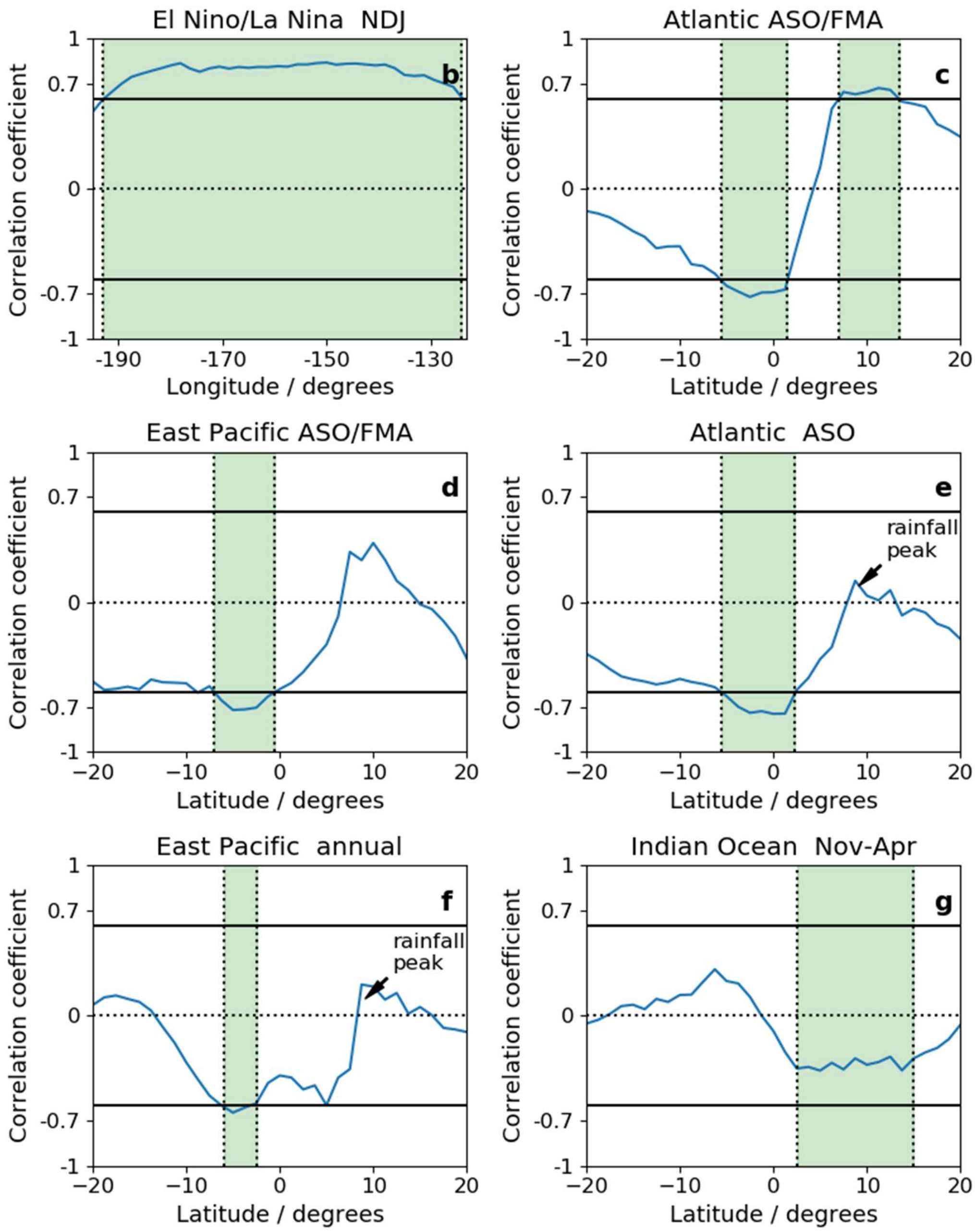


Extended data Figure 2. Testing the method of estimating k_{qsat}. **a-d**: example results of the sortav method for TRMM precipitation and HadISST SST, for different seasons: mean vectors of anomalies in (y-axis) log(precipitation) and (x-axis) q_{sat}; k_{qsat} is given by the gradient of the blue best-fit regression line. **e**, y-axis: k_{qsat} calculated after excluding the 9 years with the largest absolute value of the nino3.4 index; x-axis: default k_{qsat} (one symbol per model); k_{qsat} is on average 6% lower when ENSO years excluded, due to a small sensitivity to the ENSO characteristic spatial pattern; but the model ranking is largely unchanged (r = 0.99). **f**, k_{qsat} calculated for individual seasons versus the annual mean value; **g** k_{qsat} using only years 1995-1999 versus the full 25-year estimate; **h,** estimating variability (due to internal variability in SST patterns) in k_{qsat} estimated from 25 years of data: for each coupled ocean-atmosphere model, k_{qsat} is estimated both for the full historical run, and for all 25-year chunks. Panel shows the cumulative distribution function of absolute percentage differences between the 25-year estimates and the full estimates (95% of samples are within 8% of the long-term value from the full historical run). This panel shows results for two methods of estimating k_{qsat}: our 'sortav' method (as used throughout the manuscript), and standard OLS regression between seasonal anomalies in log(precipitation) and q_{sat}. **i** comparing $k_{qsat}^{spattemp}$ with k_{qsat}; each cross represents one CMIP5 model. **j,k** Cumulative distribution functions of **j** climatological mean precipitation and **k** log(precipitation). From HadGEM2-A, May-July season (same picture seen in other seasons).

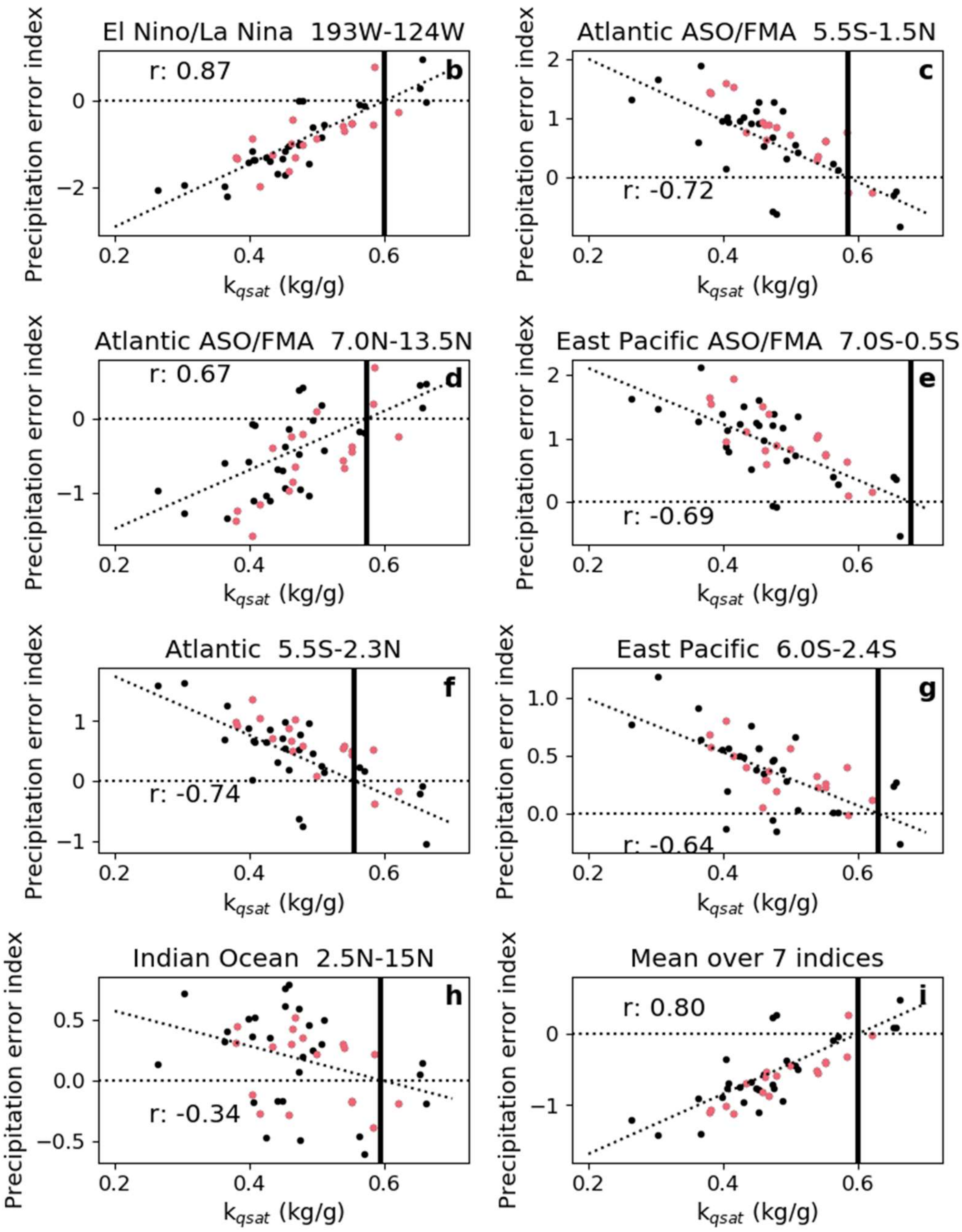


Extended data Figure 3. Model biases for the high, mid-range and low-k_{qsat} models separately. As Figure 2, for **a-f** high-k_{qsat} models; **g-l** mid-k_{qsat} models; **m-r** low-k_{qsat} models.

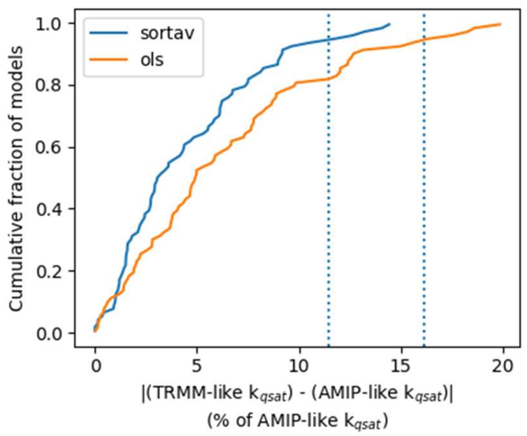
Extended data Figure 4. Testing potential errors in the satellite validation against GTMBA. **a,b** testing for regression dilution bias from error in TRMM observations: as Figure 1, but for **a** TRMM versus GPCP (both interpolated to GTMBA sites and masked as in Figure 1) and **b** GPCP versus TRMM. **c-f** testing for effects of SST uncertainty on the binning: as Figure 1, but using **c,d** ERSST and **e,f** COBE SST datasets to bin the data.



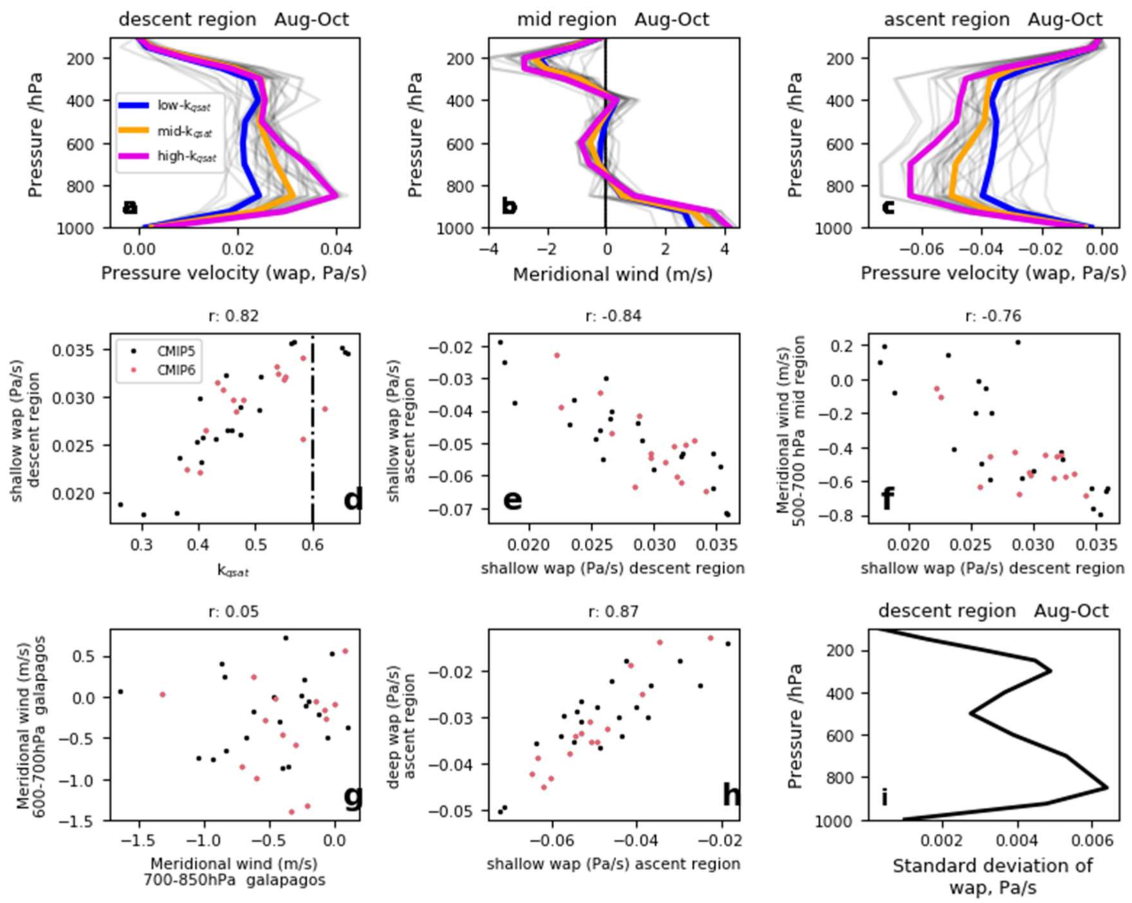
Extended data Figure 5. Regions where models are most sensitive to k_{qsat}. For each latitude of each region: y-axis shows Pearson correlation coefficients (r) between the 28 different CMIP5 model values k_{qsat}, and the 28 CMIP5 model values of the logarithm of the precipitation ratio for that latitude and region (i.e. the logarithm of the grey lines in Figure 2a-f). Green bands mark the latitude intervals chosen to estimate the observational constraints on k_{qsat} (**a-e**: intervals chosen where $|r|> 0.6$; **f**, a band of most negative r is chosen). Coefficients close to zero near 8N in the Atlantic and East Pacific spatial patterns correspond to the latitude of the precipitation peak in most models (the model spread in the precipitation peak is scaled out; coefficients are not exactly zero as there is a small model spread in the latitude of the precipitation peak).



Extended data Figure 6. Scatter plots underpinning the central observational estimate of k_{qsat}. **a-g** Precipitation error index versus k_{qsat} for each of the 7 latitude intervals highlighted in Figure 2. Y-axes: logarithm of precipitation ratio, averaged over each latitude band, minus the equivalent value for TRMM observations, for (black) CMIP5 and (red) CMIP6 models. Dotted lines: linear least-squares fits (using CMIP5 data only). Vertical black line: k_{qsat} estimate for each latitude interval, from the intercept of the green line with zero error index (dotted line). **h** Mean precipitation error index versus k_{qsat}: mean error index is averaged over the 7 indices in the other panels (after the signs of the 5 indices with negative best-fit slopes were changed, to ensure a positive correlation with k_{qsat}).



Extended data Figure 7. Supporting results for observational estimate of the k_{qsat} lower bound. Estimating error, from internal variability, due to the fact that the TRMM operational period only partly overlaps the time period simulated by the AMIP SST-forced models. Error magnitudes are estimated from the coupled ocean-atmosphere simulations, using differences between k_{qsat} estimated from all possible overlapping 17-year (TRMM-like) and 25-year (AMIP-like) periods (with the same overlap as TRMM and the 25-year SST-forced model simulations). Results are given for two methods of estimating k_{qsat}: our 'sortav' method (as used throughout the manuscript), and standard OLS regression between seasonal anomalies in log(precipitation) and q_{sat}.



Extended Data Figure 8. Atmospheric circulation measures in CMIP5 and CMIP6 models. **a-c** thick lines are CMIP5 composite means, for (magenta) high k_{qsat} subset; (blue) low k_{qsat} subset and (gold) intermediate k_{qsat}. Thin grey lines are individual models (CMIP5 and CMIP6). Descent (5S-1N), mid (1-7N) and ascent (7-13N) regions are marked by vertical dotted lines in Figure 5c-e. **d-h**: each symbol represents one CMIP5 (black) or CMIP6 (red) model. Title gives Pearson correlation coefficient. **d** shallow descent versus k_{qsat}; vertical line marks our best estimate of k_{qsat}. **e** shallow ascent versus shallow descent. **f** shallow meridional return flow versus shallow descent. **g** shallow versus very-shallow meridional wind, over Galapgos: the negligible correlation indicates different physical processes at these two levels. **h** deep versus shallow ascent. **i** standard deviation, across models, of the pressure velocity (wap) at each pressure level.

Extended Data Tables

Descent region, column integrals	High k_{qsat} mean	Mid k_{qsat} mean	Low k_{qsat} mean	High k_{qsat} – Low k_{qsat}
Net radiation + sensible heat flux	-129.4	-125.9	-127.5	-1.8
Latent heating by precipitation	14.4	22.9	35.5	-21.1
Vertical advection by mean vertical velocity, integrated over 100-1000 hPa	130.1	110.0	95.5	34.6
Vertical advection by mean vertical velocity, integrated over 600-1000 hPa	74.5	57.4	48.1	26.4
Residual advection	-15.1	-7.0	-3.5	-11.6

Extended Data Table 1. Column-integrated dry static energy budget for the descent region, Aug-Oct, averaged over the high-, mid- and low-k_{qsat} groups of CMIP5 models. Vertical advection by mean vertical velocity is calculated using seasonal mean vertical velocity. The final row, calculated as a residual from the first three columns, includes horizontal advection, and vertical advection by transient eddies.

a SST dataset	Unscaled k_{qsat} estimate (no bias correction)
HadISST	0.58
ERSST	0.49
COBE	0.51

b SST1	SST2	K1 (X = SST1)	K2 (X = SST2)	SQRT(K1*K2)
AMIP	HadISST	0.77	1.01	0.89
AMIP	COBE	0.85	1.00	0.92
AMIP	ERSST	0.76	0.97	0.86
HadISST	COBE	0.76	0.79	0.78
HadISST	ERSST	0.83	0.55	0.68
COBE	ERSST	0.92	0.54	0.70

Extended data Table 2. Supporting results for observational estimate of the k_{qsat} lower bound (see Methods for details). **a** Unscaled estimates for k_{qsat} directly estimated using TRMM precipitation and three different SST datasets (the AMIP SST dataset was not used due to limited temporal overlap with the TRMM operational period). These values are contaminated by regression dilution bias so do not represent central estimates. **b** Estimating typical values of regression dilution bias from each pair of SST datasets. K1 is the gradient from linear regression when regressing SST1 against SST2 (using the sortav regression method). K2 is the value obtained when regressing SST2 against SST1.