This is a repository copy of *High sensitivity of tropical precipitation to local sea surface temperature*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/170374/

Version: 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

21 remote effects of conditions elsewhere. Such remote effects occur, for example, from El

22 Niño Southern Oscillation (ENSO) events in the equatorial Pacific, which alter precipitation

23 across the tropics. Atmospheric circulations associated with tropical precipitation are

24 predominantly deep, extending up to the tropopause. Shallow atmospheric circulations[6–8],

25 impacting the lower troposphere, also occur, but the importance of their interaction with

26 precipitation is unclear. Uncertainty in precipitation observations[9,10], and limited

27 observations of shallow circulations[11], further obstruct understanding of the ocean's influence

28 on weather and climate. Despite decades of research, persistent biases remain in many

29 numerical model simulations[12–18], including excessively-wide tropical rainbands[14,18], the

30 'double-intertropical convergence zone (ITCZ) problem'[12,16,17] and too-weak responses to

31 ENSO[15]. These demonstrate stubborn gaps in our understanding, reducing confidence in

32 forecasts and projections. Here we show that the real world has a high sensitivity of seasonal

33 tropical precipitation to local sea-surface temperature. Our best observational estimate is

34 80% precipitation change per g/kg change in the saturation specific humidity (itself a

35 function of the ocean surface temperature). This observed sensitivity is higher than in 43 of

36 the 47 climate models studied, and is associated with strong shallow circulations. Models

37 with more realistic sensitivity have smaller biases across a wide range of metrics. Our results

38 apply to both temporal and spatial variation, over regions where climatological precipitation

39 is around 1 mm/day or greater. Novel analysis of multiple independent observations, physical

40 constraints and model data, underpin these findings. The spread in model behaviour is further

41 linked to differences in shallow convection, providing a focus for accelerated research, to

42 improve seasonal forecasts through multidecadal climate projections.

43

44

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:

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

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.

69 k_{qsat} relates most directly to the strength of percentage variations in precipitation. Writing

70 Equation 1 in exponential form, percentage precipitation differences are a function of k_{qsat}

71 and q_{sat}:

72 $(P_1 - P_0)/P_0 * 100 \approx 100 [\exp(k_{qsat} * (q_{sat} - q_{sat,0})) - 1].$ Equation 2

73 Absolute differences depend also on the reference precipitation P_0:

74 $P_1 - P_0 \approx P_0[(\exp(k_{qsat} * (q_{sat} - q_{sat,0})) - 1],$ Equation 3

75 (in absolute terms, precipitation variations are largest in regions of large mean precipitation[19].

76 However, we will show that spatial variation in P_0 itself also depends partly on k_{qsat}.

77

Validating satellite observations

79 Given uncertainty in precipitation observations[9,10], we perform a high-precision evaluation of

80 log(precipitation) (as Equation 1) from two satellite datasets: TRMM[23,24] (3B43, v7) and

81 GPCP[25] (v2.3), both for 1998-2015. We do this (Methods) using in-situ raingauge data from

82 89 buoys of the Global Tropical Moored Buoy Array (GTMBA)[26–28]. Satellite-GTMBA

83 validation is challenging: on top of satellite error[9,10], the GTMBA point observations include

84 noise from small-scale variability unresolved by satellite data, missing data, error in

85 individual raingauges and wind undercatch[29]. Our method reduces this noise considerably,

86 giving a tight relationship between GTMBA and TRMM data (Figure 1a). Critically, the best

87 fit gradient ≈ 1, so TRMM accurately retrieves differences in log precipitation. On the other

88 hand, GPCP underestimates differences in log precipitation (Figure 1b, gradient > 1;

89 differences between TRMM and GPCP emerge primarily at low precipitation[9]), although

90 GPCP is more suitable over larger spatial scales (Extended data Figure 1). Since TRMM

91 captures differences in log(precipitation) more accurately than GPCP, TRMM is used below.

92

Model precipitation simulations

94 We first highlight precipitation biases in 28 atmospheric models from the fifth Coupled

95 Model Intercomparison Project (Methods), each forced by observed SST (CMIP5 AMIP

96 experiment; Figure 2a-f). We quantify temporal, seasonal and spatial variation in

97 precipitation: the 1997-98 El Niño divided by the mean of the 1998-2000 La Niñas; Aug-Oct

98 divided by Feb-Mar seasons, and precipitation scaled by its latitudinal maximum. Our

99 metrics coincide with significant spatial or temporal differences in SST (Methods). Spatial

100 variation across the west Pacific is excluded, for example, because spatial gradients in SST

101 are weak there, so model differences in k_{qsat} will be less important for spatial variation there.

102 Given the form of Equation 1, precipitation is shown on log scales as ratios. Although some

103 models are close to the observations, in others, biases exceed a factor of five in the El

104 Niño/La Niña ratio, seasonal cycles over the Atlantic and West Pacific, and in the Atlantic

105 spatial pattern for the Aug-Oct season (Figure 2a-d). Biases over a factor of two occur in the

106 spatial patterns of the East Pacific annual mean (the long-standing 'double-ITCZ'

107 problem[12,17]) and the Indian Ocean for November-April (Figure 2e-f). Such biases are

108 known, but their causes are not well understood.

109

110 These biases (Figure 2a-f) all correspond to excessively weak spatial/temporal variations in

111 precipitation (precipitation ratios too close to 1; including excessively-wide inter-tropical

112 convergence zones[14,18]). This suggests a hypothesis (H_0), that the sensitivity of seasonal

113 precipitation to local SST (k_{qsat}) may be too weak in many models.

114

115   To test H_0 objectively, we use a method independent of Figure 2 (Figure 2 was used to

116   propose H_0). This involves estimating k_{qsat} for each model using different data.

117

118   **Evaluating k_{qsat} in models**

119   We evaluate k_{qsat} using interannual variability in seasonal mean precipitation and SST (the

120   AMIP SST dataset[30] used to drive the model experiments; using years 1980-2005). k_{qsat} is

121   calculated using gridpoint values of seasonal precipitation and q_{sat}, from each location in the

122   study region, and for each year. With these data, Equation 1 becomes a model of the effect of

123   local interannual SST variability on precipitation:

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

125   where (x,t) indicates values for each gridpoint and year; and here, P_0 and $q_{sat,0}$ are the

126   corresponding climatological means for each gridpoint. We estimate k_{qsat} from these data

127   using a modified regression approach (detailed in Methods), minimising the influence of

128   other processes in Equation 1.

129

130   To minimise observational error, we exclude the 30% of the tropical oceans with the lowest

131   climatological mean SST (Figure 3b-e, area outside white contour).

132

133   Taking logarithms means that all areas of our study region contribute relatively equally to our

134   k_{qsat} estimate (Methods). Consequently, k_{qsat} is relevant over most of the tropical oceans

135   (Figure 3a, correlations are high except for the left bar). k_{qsat} is inapplicable over the coolest,

136   driest ocean regions (Figure 3a, left bar; area masked in Figure 3b-e). The applicable region

137   corresponds to climatological precipitation > ~1 mm/day (Figure 3b-e, orange contour).

138

139   k_{qsat} is intended to be independent of large-scale SST spatial patterns. To avoid bias from the

140   large, recurrent ENSO pattern, our 'sortav' regression method first processes the data so all

141   years contribute equally. Linear regression is then applied to obtain k_{qsat}. Rankings of

142   CMIP5 models by k_{qsat} are robust: insensitive to season, to using fewer years of data, or to

143   excluding ENSO years - Extended data Figure 2e-g. Calculated this way, k_{qsat} is less

144   sensitive to the time period used than with simple least squares regression (Extended Data

145   Figures 2h, 7).

146

147   We find that the sensitivity of precipitation to local SST variability is much stronger in some

148   models than in others: k_{qsat} varies across CMIP5 models by a factor of 2.5 (0.26-0.66 kg/g;

149   median = 0.46). We group the models into 'high-k_{qsat}' (the 6 models with the largest k_{qsat}

150   values), 'low-k_{qsat}' (the lowest 6 k_{qsat} values) and 'mid-range' subsets.

151

152   In Equation 2, setting (q_{sat} - $q_{sat,0}$) to 1, expresses k_{qsat} as the percentage precipitation change

153   per g/kg change in saturation specific humidity (q_{sat}):

154

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

156

157 Expressed this way, the precipitation sensitivity in CMIP5 models spans 30-93% per g/kg

158 (median = 58%). For context, q_{sat} can vary by a few g/kg, 10° either side of the East Pacific

159 ITCZ during Aug-Oct, and anomalies during ENSO events have a similar magnitude.

160

High sensitivity of precipitation to SST

162 We hypothesised above (H_0) that k_{qsat} may be too low in most models. To begin testing this,

163 the results in Figure 2a-f are replotted, but with the 'high-k_{qsat}' subset of models highlighted in

164 magenta (Figure 2g-l). The 'high-k_{qsat}' subset shows much better agreement with TRMM than

165 the full ensemble, in all six panels. Conversely, the 'low-k_{qsat}' subset performs much worse

166 (Extended data Figure 3). Next, we calculate $k_{qsat}^{(spatial)}$ (Methods): as k_{qsat}, but using spatial

167 patterns in climate means, rather than internal variability (Figure 4b). Again, models closest

168 to the observations (Figure 4b, horizontal line) tend to be those with high k_{qsat}. These results

169 all imply that k_{qsat} should be high in the real world (H_0).

170

171 These results also show that k_{qsat} is relevant to both spatial and temporal variations in

172 precipitation. We emphasise this by quantifying the overall sensitivity of precipitation to

173 local SST ($k_{qsat}^{spattemp}$), including both spatial and temporal variations (including spatial

174 variation in P_0, Methods). $k_{qsat}^{spattemp}$ is well correlated with k_{qsat} (Extended data Figure 2i).

175 This confirms that k_{qsat} is a useful measure of the underlying sensitivity of precipitation to

176 local SST, relevant to spatial and temporal variations. k_{qsat} does not give information about

177 tropical mean precipitation, which is governed by different processes[31]. k_{qsat} remains our

178 primary measure of precipitation sensitivity to local SST, because it is insensitive to details of

179 SST patterns. In contrast, $k_{qsat}^{spattemp}$ and $k_{qsat}^{(spatial)}$ may be sensitive to the specific spatial

180   patterns in climatological SST (Methods), explaining some of the noise in Figure 4b and

181   Extended data Figure 2i.

182

183   We estimate a lower bound for k_{qsat}, using observed interannual variability (independent of

184   Figure 2; Methods). Three values of k_{qsat} are estimated, exactly as for the models, but using

185   TRMM precipitation, and q_{sat} from each of three different SST datasets (HadISST[32] version

186   1.1, ERSST[33] version 4 and COBE[34] version 2). Uncertainties are estimated, from SST error

187   (including regression dilution bias) and internal variability (the TRMM observational period

188   only partly overlaps the model simulation period). A lower observational bound (95%

189   confidence) of 0.51 kg/g for k_{qsat} is obtained.

190

191   For a central observational estimate of k_{qsat} (details in Methods), we return to Figure 2. We

192   ask: if all CMIP5 models had the same value of k_{qsat}, with what value would they best

193   reproduce the observations in Figure 2? We first find where, geographically, the models are

194   most sensitive to k_{qsat}. This reveals seven intervals (shaded in Figure 2g-l). For each interval,

195   model errors relative to TRMM are regressed against modelled k_{qsat} (Extended data Figure 6).

196   For each interval, k_{qsat} is estimated as where the regression line intercepts the x-axis (the

197   value for a theoretical model with zero precipitation error). These seven estimates of k_{qsat}

198   range from 0.56 to 0.68 kg/g (Figure 4a, white dashed lines), all larger than our lower bound

199   estimate. The spread of estimates comes from uncertainty in processes not quantified by

200   k_{qsat}. The range of conditions used, covering spatial, seasonal and temporal variability across

201   different locations, helps to quantify and mitigate this uncertainty. Robustness is tested by

202   plotting results from the sixth model intercomparison project (CMIP6, not used to select the

203   seven intervals) on Extended data Figure 6. The mean of the seven k_{qsat} values (0.6 kg/g; or,

204 using Equation 4, 80% per g/kg) is our central estimate (for 1980-2005; other periods would

205 give slightly different values, from internal SST variability - Extended data Figure 2h).

206

207 These results, from two independent methods, suggest that most models underestimate k_{qsat}.

208 Our central estimate (0.6 kg/g, Figure 4a, horizontal black line; Figure 4b, vertical line) is

209 greater than 43 of the 47 model values from CMIP5 and CMIP6. This implies that models

210 underestimating the sensitivity of precipitation to local SST underlies a range of model biases

211 over tropical oceans. CMIP5 and CMIP6 have similar ranges of k_{qsat} values (Figure 4a),

212 highlighting the need for accelerated model development.

213

214 Other studies[35] have found biases in a different aspect of the SST-precipitation relationship:

215 model precipitation often tracks SST maxima more closely than in observations. We quantify

216 this in each CMIP5 model as the correlation coefficient between climatological spatial

217 patterns of precipitation and SST, for each season, then average the four seasonal values.

218 This 'spatial-correlation index' is uncorrelated with k_{qsat} (r = 0.01; i.e. models with a high

219 spatial-correlation index can have high, low or intermediate k_{qsat}), so it involves different

220 processes.

221

222 **Processes behind uncertainty in k_{qsat}**

223 To guide model improvements, we explore what causes model differences in k_{qsat}, revealing

224 links to shallow atmospheric circulations. We first note that k_{qsat} involves processes

225 unrelated to tropical mean precipitation: the correlation across CMIP5 models between the

226   two measures is 0.03. Energy budgets constrain tropical mean precipitation[31], while the value

227   of k_{qsat} affects precipitation variation in both time and space.

228

229   Beginning with interannual variability, we define $k_{qsat}^{wap}(p)$: the sensitivity of the vertical

230   pressure velocity (wap) to local SST change, at each pressure level (p). This is evaluated like

231   k_{qsat}, using data from all seasons across the tropical oceans, but using wap(p) instead of

232   log(precipitation). A deep mode dominates tropical variability[36], so the CMIP5 mean profile

233   of k_{qsat}^{wap} peaks around 450 hPa (Figure 5a). In contrast, model spread in k_{qsat} is linked to

234   shallow circulations: correlations between $k_{qsat}^{wap}(p)$ and k_{qsat} (Figure 5b) peak near 700hPa (r

235   = -0.9; correlations are small at 1000hPa, as wap approaches zero near the surface). That is,

236   in models with high k_{qsat} (as in the real world), shallow circulations respond strongly to SST

237   anomalies. Although deep circulations are important[36] in all models, and shallow circulations

238   have been linked to mean precipitation, especially over the Eastern Pacific[7], our novel finding

239   is that shallow circulations are central to model uncertainty in SST-driven precipitation

240   variability, across the tropics.

241

242   Shallow circulations are further linked to model differences in climate means. We study

243   zonal, ocean-only means over 180W-10E (most of the Pacific, entire Atlantic), in Aug-Oct,

244   when meridional SST gradients are strong. CMIP5 results are used to define key regions

245   (black and orange lines in Figure 5c-e), and CMIP6 used to test the conclusions. As

246   expected[4,7], CMIP5 mean vertical velocity profiles are bottom heavy, but extend throughout

247   the troposphere (Figure 5c, colours). Meridional wind (white contours) peaks near the

248   tropopause, with a weaker shallow flow between 500-700hPa[37]. Again, however, inter-model

249   spread in k_{qsat} is associated with shallow circulations: models with high k_{qsat} have stronger

250   shallow descent south of the ITCZ between 600-850hPa (Figure 5d, yellow; Extended Data

251   Figure 8a, magenta line; Extended Data Figure 8d). They also tend to have stronger shallow

252   ascent in the ITCZ (Figure 5d, blue region; Extended Data Figure 8c, magenta line), stronger

253   trade winds and stronger return flow between 500-700hPa (Figure 5d, white contours;

254   Extended Data Figure 8b, magenta line). This shallow circulation is weak in the low-k_{qsat}

255   mean (Extended Data Figure 8a-c, blue lines).

256

257   The weak link between model differences in k_{qsat} and deep circulations may arise partly from

258   physical constraints. In descending air, differences in vertical velocity are largest below

259   about 600 hPa (Extended Data Figure 8a,i). This is partly explained by the vertically-

260   integrated dry static energy (DSE) budget. This budget constrains vertical velocities,

261   requiring balance between radiative, sensible and latent heating, and advection of DSE

262   (Methods). In descending air above 600 hPa, there are fewer uncertain processes affecting

263   this budget, with negligible energy input from cloud and precipitation. Here, therefore,

264   downward advection of DSE is balanced mostly by dry clear-sky radiative cooling (Extended

265   Data Table 1). Further, vertical temperature profiles are similar across CMIP5 models. This

266   constrains both radiative cooling and vertical gradients of DSE, limiting model differences in

267   vertical velocity. Model temperatures are constrained, near the surface by SST, and near the

268   tropopause as modelling groups aim to reproduce observed outgoing longwave radiation.

269   Below 600hPa, there are additional sources of uncertainty, from cloud and precipitation[38],

270   leading to larger model differences in shallow descent, and so stronger links to differences in

271   k_{qsat}.

272

273   The depth of the meridional return flow (500-700hPa, Figure 5e, white contours) suggests a

274   circulation driven by precipitating shallow convection[37,39]. The alternative, sea-breeze

275   mechanism reaches lower levels[37]. Both circulation types may exist over the Galapagos

276   (Extended Data Figure 8g): here, model meridional winds between 600-700hPa (shallow-

277   convection-type) are uncorrelated with those between 700-850hPa (sea-breeze-type),

278   indicating different physical processes at these two levels. The column dry static energy

279   budget (Methods) also implicates precipitating shallow convection. Model differences in this

280   budget are predominantly a balance between vertical advection integrated over 600-1000 hPa,

281   and precipitation latent heating (Extended Data Table 1, final column): in descending air

282   below 600 hPa, high-k_{qsat} models have stronger shallow advective warming from stronger

283   shallow descent, balancing weaker warming from weaker precipitation. As expected, the

284   ensemble mean in descending air is mostly a balance between radiative cooling and advective

285   warming.

286

287   Causality is hard to fully establish, but the most likely explanation of our results is that model

288   differences in k_{qsat} mostly originate from model differences in the behaviour of shallow

289   precipitating convection. Such differences would affect the sensitivity of precipitation to

290   SST directly. The consequent differences in shallow latent heating also appear to lead to

291   differences in shallow circulations. This couples shallow vertical velocities in descending

292   and ascending air (Extended Data Figure 8e), further modifying the sensitivity of

293   precipitation to SST.

294

295   Significant model differences in the physical representation of convection[40,41], including its

296   coupling to circulation[5,42], are well established. Improvements in k_{qsat} in the CNRM model

297 from CMIP5 (0.43 kg/g) to CMIP6 (0.54 kg/g) are largely associated with convection scheme

298 changes. Running CNRM-CM6 with the CNRM-CM5 convection scheme gives k_{qsat}=0.36

299 kg/g, even smaller than that of CNRM-CM5. That is, the effect of changing the convection

300 scheme is partly offset by changes in other schemes. Changes from CM5 to CM6 include the

301 shallow convection scheme and the transition from shallow to deep convection[43]. These, and

302 other physics schemes, including boundary layer, deep or mid-level convection, or

303 microphysics, could all affect how shallow precipitating convection responds to SST.

304

305 Other processes may have smaller contributions. Dry shallow circulations[44] are more

306 important over hot, dry land. A limited role for differences in radiation (Extended Data Table

307 1, final column), suggests cloud parameterizations are not dominant. Differences in

308 dynamical schemes are thought to be less likely to be important for tropical precipitation

309 biases, although coupling of dynamics to physics is important[45–47].

310

311 Models with stronger shallow circulations can also import more moist static energy in

312 ascending air, driving enhanced deep convection[48]. Our results support this (Extended Data

313 Figure 8h), showing that model differences in shallow ascent are strongly positively

314 correlated with differences in deep ascent. This is an indirect link to k_{qsat}, which is more

315 weakly associated with deep ascent rates (Figure 5d).

316

317 **Strong real-world shallow circulations**

318 Shallow circulations are challenging to observe[11], but our results suggest they are stronger in

319 the real world than in most models. Models with high k_{qsat} (as in the real world) tend to have

320   strong shallow circulations (in both climate means and internal variability). We test this

321   further with two independent observations, in Aug-Oct. In models, there is a strong (r=0.86)

322   relationship between shallow descent, and northward trade winds (Figure 4c). If the shallow

323   descent is strong in the real world, we would expect strong northward trades – and this is

324   confirmed by QuikSCAT satellite observations[49] (SeaWinds scatterometer, Level 3 product,

325   years 1999-2009 - horizontal line in Figure 4c). There is, similarly, a strong (r=0.81)

326   relationship (Figure 4d) between 500-700hPa meridional winds over our study region, and

327   the mean 600-700hPa meridional wind over Galapagos and Christmas Islands (915 MHZ

328   wind profiler[8,50] observation sites, years 1994-2005 at Galapagos, 1990-2002 at Christmas

329   Island; few observations reach above 600hPa). If the large-scale 500-700hPa wind is strong

330   in the real world, we would expect strong southward winds in the observations, and this is

331   what the wind profilers show (horizontal line in Figure 4d). These results, using multiple

332   observations, confirm a previous suggestion based on a single reanalysis product[5].

333

334   **Conclusions**

335   Our results show that k_{qsat} is linked to a range of model biases in precipitation and

336   atmospheric circulation. Improving k_{qsat} should reduce those biases, giving greater

337   confidence in seasonal through multi-decadal model projections. k_{qsat} affects the strength of

338   precipitation variation in both time and space. Other biases, in large-scale energy budgets,

339   and teleconnections, also affect precipitation. We show that k_{qsat} can be constrained by

340   observations, and give evidence that improving the representation of shallow tropical

341   precipitating convection, and its coupling to SST and circulation, could improve k_{qsat}. This

342   identifies specific model development goals and gives new ways of linking these to

343   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).

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

385 20. Rushley, S. S., Kim, D., Bretherton, C. S. & Ahn, M. S. Reexamining the Nonlinear Moisture-
386   Precipitation Relationship Over the Tropical Oceans. *Geophys. Res. Lett.* **45**, 1133–1140
387   (2018).

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

390 22. Kanemaru, K. & Masunaga, H. A Satellite Study of the Relationship between Sea Surface
391   Temperature and Column Water Vapor over Tropical and Subtropical Oceans. *J. Clim.* **26**,
392   4204–4218 (2013).

393 23. TRMM. Tropical Rainfall Measuring Mission, TRMM (TMPA/3B43) Rainfall Estimate L3 1
394   month 0.25 degree x 0.25 degree V7, Greenbelt, MD, Goddard Earth Sciences Data and
395   Information Services Center (GES DISC), Accessed 2017. 10.5067/TRMM/TMPA/MONTH/7
396   (2011).

397 24. Huffman, G. J. *et al.* The TRMM multisatellite precipitation analysis (TMPA): Quasi-global,
398   multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **8**, 38–55
399   (2007).

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

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

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

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

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

410 30. Taylor, K. E., Williamson, D. & Zwiers, F. W. The sea surface temperature and sea-ice
411   concentration boundary conditions for AMIP II simulations. *PCMDI Rep. No. 60, Progr. Clim.*
412   *Model Diagnosis Intercomparison, Lawrence Livermore Natl. Lab. Livermore, California, 25 pp*
413   (2000).

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

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

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

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

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

423   *Clim.* **19**, 935–958 (2006).

424   36.   Holloway, C. E. & Neelin, J. D. The convective cold top and quasi equilibrium. *J. Atmos. Sci.*
425   (2007) doi:10.1175/JAS3907.1.

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

428   38.   Knutson, T. R. & Manabe, S. Time-mean response over the tropical Pacific to increased CO2 in
429   a coupled ocean-atmosphere model. *J. Clim.* (1995) doi:10.1175/1520-
430   0442(1995)008<2181:TMROTT>2.0.CO;2.

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

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

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

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

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

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

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

445   46.   Oueslati, B. & Bellon, G. Convective Entrainment and Large-Scale Organization of Tropical
446   Precipitation: Sensitivity of the CNRM-CM5 Hierarchy of Models. *J. Clim.* **26**, 2931–2946
447   (2013).

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

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

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

457   50.   Gage, K. S., Williams, C. R. & Ecklund, W. L. UHF wind profilers: a new tool for diagnosing
458   tropical convective cloud systems. *Bull. - Am. Meteorol. Soc.* (1994) doi:10.1175/1520-
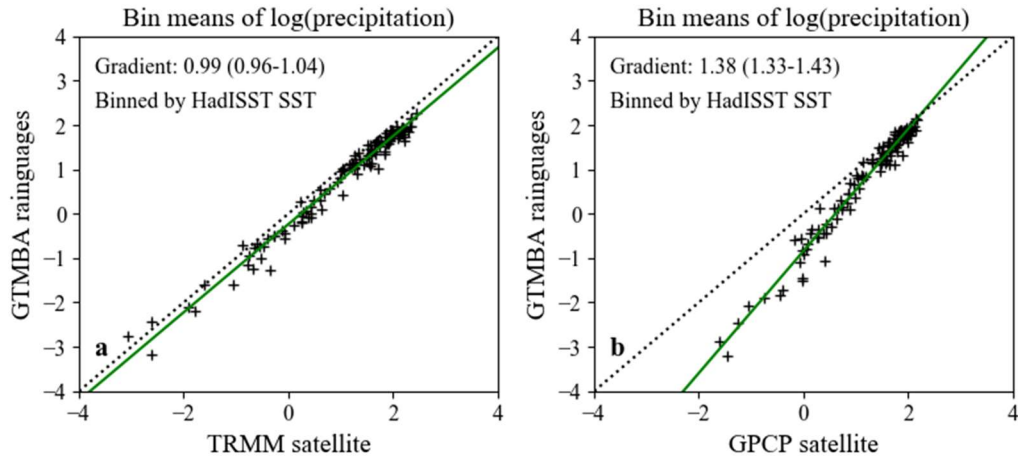459   0477(1994)075<2289:UWPANT>2.0.CO;2.

460

461 **Figures**

462

Figure 1.

464 Validating observations of log precipitation from satellites. GTMBA in-situ raingauge

465 observations versus satellite observations from **a** TRMM and **b** GPCP. Each symbol

466 represents the mean of all seasonal mean data within a given SST bin (Methods). Green line:

467 best fit line (gradient and its 95% confidence interval quoted in each figure); dotted line: 1:1

468 line.

469

470

471
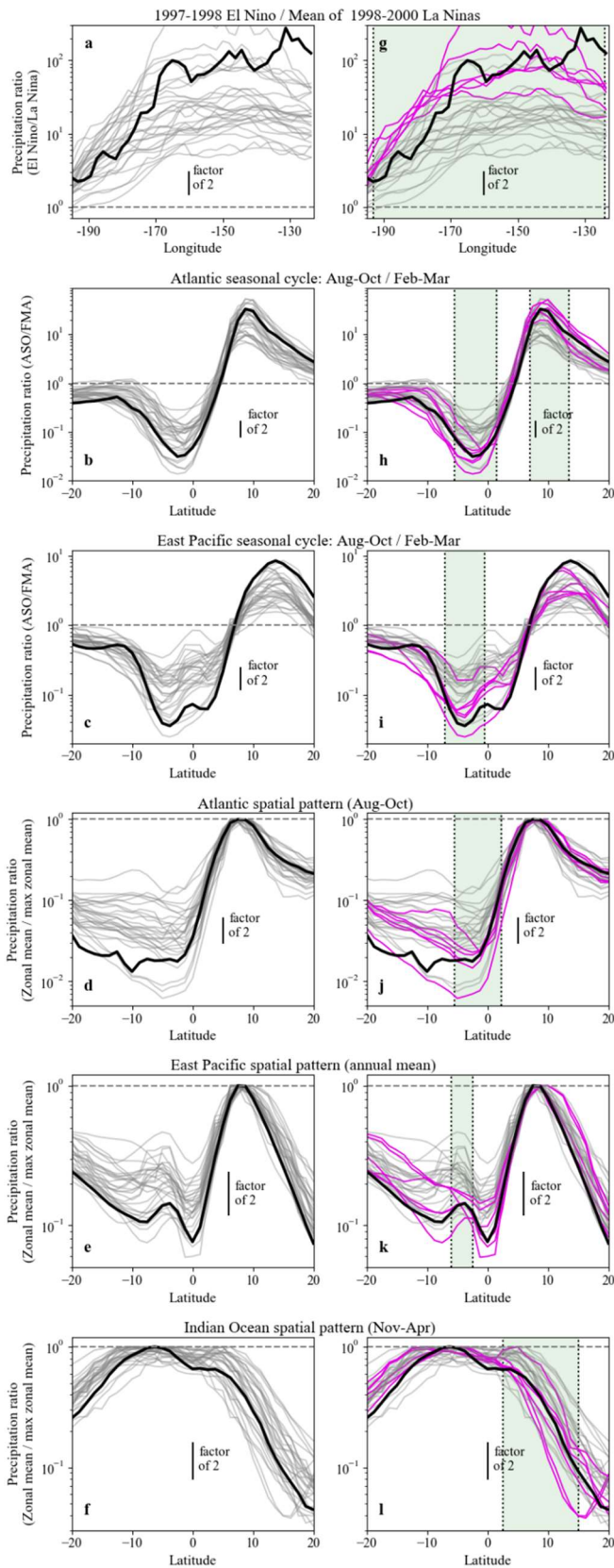
Figure 2. Model precipitation biases. (black) TRMM observations. Horizontal dashed line

marks precipitation ratio=1. **a-f** all CMIP5 models are shown in grey lines; **g-i** magenta:

474   'high-k_{qsat}' subset; grey: other models. Spatial patterns (bottom 3 rows) given by scaling

475   zonal mean precipitation by its latitudinal maximum. Green shading marks the intervals used

476   for the 7 estimates of k_{qsat}. These examples were chosen as they feature large

477   differences/gradients in SST. Precipitation ratios are plotted because of the form of Equation

478   1.

479

480
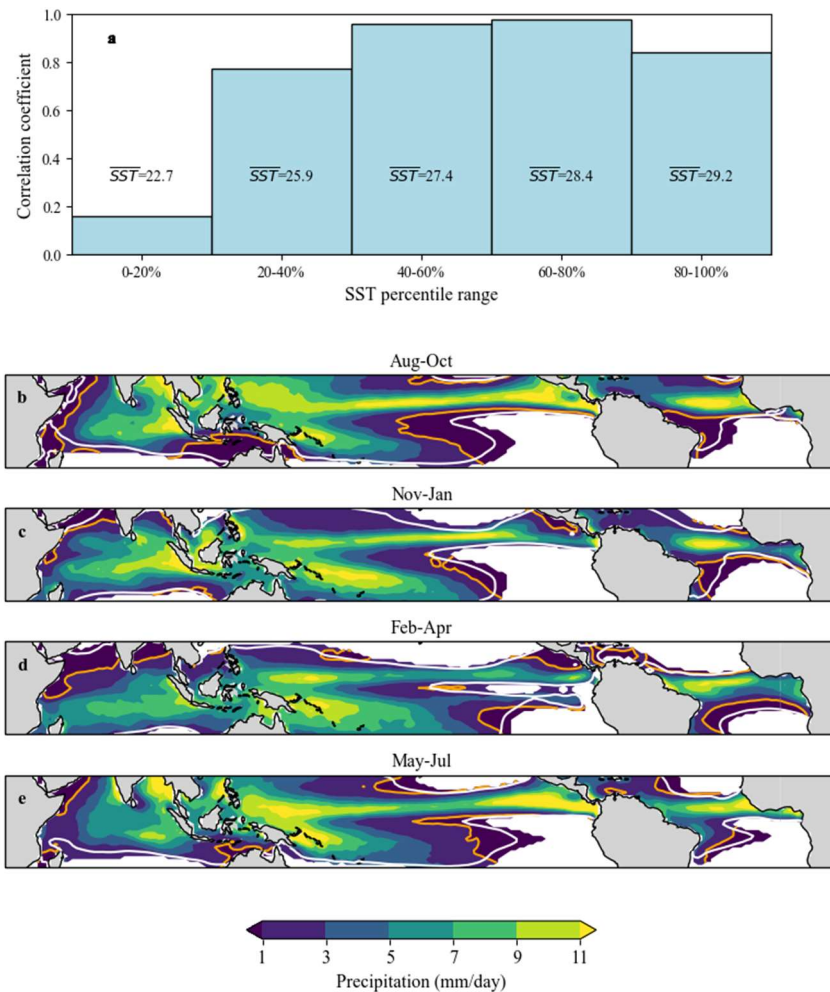
Figure 3. The region of applicability of kqsat. (**a**) each bar represents a climatological zone covering 20% of the tropical oceans, defined by the seasonal climatological SSTs (e.g. the left bar is the zone with the coolest 20% of SSTs – white masked ocean in the maps below). Climatological zones 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.

493
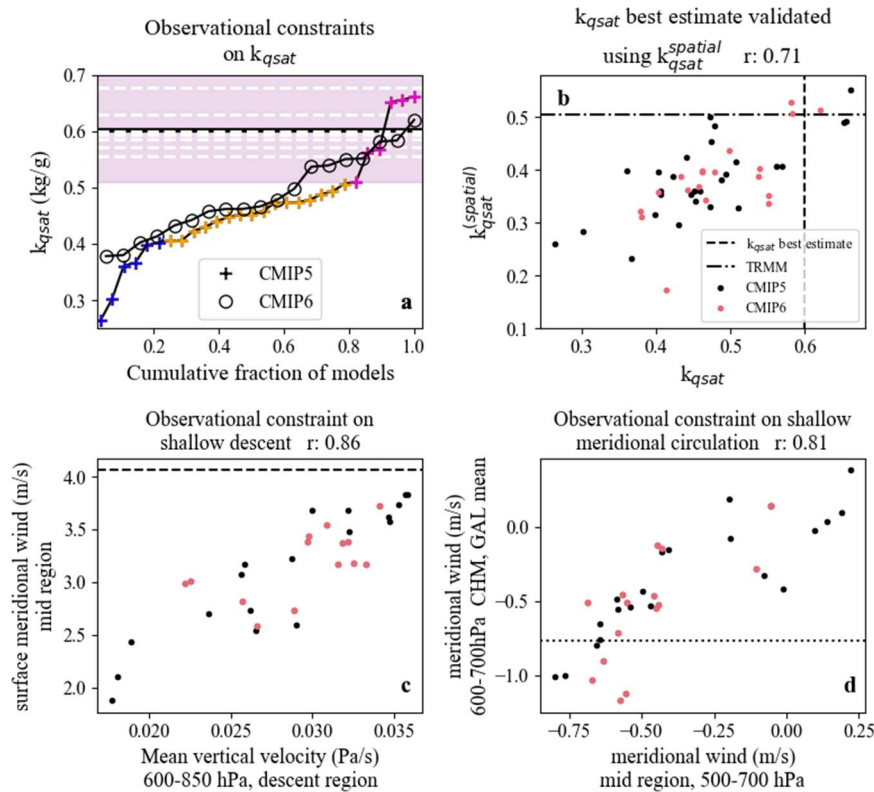
Figure 4. High sensitivity of precipitation to SST, and strong shallow circulations, in the real

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.

507
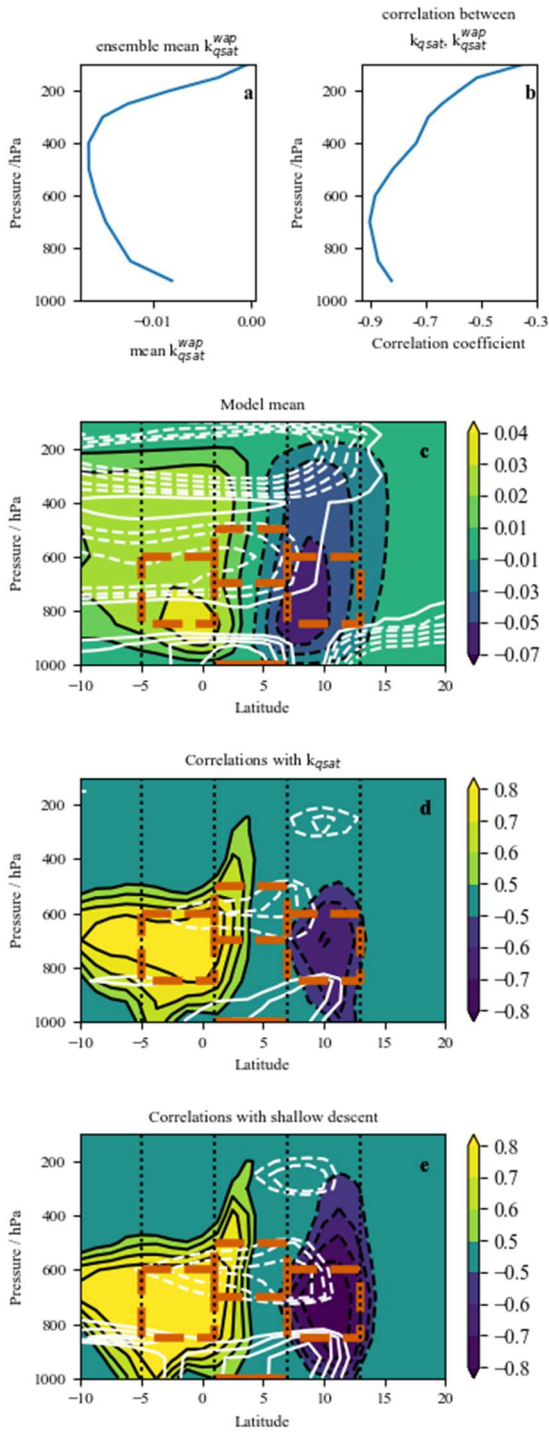
Figure 5. Linking k_{qsat} to shallow circulations. **a,b** quantifies internal variability, and **c-e**

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

513   wind. **d** inter-model correlations between k_{qsat} and mean vertical velocity (colours) and

514   between k_{qsat} and mean meridional wind (white contours). **e** as **d**, but for correlations with

515   the shallow descent index instead of k_{qsat} (shallow descent index = vertical velocity averaged

516   over left-hand orange-dashed box: 5S-1N, 850-600 hPa).

517

518

519

520

Methods

522

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

543   missing). Seasonal means were taken only where three consecutive months had non-missing

544   data.

545   Monthly mean SSTs are taken from four different datasets. The CMIP5 AMIP dataset[30] (the

546   dataset used to drive the AMIP SST-forced model runs) is available only for 1980-2005.

547   This was used to calculate k_{qsat} for each model. For our observational estimate of a lower

548   bound on k_{qsat} (using TRMM precipitation), three SST datasets were used: HadISST[32] version

549   1.1, ERSST[33] version 4 and COBE[34] version 2 (the AMIP dataset used by the model

550   simulations was not used, due to its limited temporal overlap with the TRMM operational

551   period).

552   Surface meridional wind observations are from SeaWinds on QuikSCAT Level 3, for the

553   period Aug 1999-Oct 2009. Wind profiler observations at San Cristóbal, Galápagos (0.9°S,

554   89.7°W, 1994-2005) and Christmas Island (2.8°N, 157.5°W, 1990-2002) used 915 MHz in

555   low mode, as used in other studies of shallow circulation[8].

556   **Model data**

557   All model results are from atmosphere-only AMIP runs (one run per model version) forced

558   by observed SST, corresponding to the period 1980-2005. This includes 28 models from

559   CMIP5 (ACCESS1-0, ACCESS1-3, BNU-ESM, CCSM4, CESM1-CAM5, CNRM-CM5,

560   CSIRO-Mk3-6-0, CanAM4, GISS-E2-R, HadGEM2-A, IPSL-CM5A-LR, IPSL-CM5B-LR,

561   MIROC-ESM, MIROC5, MRI-AGCM3-2H, MRI-AGCM3-2S, MRI-CGCM3, NorESM1-M,

562   inmcm4, bcc-csm1-1-m, bcc-csm1-1, EC-EARTH, MPI-ESM-LR, MPI-ESM-MR,

563   FGOALS-s2, FGOALS-g2, GFDL-CM3, GFDL-HIRAM-C180) and 19 from CMIP6 (BCC-

564   CSM2-MR, BCC-ESM1, CAMS-CSM1-0, CanESM5, CNRM-CM6-1, CNRM-ESM2-1,

565   EC-Earth3-Veg, INM-CM4-8, INM-CM5-0, IPSL-CM6A-LR, MIROC6, HadGEM3-GC31-

566   LL, UKESM1-0-LL, MRI-ESM2-0, GISS-E2-1-G, GFDL-AM4, GFDL-CM4, NESM3,

567   SAM0-UNICON).

568   Evaluation of satellite precipitation using GTMBA raingauge data

569   First, seasonal mean satellite precipitation data (for all seasons) were interpolated linearly to

570   the GTMBA locations. Logarithms of seasonal mean precipitation were then taken, and all

571   datasets masked at times and locations where any data (GTMBA or satellite) were missing.

572   This resulted in 1723 observations of seasonal mean precipitation from each dataset, covering

573   the period 1998-2015 (the overlap between TRMM and GTMBA operational periods). For

574   the remainder of the analysis, the observation location and time are ignored.

575   To reduce noise effectively, while retaining the signal of interest, we use the fact that

576   precipitation tends to increase with SST, but the noise (as defined here) is largely

577   independent of SST. For each dataset, the 1723 observations were grouped into 120 bins (14

578   observations per bin). This was done by ranking the observations by seasonal mean SST (the

579   14 observations corresponding to the 14 lowest SST values were placed in the first bin, and

580   so on). The mean across each bin was then taken, giving 120 bin means of log(seasonal

581   precipitation): giving 120 symbols in Figure 1. 120 bins were chosen, as a mean over 14

582   observations is sufficient to reduce noise significantly, while retaining a large number of

583   symbols in Figure 1 to assess the method visually. Doubling the number of bins has

584   negligible effect on the gradient in Figure 1. Results are insensitive to which SST dataset is

585   used to bin the data (compare Extended data Figure 4, bottom two rows, with Figure 1).

586   We demonstrate that regression dilution bias is likely to be small in Figure 1a (method

587   justification in Methods subsection 'Estimating regression dilution bias' below). We do this

588   by repeating the analysis in Figure 1a, but regressing first TRMM against GPCP (Extended

589   data Figure 4, top left), then GPCP against TRMM (Extended data Figure 4, top right). The

590   product of the two regression gradients is 0.98 (close to 1), suggesting that this bias is small.

591   We also tested sensitivity of the validation to potential undercatch by the GTMBA raingauges

592   in windy conditions[29]. This issue could only bias the gradients in Figures 1a,b if the

593   percentage undercatch varied systematically from low to high precipitation (because Figure 1

594   shows log precipitation). To test this, we recalculated the gradients in Figures 1a,b, but after

595   masking the data according to the seasonal mean wind speed (also observed by GTMBA

596   buoys). Gradients calculated for low wind (0-4 m/s; 22% of all data) and high wind (6-10

597   m/s; 27% of all data) show no significant differences from Figure 1a,b (for TRMM,

598   confidence intervals consistently spanning 1 and best estimate within 5% of 1; for GPCP,

599   confidence intervals consistently excluding 1). This suggests that the satellite validation is

600   insensitive to wind undercatch.

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

602   For Figure 2, all data is first regridded by area-averaging to a common grid (resolution: 1.25º

603   latitude by 1.875º longitude).

604   The El Niño/La Niña ratio is based on large ENSO episodes in years both simulated by the

605   models and observed by TRMM: the 1997-1998 El Niño divided by the mean of 1998-1999

606   and 1999-2000 La Niñas, for the El Niño peak season (November-January), averaged over

607   10S-10N.

608   Other data are zonal means over the following longitude bands (with land masked out):

609   (Atlantic) 70W-25E; (East Pacific) 150-100W; (Indian Ocean) 50-100E.

610   Seasonal cycles over Atlantic and Pacific are calculated as the zonal mean for August-

611   October divided by the zonal mean for February-April). These seasons were chosen because

612 they show large differences in SST, but small differences in solar zenith angle (the latter can

613 affect precipitation by altering land temperature).

614 Spatial patterns are calculated, for each model/observational dataset, as the zonal mean at

615 each latitude, divided by the maximum, zonal mean for the same model/observational dataset.

616 This was calculated for August-October (ASO) for the Atlantic, due to the large meridional

617 SST gradient for this season. November-April was used for the Indian Ocean, as this basin

618 has a significant meridional SST gradient for this period.

Estimating k_{qsat}, part 1: data preparation

620 k_{qsat} is used here specifically to rank the models and compare with observations. Therefore,

621 the method of calculation needs to be consistent across models and observations, and to

622 minimise the potential for observational error.

623 For each year, for a given season, the logarithm of seasonal mean precipitation is calculated.

624 The spatial pattern of climatological mean precipitation (Extended data Figure 2j) is

625 dominated by a small area of large precipitation (occupying around 10% of the area). Hence,

626 if we evaluated k_{qsat} without taking the logarithm, our result would be dominated by this

627 small area of the tropics. The spatial distribution of log(precipitation) is much more uniform

628 (Extended data Figure 2k), except for the driest 10% of the tropics (which is eliminated from

629 our analysis as we mask the region of coolest SST). Hence, using log(precipitation) to

630 calculate k_{qsat} ensures that the result is influenced fairly equally by all parts of our analysis

631 region (confirmed in Figure 3a). Our results use seasonal mean precipitation. Use of other

632 timescales would alter k_{qsat}, due to the (nonlinear) logarithm in Equation 1.

633 For models, k_{qsat} is calculated using data on each model's native grid. For observations, q_{sat}

634 is regridded linearly to the high resolution TRMM horizontal grid.

635   Before estimating k_{qsat}, to minimise observational error, we exclude the 30% of the tropical

636   oceans with the lowest climatological mean SST. An advantage of using a logarithm in

637   equation 1 is that k_{qsat} estimates are not dominated by the narrow ITCZ region. However, it

638   could mean that error in observing the very lowest rainfall rates could cause large error in our

639   real-world estimate of k_{qsat}. Therefore we mask the regions with coolest SST on average.

640   This masking is only done in calculating k_{qsat}. It is not done in Figure 2, as the climatological

641   means reduce observational error there.

642   At each location, anomalies relative to climatological means are calculated for each year, for

643   both q_{sat} and log(precipitation). Locations that have missing data in any year are excluded.

644   These data are used in the sortav method, described below.

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

646   Once the data is prepared as above, our regression method for estimating k_{qsat} (denoted

647   'sortav') is applied. For python code for this method, and an illustrative example, see Code

648   Availability Statement. The sortav method is designed to prevent dominance from the SST

649   spatial pattern associated with ENSO (an issue because ENSO features large SST anomalies

650   in a consistent pattern). In estimating k_{qsat} from inter-annual variability, our aim is to reduce

651   the 'other-processes' term in Equation 1, by averaging over different SST patterns, with

652   different patterns of large-scale circulation anomalies. If a single SST pattern (ENSO) was

653   allowed to dominate, this would not be effective. If standard linear regression was used, the

654   ENSO pattern would dominate, because ENSO features large SST anomalies. In addition,

655   tropical means of precipitation and SST can vary over time, involving different processes

656   than those represented by k_{qsat}. Our method avoids these issues.

657   The first step sorts each year of data. Say the data has n locations and y years. For each year,

658   the n anomalies in log(precipitation) are sorted in order of increasing q_{sat} anomaly. This

659   produces, for each year, a vector of length n, with the first element corresponding to the

660   location with the most negative q_{sat} anomaly, and the last element being that with the most

661   positive q_{sat} anomaly. If Equation 1 was exactly true, with zero noise, each vector would be

662   sorted in order of increasing anomaly in log(precipitation). Because of the noise (from large-

663   scale processes), this is not, in general, true. This gives y sorted vectors, each of length n.

664   We then average over the y years, to produce one mean vector of length n (e.g. the 1st

665   element of this vector is the mean of the y precipitation anomalies found over the most

666   negative q_{sat} anomaly from each year). This averaging removes much of the noise, because

667   the noise is largely independent of q_{sat}. In this mean, years with large SST anomalies have

668   the same weighting as other years, avoiding dominance by ENSO.

669   The same process is repeated for q_{sat}. This gives two mean vectors, each of length n: for

670   anomalies in log(precipitation) and in q_{sat}.

671   The relationship between the averaged anomalies in log(precipitation) and q_{sat} is relatively

672   linear in both models and observations (e.g. Extended data Figure 2a-d), suggesting that

673   Equation 1 is a useful approximation in this context. k_{qsat} is then estimated from these vectors,

674   by ordinary least squares linear regression (e.g. the gradients of the best fit lines in Extended

675   data Figure 2a-d).

676   To compare with our sortav method, alternative estimates of k_{qsat} (marked OLS in Extended

677   Data Figures 2h,7) use standard linear regression between seasonal anomalies in

678   log(precipitation) and q_{sat} (without sort-averaging).

679

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

681      $k_{qsat}{}^{(spatial)}$ is calculated using the same method as k_{qsat}, except that it quantifies seasonal and

682   spatial variation in time-mean climate (in contrast with k_{qsat}, which quantifies interannual

683   variability). First, time means are taken for each dataset and season (giving 4 season means

684   per grid point per dataset). For each dataset and each season separately, precipitation is

685   divided by the tropical mean, before taking logarithms. This is done to scale out model

686   variation in the tropical mean (which is controlled by the large-scale energy budget). For q_{sat},

687   for each dataset and season separately, anomalies are taken with respect to the tropical mean.

688   Masking, to exclude regions with low q_{sat}, is based on annual mean q_{sat}.

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

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

691   This approximates variation in precipitation relative to the tropical mean, driven by variation

692   in q_{sat} relative to its tropical mean. The overbar represents the tropical mean for the current

693   season of the current year, and q_{sat}' is specifically the q_{sat} anomaly with respect to the

694   tropical mean:

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

696   Anomalies expressed this way capture temporal and spatial variability associated with

697   variation in local SSTs, but exclude temporal variability in tropical mean precipitation (which

698   is constrained by the large-scale atmospheric energy budget).

699      $k_{qsat}^{spattemp}$ is estimated using the same approach as kqsat, but anomalies of log(P) and qsat

700   are taken relative to their tropical means for the corresponding season and year (kqsat is

701   evaluated using anomalies with respect to local climatological means for each location).

702   A disadvantage of both $k_{qsat}^{(spatial)}$ and $k_{qsat}^{spattemp}$ is that they have some sensitivity to other

703   processes (teleconnections) associated with the specific spatial patterns in climatological

704   mean SST (these spatial patterns are filtered out less effectively as there are only 4 seasons,

705   compared to the 25 years of internal variability used to estimate kqsat).

706

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

708   This method (estimated lower bound on k_{qsat}) accounts for three forms of observational error,

709   combined using Monte Carlo sampling.

710   First, systematic error in the observed magnitudes of seasonal mean SST anomalies could

711   bias k_{qsat}. To explore this, we first estimated k_{qsat} (with the sortav method) using TRMM

712   precipitation, and each of the three SST datasets (HadISST, ERSST, COBE) that cover the

713   whole TRMM operational period, giving three direct, unscaled estimates of k_{qsat} (Extended

714   data Table 2a).

715   Second, error in the SST spatial pattern will cause a low bias in k_{qsat} (regression dilution

716   bias). Typical magnitudes of this bias are estimated (see more detail in Methods section

717   'Estimating regression dilution bias'). This is done by regressing pairs of SST datasets

718   against each other, with regression coefficients calculated using the sortav method (as used

719   for k_{qsat}). Extended data Table 2b shows the results for each pair of SST datasets. The final

720   column of Extended data Table 2b gives the 6 different estimates of regression dilution bias.

721   Third, the TRMM operational period does not fully overlap the AMIP SST-forced model

722   simulation period. Thus, k_{qsat} estimated for the TRMM period will be different from that

723   obtained if TRMM were operational throughout the AMIP period (due to a different set of

724   SST patterns during the TRMM and AMIP period). Estimates of typical magnitudes of this

725   error were obtained using samples from coupled ocean-atmosphere simulations (Extended

726   data Figure 7), giving 85 samples of percentage error.

727   10000 estimates of k_{qsat} were then generated using Monte Carlo sampling. For each estimate,

728   one of the three direct estimates (Extended data Table 2a) was selected at random, 'corrected'

729   by a random selection from the 85 samples of percentage error (Extended data Figure 7), and

730   further corrected using a random selection from the 6 estimates of regression dilution bias

731   (Extended data Table 2b).

732   The lower bound of the 95% confidence interval of these 10000 estimates is then taken. We

733   only use the lower bound for the following reason: it seems unlikely that the true regression

734   dilution bias is weaker than the minimum value estimated here (around 10%, Extended data

735   Table 2b). This is because we do not expect the SST datasets to be significantly closer to the

736   real SST than they are to each other. However, it is plausible that the regression dilution bias

737   could be larger than estimated, so we do not quote an upper bound for k_{qsat}.

738   The appropriate lower bound for an atmosphere-only AMIP model may be even higher than

739   our quoted result, as the observed k_{qsat} value may be reduced by the effect of atmospheric

740   internal variability on SST[19]. kqsat values from AMIP models are similar to or higher than

741   from the equivalent coupled models. In contrast, this coupling issue should be small for our

742   central estimate of k_{qsat} (next section), as that is based on metrics where ocean dynamics (for

743   ENSO) or forcing associated with the mean state/seasonal cycle dominate the SST

744   differences/gradients.

745   **Central estimate of k_{qsat} in the real world**

746   We start by finding where, geographically, the models are most sensitive to k_{qsat}. This is

747   done using correlation coefficients (r) between k_{qsat} and the log of precipitation ratios

748   (Extended data Figure 5) for each latitude (or longitude for ENSO) in each panel in Figure

749   2. This reveals six discrete intervals (shaded in Figure 2) where $|r| > 0.6$ (about 50% variance

750   explained), and a seventh, showing weaker correlations, over the Indian Ocean. For each

751   model, we average the log of precipitation ratios over each interval, and take the difference

752   from the equivalent value for TRMM. This gives seven error indices for each model (y-axes

753   in Extended data Figure 6). For each interval, the 28 model error indices are regressed

754   against k_{qsat} (x-axes in Extended data Figure 6), and k_{qsat} is estimated from where the line of

755   best fit crosses zero error. This gives seven estimates of k_{qsat}; their mean is our central

756   estimate.

757   We did not use the coupled ocean-atmosphere models to estimate k_{qsat}, due to the evident

758   residual biases from SST error in these models, and because it isn't possible to use the ENSO

759   response for coupled models, due to model differences in simulations of ENSO SST

760   responses.

761   **Estimating regression dilution bias**

762   Regression dilution bias[51] arises when there is random error in the independent variable (e.g.

763   q_{sat} in estimating k_{qsat}). This causes the regression gradient to be biased low. This bias

764   reduces the gradient by a factor (β) that depends only on the characteristics of the

765   independent variable. We estimate typical magnitudes of this bias using different

766   observations of the independent variable (e.g. different SST datasets), as follows.

767   Say the vector of true values of the independent variable is **x**, and we have two different

768   observational estimates, **x1** and **x2**. We first regress **x1** against **x2**. This regression uses the

769   same methodology as when regressing the dependent variable against **x**: i.e. for estimating

770   dilution bias in observations of k_{qsat}, the sortav method is applied; and for the satellite

771   precipitation validation, the SST binning is used. The regression gradient (f_{12}) obtained from

772 regressing **x1** against **x2**, will be biased low by a factor β_2 (from error in **x2**). We then

773 regress **x2** against **x1**. This regression gradient (f_{21}) will be biased low by a factor β_1 (from

774 error in **x1**).

775 We then estimate the dilution bias as:

776 $\beta \approx$ sqrt($\beta_2 \cdot \beta_1$) = sqrt($f_{12} \cdot f_{21}$)

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

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

779 **Column dry static energy budget (DSE)**

780 Dry static energy (DSE) is given by $s = c_p T + gz$, where c_p is the specific heat at constant

781 pressure, T is the temperature, g is the gravitational acceleration, and z is altitude). DSE is

782 affected by advection, precipitation, radiation and sensible heat:

783 $$-\langle \omega \frac{\partial s}{\partial p} \rangle - \langle v. \nabla s \rangle + LP + R + Q_{turb} = 0,$$

784 where ω is the vertical pressure velocity, p is pressure, v is horizontal wind. Angle brackets

785 represent the mass-weighted vertical integral from 1000-100hPa. The first term represents

786 import of DSE via column-integrated vertical advection; the second is horizontal advection.

787 P is the total surface precipitation and L the latent heat of condensation; R is net radiation

788 into the atmospheric column; and Q_{turb} is the surface sensible heat flux.

789

790 **Methods References**

791

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

808

809

810
811

End Notes

813

833   CNRM model simulations. P.G., R.C., C.E.H. and R.R. contributed understanding on physical

834   processes. J.K. provided knowledge of SST observational uncertainty and datasets.
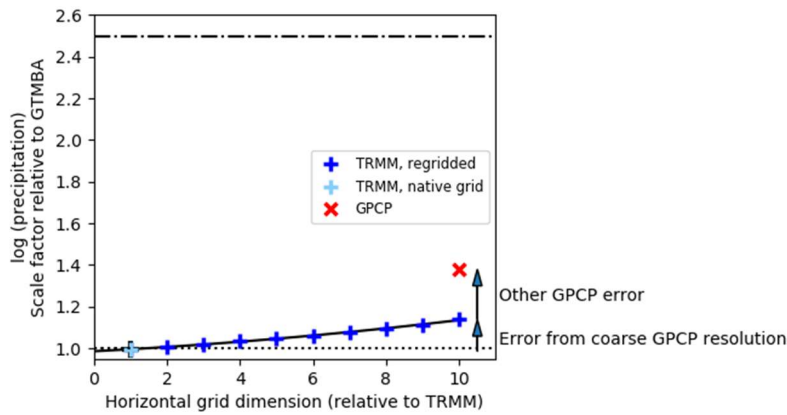
838

839

840

841

842

843

Extended Data Figures

845

846

847 Extended data Figure 1. Effect of low spatial resolution in GPCP satellite observations of

848 log(seasonal precipitation). y-axis: regression gradient in validation against GTMBA

849 raingauge data (i.e. gradients in Figure 1 for light blue and red symbols). x-axis: horizontal

850 grid dimension relative to TRMM (e.g. the TRMM resolution is 0.25°, ten times smaller than

851 the GPCP resolution of 2.5°, so the red symbol is placed at x=10). Dark blue symbols:

852 results when TRMM data is regridded (by area averaging) to coarser grids. The coarser grids

853 are chosen so the grid box edges overlap edges of the native TRMM grid. To give the errors

854 context, the dash-dot line marks the ratio between the largest and smallest model values of

855 k_{qsat} (2.5). Solid black line is a quadratic least-squares best fit line through the TRMM-based

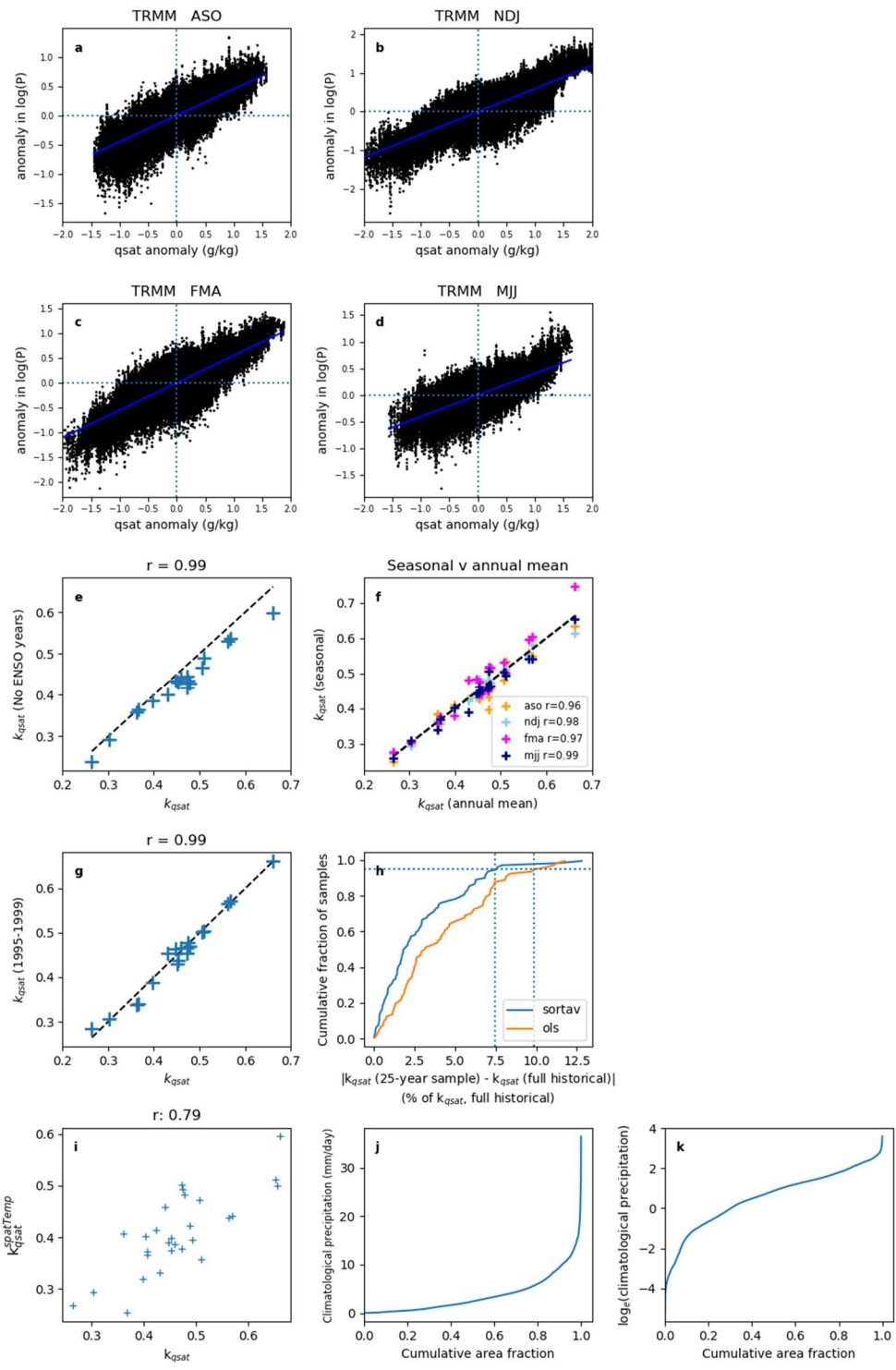856 data. The intercept of the TRMM best-fit curve at x=0 (i.e. infinitely fine grid) is very close

857 to the value estimated on the TRMM native grid (light blue symbol), indicating that the

858 TRMM grid is sufficiently fine for comparison with the rangauge data on seasonal

859 timescales.

860

861

862

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

867   years with the largest absolute value of the nino3.4 index; x-axis: default k_{qsat} (one symbol

868   per model); k_{qsat} is on average 6% lower when ENSO years excluded, due to a small

869   sensitivity to the ENSO characteristic spatial pattern; but the model ranking is largely

870   unchanged (r = 0.99). **f,** k_{qsat} calculated for individual seasons versus the annual mean value;

871   **g** k_{qsat} using only years 1995-1999 versus the full 25-year estimate; **h,** estimating variability

872   (due to internal variability in SST patterns) in k_{qsat} estimated from 25 years of data: for each

873   coupled ocean-atmosphere model, k_{qsat} is estimated both for the full historical run, and for all

874   25-year chunks. Panel shows the cumulative distribution function of absolute percentage

875   differences between the 25-year estimates and the full estimates (95% of samples are within

876   8% of the long-term value from the full historical run). This panel shows results for two

877   methods of estimating k_{qsat}: our 'sortav' method (as used throughout the manuscript), and

878   standard OLS regression between seasonal anomalies in log(precipitation) and q_{sat}. **i**

879   comparing $k_{qsat}^{spattemp}$ with k_{qsat}; each cross represents one CMIP5 model. **j,k** Cumulative

880   distribution functions of **j** climatological mean precipitation and **k** log(precipitation). From

881   HadGEM2-A, May-July season (same picture seen in other seasons).

882

883

884
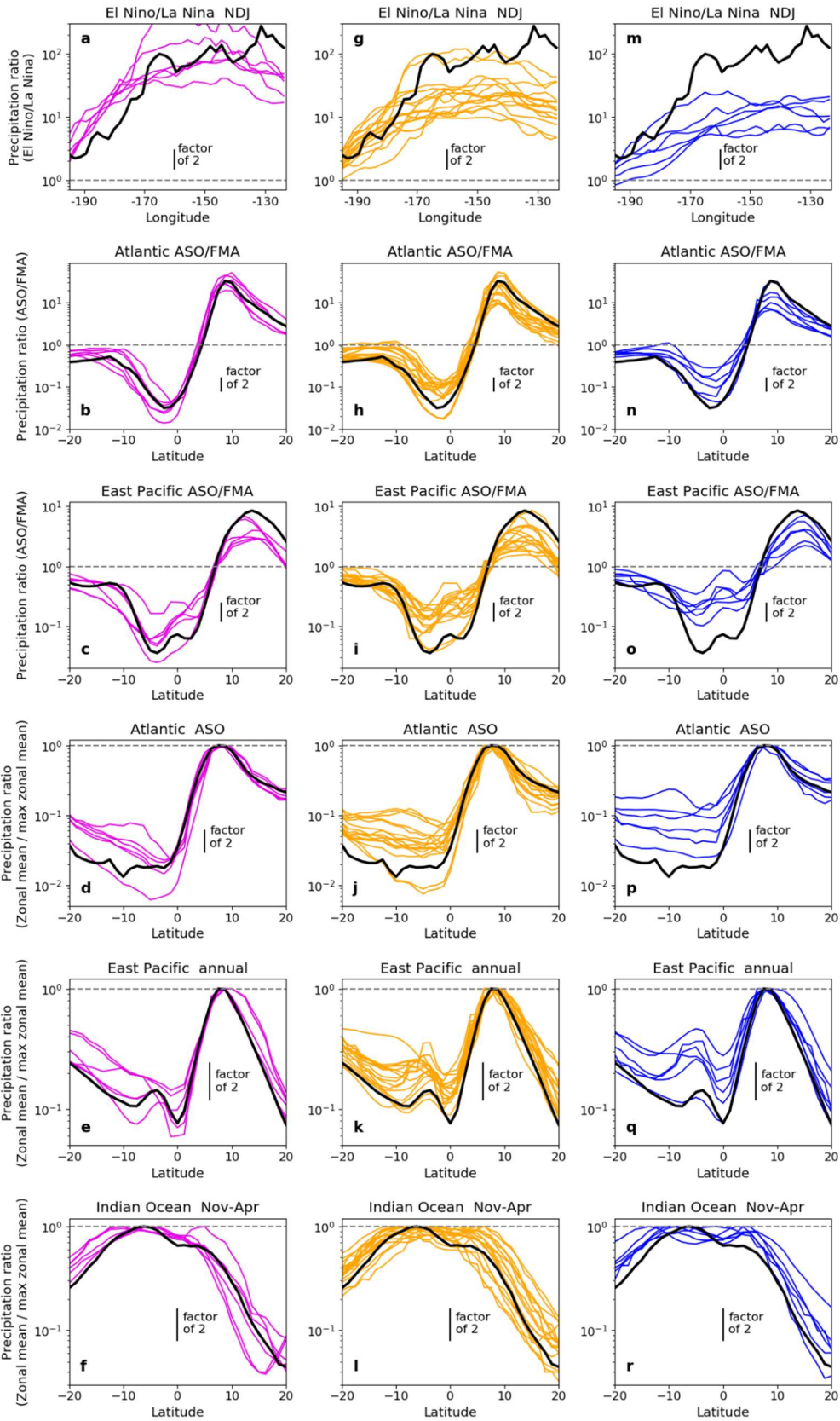
Extended data Figure 3. Model biases for the high, mid-range and low-k_{qsat} models

separately. As Figure 2, for **a-f** high-k_{qsat} models; **g-l** mid-k_{qsat} models; **m-r** low-k_{qsat} models.

887

888   Extended data Figure 4.   Testing potential errors in the satellite validation against GTMBA.

889   **a,b** testing for regression dilution bias from error in TRMM observations: as Figure 1, but for
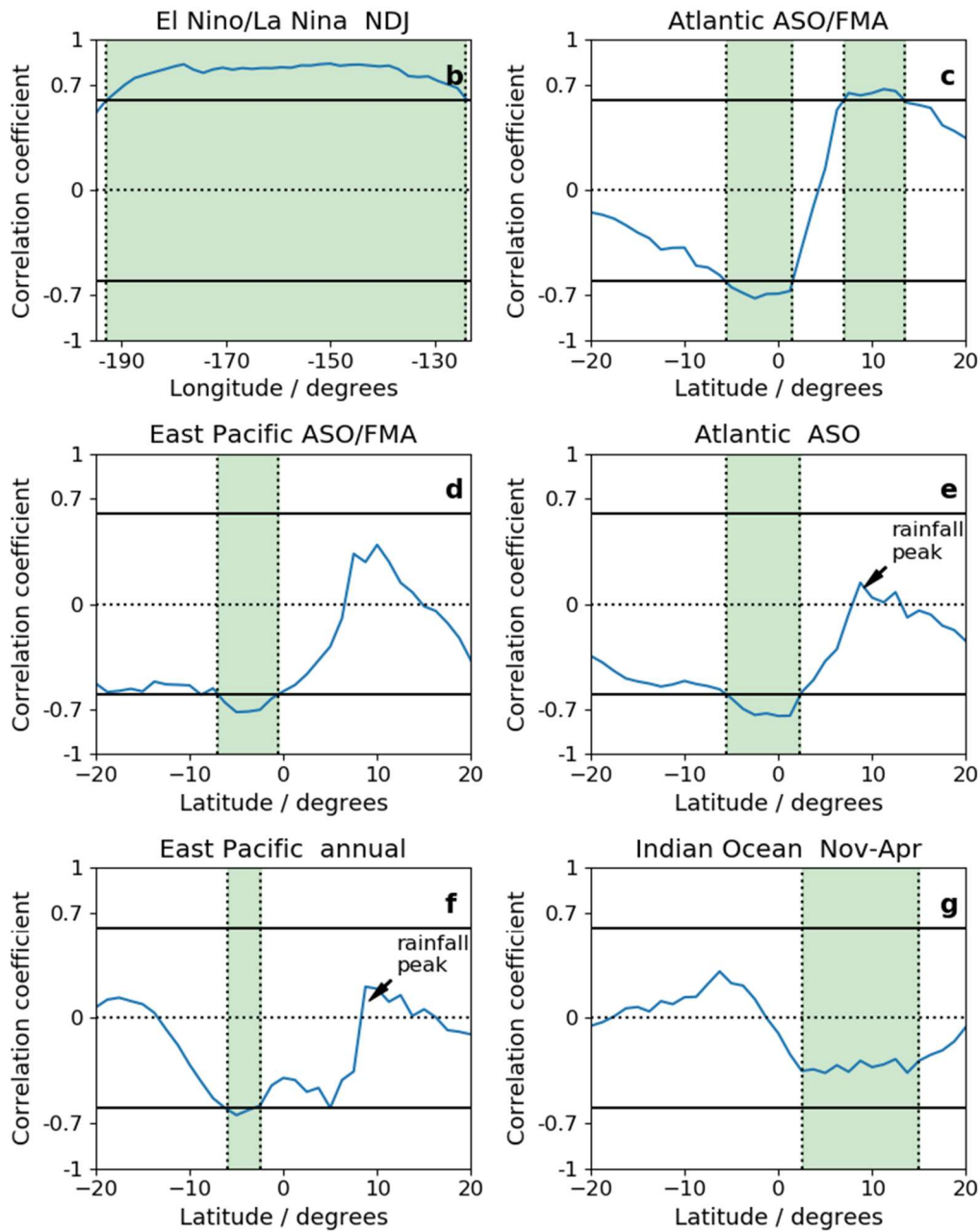
890   **a** TRMM versus GPCP (both interpolated to GTMBA sites and masked as in Figure 1) and **b**

891   GPCP versus TRMM. **c-f** testing for effects of SST uncertainty on the binning: as Figure 1,

892   but using **c,d** ERSST and **e,f** COBE SST datasets to bin the data.

893

894

895

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

903   to the latitude of the precipitation peak in most models (the model spread in the precipitation

904   peak is scaled out; coefficients are not exactly zero as there is a small model spread in the

905   latitude of the precipitation peak).

906

907

908
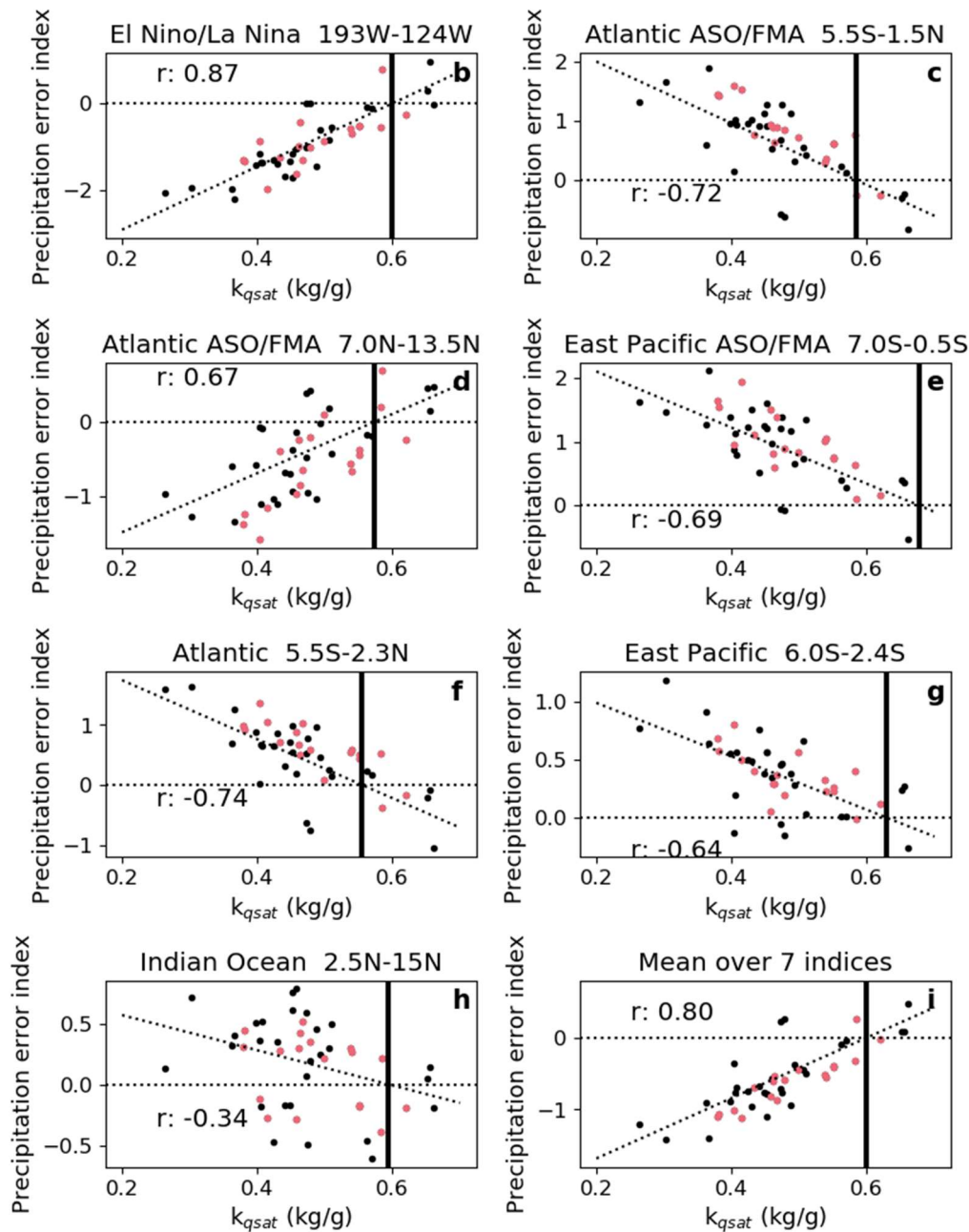
Extended data Figure 6. Scatter plots underpinning the central observational estimate of k_{qsat}.

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

916   the 7 indices in the other panels (after the signs of the 5 indices with negative best-fit slopes

917   were changed, to ensure a positive correlation with k_{qsat}).

918

919
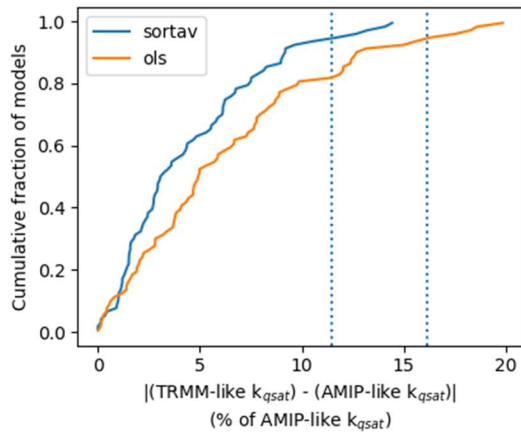
920

921   Extended data Figure 7. Supporting results for observational estimate of the k_{qsat} lower

922   bound. Estimating error, from internal variability, due to the fact that the TRMM operational

923   period only partly overlaps the time period simulated by the AMIP SST-forced models. Error

924   magnitudes are estimated from the coupled ocean-atmosphere simulations, using differences

925   between k_{qsat} estimated from all possible overlapping 17-year (TRMM-like) and 25-year

926   (AMIP-like) periods (with the same overlap as TRMM and the 25-year SST-forced model

927   simulations). Results are given for two methods of estimating k_{qsat}: our 'sortav' method (as

928   used throughout the manuscript), and standard OLS regression between seasonal anomalies in

929   log(precipitation) and q_{sat}.

930

931

932
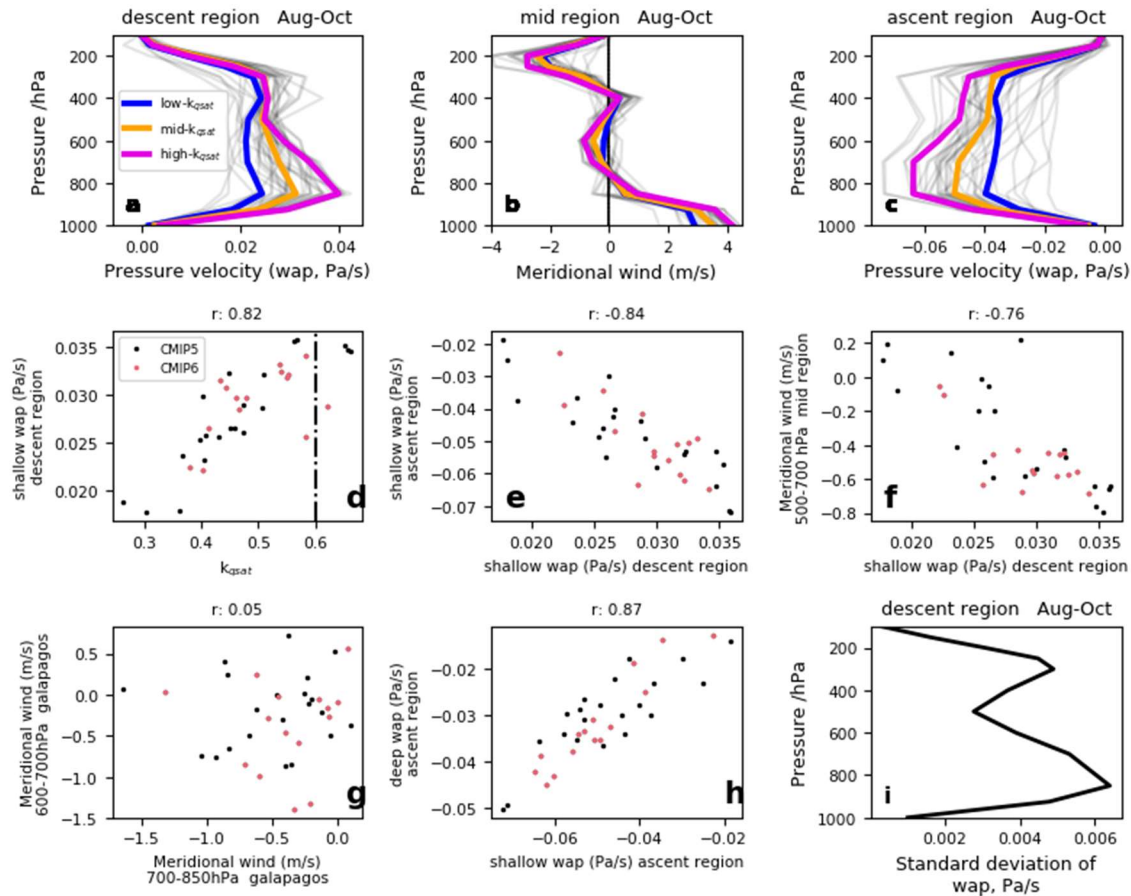
Extended Data Figure 8. Atmospheric circulation measures in CMIP5 and CMIP6 models.

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.

944

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

957

958   Extended data Table 2. Supporting results for observational estimate of the k_{qsat} lower bound

959   (see Methods for details). **a** Unscaled estimates for k_{qsat} directly estimated using TRMM

960   precipitation and three different SST datasets (the AMIP SST dataset was not used due to

961   limited temporal overlap with the TRMM operational period). These values are

962   contaminated by regression dilution bias so do not represent central estimates. **b** Estimating

963   typical values of regression dilution bias from each pair of SST datasets. K1 is the gradient

964   from linear regression when regressing SST1 against SST2 (using the sortav regression

965   method). K2 is the value obtained when regressing SST2 against SST1.

966