Strong cloud-circulation coupling explains weak trade cumulus feedback

Raphaela Vogel[1], Anna Lea Albright[2], Jessica Vial[2], Geet George[3], Bjorn Stevens[3], and Sandrine Bony[2]

[1]LMD/IPSL, Sorbonne Université, now at: Universität Hamburg
[2]LMD/IPSL, Sorbonne Université, CNRS
[3]Max Planck Institute for Meteorology

November 26, 2022

Abstract

Shallow cumulus clouds in the trade-wind regions cool the planet by reflecting solar radiation. The response of trade cumulus clouds to climate change is a major uncertainty in climate projections. Trade cumulus feedbacks in climate models are governed by changes in cloud fraction near cloud base, with high climate-sensitivity models suggesting a strong decrease in cloud-base cloudiness due to increased lower-tropospheric mixing. Here we show that novel observations from the EUREC4A field campaign refute this mixing-desiccation hypothesis. We find the dynamical increase of cloudiness through mixing to overwhelm the thermodynamic control through humidity. Because mesoscale motions and the entrainment rate contribute equally to variability in mixing, but have opposing effects on humidity, mixing does not desiccate clouds. The magnitude, variability, and coupling of mixing and cloudiness differ drastically among climate models and with the EUREC4A observations. Models with large trade cumulus feedbacks tend to exaggerate the dependence of cloudiness on relative humidity as opposed to mixing, and also exaggerate variability in cloudiness. Our observational analyses render models with large positive feedbacks implausible, and both support and explain at the process scale a weak trade cumulus feedback. Our findings thus refute an important line of evidence for a high climate sensitivity.

Strong cloud-circulation coupling explains weak trade cumulus feedback

Raphaela Vogel[1,2*], Anna Lea Albright[1], Jessica Vial[1],
Geet George[3] , Bjorn Stevens[3], Sandrine Bony[1]

[1]LMD/IPSL, Sorbonne Université, CNRS, Paris, France
[2]now at: Universität Hamburg, Hamburg, Germany
[3]Max Planck Institute for Meteorology, Hamburg, Germany

[*]To whom correspondence should be addressed; E-mail: raphaela.vogel@uni-hamburg.de

August 30, 2022

Shallow cumulus clouds in the trade-wind regions cool the planet by reflecting solar radiation. The response of trade cumulus clouds to climate change is a major uncertainty in climate projections[1–4]. Trade cumulus feedbacks in climate models are governed by changes in cloud fraction near cloud base[5,6], with high climate-sensitivity models suggesting a strong decrease in cloud-base cloudiness due to increased lower-tropospheric mixing[5–7]. Here we show that novel observations from the EUREC^{4}A field campaign[8,9] refute this *mixing-desiccation* hypothesis. We find the dynamical increase of cloudiness through mixing to overwhelm the thermodynamic control through humidity. Because mesoscale motions and the entrainment rate contribute equally to variability in mixing, but have opposing effects on humidity, mixing does not desiccate clouds. The magnitude, variability, and coupling of mixing and cloudiness differ drastically among climate models and with the EUREC^{4}A observations. Models with large trade cumulus feedbacks tend to exaggerate the dependence of cloudiness on relative humidity as opposed to mixing, and also exaggerate variability in cloudiness. Our observational analyses render models with large positive feedbacks implausible, and both support and explain at the process scale a weak trade cumulus feedback. Our findings thus refute an important line of evidence for a high climate sensitivity[10,11].

Introduction

Earth's climate strongly depends on the abundance and behavior of its smallest clouds. Shallow trade-wind cumulus clouds are rooted in the turbulent sub-cloud layer and form when thermals rise above the lifting condensation level[12]. They may grow only a few 100 m high in dry environments, or become positively buoyant and rise up to the trade-wind inversion, where they detrain condensate into stratiform cloud layers. Trade cumuli populate the majority of subtropical oceans and cool the planet by reflecting the incoming solar radiation. Due to their large geographical extent, small errors in predicting the way trade cumuli respond to warming can have a large effect on the global radiative budget. This explains why shallow cumuli in the trades are a major source of spread in climate models' estimates of climate sensitivity[1–4].

Cloudiness near the base of the cumulus layer makes up two-thirds of the total cloud cover in the trades[13] and its change with warming governs the strength of the trade cumulus cloud feedback in climate models[5,6]. Reductions in cloud-base cloudiness in climate models are tightly coupled with increases in lower-tropospheric mixing due to convective and large-scale circulations[5–7]. Based on this strong negative coupling between mixing and cloudiness, the hypothesis emerged that enhanced convective *mixing* might *desiccate* the lower cloud layer and reduce cloudiness in the trades[7]. This *mixing-desiccation* hypothesis suggests that the moisture transported by convection from the sub-cloud layer to the trade inversion is compensated by downward mixing of drier air and evaporation of clouds near cloud base. The mechanism—which is expected to become more pronounced with warming due to the nonlinear Clausius-Clapeyron relationship—is consistent with idealized high-resolution simulations of non-precipitating trade cumuli[14] and with the behavior of climate models that have a strongly positive trade cumulus feedback[5,7,15]. However, the mixing-desiccation hypothesis has never been tested with observations. Using the convective mass flux at cloud base, M, as a proxy for lower-tropospheric convective mixing, the hypothesis can be tested by analyzing the relationship between M and the mean relative humidity ($\mathcal{R}$) and cloud fraction (C) at cloud base in observations, with $C \propto \mathcal{R} \propto M^{\beta}$ and $\beta < 0$ suggesting the mixing-desiccation mechanism to be present in nature (Fig. 1a).

The mixing-desiccation mechanism is based on a number of assumptions that might not be operating in nature. M is commonly defined as the product of the cloud fraction and the in-cloud vertical velocity, and its variability is mostly governed by the area coverage of active clouds[16,17], defined as saturated and buoyant updrafts that ventilate the sub-cloud layer. If variability in the in-cloud vertical velocity near cloud-base is small, a positive relationship between C and M is expected ($\beta > 0$, Fig. 1b). This was demonstrated for non-precipitating trade cumuli using Doppler radar data[17,18] and appears at odds with the mixing-desiccation hypothesis. Yet active clouds represent only half of the total C[19,20] and the lifetime and variability of passive clouds, such as the detritus of decaying clouds, might be more sensitive to $\mathcal{R}$ and mixing-induced drying of their environment than to M.

The sub-cloud layer mass budget provides a theoretical basis for interpreting the mixing-desiccation mechanism. It can be expressed as a budget of the sub-cloud layer height h,

$$\frac{\partial h}{\partial t} + V_h \cdot \nabla h = E + W - M, \tag{1}$$

where the entrainment rate, E, representing the mass source due to the entrainment of dry and warm cloud layer air, and the mesoscale vertical velocity, W, are balanced by the mass export due to the convective mass flux, M[20]. Note that we define M as the (mass) specific mass flux, which has units of velocity (see Methods). E is the only term directly affecting the sub-cloud layer moisture and heat budgets[21,22]. If an increase in M is mostly balanced by an increase in E, a drying and warming of the sub-cloud layer and a reduction in $\mathcal{R}$ and C is expected (Fig. 1a). The trades, however, exhibit strong mesoscale convective organization, which is linked to the presence of mesoscale circulations and substantial variability in W[20,23–25]. This variability in W could contribute to variability in M without directly affecting $\mathcal{R}$ (Fig. 1b). An increase in M could also produce increased inversion cloudiness and thus increased total cloud cover, compensating the radiative effects of a potential decrease in C. The diversity of cloud types and the large variability in W in the trades thus call into question the mixing-desiccation mechanism as the dominant control of C and trade cumulus feedbacks.

The EUREC^{4}A (*Elucidating the role of clouds-circulation coupling in climate*) field campaign was conceived to test the mixing-desiccation hypothesis[8,9]. EUREC^{4}A took place in January and February 2020 near Barbados, a region selected as a source of data because clouds in its vicinity are representative for the entire trade-wind belt[26]. During EUREC^{4}A we made measurements designed to quantify the magnitude and (co-)variability of M, C, and $\mathcal{R}$ over one month, which was characterized by substantial variability in the mesoscale convective organization[27] and the large-scale circulation[9] (see Methods). With the help of these measurements, we are able to test the mixing-desiccation hypothesis with observations for the first time.

Observations of M, C, and $\mathcal{R}$ co-variations

During EUREC^{4}A we dropped more than 800 dropsondes from the HALO aircraft flying at about 10 km altitude along 1 h circles of 220 km diameter[28,29]. We use the dropsonde data to estimate M at the sub-cloud layer top as a residual of the mass budget (Eq. 1) on the 3 h-scale of 3 consecutive circles (see Methods). Fig. 2a shows a large day-to-day variability of M, with higher values at the beginning and end of the campaign, and a campaign-mean of $17.4 \pm 7.5\,\mathrm{mm\,s^{-1}}$ (mean $\pm$ standard deviation σ). M shows a pronounced diurnal cycle (Extended Data Fig. 1), with larger values around sunrise and smaller values in the afternoon (consistent with[20,30]). The mass budget estimates are robust to changes in the estimation procedure and consistent with independent data (Methods and Extended Data Fig. 2).

The entrainment rate E is computed as the ratio of the scaled surface buoyancy flux and the buoyancy-jump across h (Eq. S1, Extended Data Fig. 3). E averages to $18.3 \pm 6.4\,\mathrm{mm\,s^{-1}}$ across the campaign (Fig. 2b) and also shows a pronounced diurnal variability (Extended Data Fig. 1). E is mostly controlled by variability in the surface buoyancy flux (Extended Data Fig. 4b). It is the strengthening of winds and surface fluxes that contributes most to the increase in E and M in the second half of EUREC^{4}A. W is, with $-0.9 \pm 6.7\,\mathrm{mm\,s^{-1}}$, on average nearly zero. Variability in W, however, is substantial and contributes slightly more to variability in M compared to E (Extended Data Fig. 4a). So while $M \sim E$ holds on average, consistent with the mixing-desiccation hypothesis (Fig. 1a), variability in M is both controlled by E and W.

Fig. 2c shows the novel measurements of the cloud-base cloud fraction C from combined horizontally-staring lidar and radar on board the ATR aircraft flying near cloud base[31]. C is, with $5.4 \pm 3.1\%$, both small and highly variable. The variability of C on the 3 h-scale is substantially larger than variability on synoptic and longer timescales[13]. The robustness of C is demonstrated by the internal consistency among complementary and independent measurements in terms of measurement techniques and spatial sampling[31]. The $\mathcal{R}$ at cloud base is robustly around 86% (Fig. 2d), except for a few outliers. Three data points with much lower $\mathcal{R}$ for ATR compared to HALO (marked with *x* in Fig. 2d) are excluded in the following analyses, as these situations were associated with air masses that were sampled differently by the two aircraft (see Methods and Fig. A2 in ref[31]).

Despite being fundamental quantities to understand climate sensitivity, the challenging nature of observing M and C so far prevented an observational analysis of the relationship between mixing and cloud-base cloudiness. With the EUREC^{4}A observations presented here, we are now able to test the mixing-desiccation hypothesis with data.

Data refute mixing-desiccation hypothesis

The cloud-base cloud fraction is hypothesized to be controlled both dynamically through M and thermodynamically through $\mathcal{R}$. We can therefore express C as a multiple linear regression $\widehat{C} = a_0 + a_{\mathrm{M}}\widetilde{M} + a_{\mathcal{R}}\widetilde{\mathcal{R}}$, where $\widetilde{(\,)}$ represents standardized values (e.g., $\widetilde{M} = M/\sigma_{\mathrm{M}}$). Fig. 3a shows that the observed C and the reconstructed $\widehat{C}$ agree very well (r=0.83 [0.80, 0.91], with values in the square brackets representing the 25th and 75th quartile of bootstrapped correlations), demonstrating that M and $\mathcal{R}$ dominate variability in C.

The mixing-desiccation mechanism contends that as M increases, E increases and leads to a reduction in $\mathcal{R}$. The anti-correlation of E and $\mathcal{R}$ is confirmed by the observations ($\mathrm{r}_{E,\mathcal{R}} = -0.47$ [-0.62, -0.32], Extended Data Fig. 4d). But W is also correlated to $\mathcal{R}$ ($\mathrm{r}_{W,\mathcal{R}}$ = 0.48 [0.29, 0.62], Extended Data Fig. 4e). W does not directly affect the thermodynamic properties of the sub-cloud layer[22], as it transports mass with the same properties of the well-mixed sub-cloud layer. The positive correlation between W and $\mathcal{R}$ is thus likely connected to a self-aggregation feedback leading to a net convergence of moisture into areas that are already moist[25,32,33]. The opposing correlations of E and W with $\mathcal{R}$ lead to a negligible anti-correlation of M and $\mathcal{R}$ (r=-0.08 [-0.26, 0.10], Fig. 3b). While this makes M and $\mathcal{R}$ independent predictors of C, it contrasts with the expected desiccation effect of increased mixing. The basic premise of the mixing-desiccation hypothesis thus breaks down in the presence of strong variability in W.

Fig. 3c further shows a pronounced positive correlation between C and M (r=0.72 [0.64, 0.81]), demonstrating that M explains more than 50% of variability in C. The EUREC^{4}A data are therefore in line with a more direct relation $C \propto M^{\beta}$ and a $\beta > 0$ (Fig. 1b). The tight connection between C and M is also consistent with physical understanding represented in the scaling $C \sim 2C_{\mathrm{core}} \propto 2M/w^{*}$, where C_{core} is the area fraction of active cloud cores and w^{*} the Deardorff vertical velocity scale (see Methods and ref[24]). The correlation of C with $\mathcal{R}$ is weaker (r=0.36 [0.16, 0.56], Fig. 3d). These conclusions are robust to changes in the estimation procedure of M and to independent estimates of C (Extended Data Fig. 5).

The relationships exposed by the EUREC^{4}A data are thus in opposition to the mixing-desiccation hypothesis, which contends that increasing mixing (larger M) leads to a desiccation of the lower cloud layer (smaller $\mathcal{R}$) and a reduction in cloud-base cloudiness (smaller C). We also find a positive relationship between C and another indicator of lower-tropospheric mixing (Extended Data Fig. 4f) and a weak positive correlation between M and the total projected cloud cover (Extended Data Fig. 6). Hence, the EUREC^{4}A data emphasizes dynamic factors—the convective mass flux M and the mesoscale vertical velocity W—as dominant controls of C, rather than thermodynamic factors related to the mixing-desiccation mechanism.

Models underestimate strong cloud-circulation coupling

How consistent is the present generation of climate models with our observations? To assess how climate models represent the relationship between mixing and cloudiness, we use 10 models from the Cloud Feedback Model Intercomparison Project CFMIP[34] that provide the necessary point-wise M, C, and $\mathcal{R}$ output at high temporal resolution near the EUREC^{4}A domain (see Methods). In contrast to the consistency among many independent EUREC^{4}A observations, Fig. 4a shows that the models strongly differ regarding their magnitude and variability of M and C. While some models predict unrealistically low M (CanAM4, MIROC6, and MPI-ESM), the IPSL-CM6A has a 5-times larger mean M compared to the EUREC^{4}A observations. Except IPSL-CM6A, all models strongly overestimate variability in C (see also ref[35]), and 8 of 10 models also overestimate the magnitude of C. This is partly due to the tendency of models to produce stratocumulus clouds in this shallow cumulus regime[36,37] (evident in the strong increases in C (up to 50-100%) above a critical $\mathcal{R}$ of about 94%, see Extended Data Fig. 7). In contrast, the observations indicate no occurrence of C>13% or $\mathcal{R}$>94%. The models that produce such more stratocumulus-like conditions with $\mathcal{R}$>94% more than 15% of the time (Extended Data Fig. 8a) are labeled with open symbols in Fig. 4.

Only the BCC-CSM2 model represents the pronounced positive correlation between C and M observed during EUREC^{4}A at the 3 h-scale (Fig. 4b). Six of the other models have a correlation coefficient r<0.05, of which three models even show a negative correlation. The majority of models thus strongly underestimate the tight coupling between clouds and convection observed in EUREC^{4}A. Instead, these six models are more in line with the mixing-desiccation mechanism and a $\beta < 0$ (Fig. 1a), even though this is not mediated by a pronounced negative correlation between M and $\mathcal{R}$ (Extended Data Fig. 8c). All the models also strongly underestimate variability in W (Extended Data Fig. 8b), as they do not represent the sub-grid processes leading to the observed variability in the mesoscale vertical velocity (e.g., shallow circulations driven by differential radiative cooling[38] or local SST gradients[39]). The relationships between C and $\mathcal{R}$ are more consistent among most models (Fig. 4b), and also more consistent with the observations compared to the relationships between C and M.

In contrast with the observations, clouds as parameterized by climate models are more thermodynamically than dynamically controlled. The misrepresentation of the relative sensitivity of C to changes in M or $\mathcal{R}$ by all models is encapsulated in the ratio of the standardized regression coefficients $a_{\mathrm{M}}/a_{\mathcal{R}}$ from the regression $\widehat{C} = a_0 + a_{\mathrm{M}}\widetilde{M} + a_{\mathcal{R}}\widetilde{\mathcal{R}}$. The model samples lie completely outside the EUREC^{4}A data (Fig. 4c). All models, with one exception, substantially underestimate the value of $a_{\mathrm{M}}/a_{\mathcal{R}}$ compared to the observations, highlighting that in the climate models, variability in C is primarily controlled by variations in $\mathcal{R}$ rather than variations in M. Whereas BCC-CSM2 appears credible in terms of the magnitude and relationship of C and M, its credibility is eroded by its unrealistic relationship between C and $\mathcal{R}$ (Extended Data Fig. 7), and thus an implausible $a_{\mathrm{M}}/a_{\mathcal{R}}$ of -5.2. At odds with the observations, in most models M and $\mathcal{R}$ are only weak predictors of C, as evident in the low coefficient of determination

(r^2) of the multiple linear regression of $\widehat{C}$ (Extended Data Fig. 8c). The cloud parameterizations of the models thus fail in capturing the key relationships between C and the dynamic and thermodynamic environment observed in nature.

Implications for trade cumulus feedbacks

The EUREC^{4}A observations provide robust estimates of the mean, the variability, and the coupling of M, C, and $\mathcal{R}$ in contrasted trade cumulus environments. While the observed variability is substantial, the variability simulated by climate models is unrealistic, as are the drivers of this variability. The EUREC^{4}A data thus provide a physical test of the capacity of models to represent the interplay of the processes active in regulating trade-wind cloud amount, and may guide future model development. Moreover, the fact that the relationships at the 3 h process scale are consistent with the relationships at the monthly timescale (r$\geq$0.84, Extended Data Fig. 8e,f) suggests that the underlying fast physical processes that couple M, $\mathcal{R}$ and C in the models are largely invariant with the timescale. The relationships derived from the EUREC^{4}A observations can therefore also be used to evaluate the credible range of trade cumulus feedbacks in the climate models.

Fig. 4b demonstrates that all models with a strong trade cumulus feedback represented by a change in the cloud radiative effect (ΔCRE) with warming exceeding $0.37\,\mathrm{W\,m^{-2}\,K^{-1}}$ (reddish colors, Fig. 4c) represent the refuted mixing-desiccation mechanism with a negative (or very weak) correlation between M and C. Also, these four models exaggerate both the coupling of C to $\mathcal{R}$ (small $a_{\mathrm{M}}/a_{\mathcal{R}}$, Fig. 4c) and the variability in C (σ_{C}, Extended Data Fig. 8d). Contrastingly, the models that are closer to the observations tend to have a weaker positive ΔCRE with warming. The EUREC^{4}A observations of the physical processes that drive the short-term variability of C thus rule out the mechanism that leads to the largest positive trade cumulus feedbacks in current climate models.

By showing that mesoscale motions inhibit the mixing-desiccation mechanism, we refute an important physical hypothesis for a large trade cumulus feedback. In the spirit of the *story-line* approach for constraining equilibrium climate sensitivity[10], our findings thus refute an important line of evidence for a strong positive cloud feedback and thus a large climate sensitivity. The EUREC^{4}A observations therefore support recent satellite-derived constraints from observed natural variability[37,40] and climate-change experiments using idealized high-resolution simulations[41,42], which suggest that a weak trade cumulus feedback is more plausible than a strong one. Moreover, for the first time we take into account all types of clouds present in the trades, including the optically thinnest ones that are usually missed in satellite observations[43] and consider the full spectrum of mesoscale variability that was not represented in idealized simulations of cloud feedbacks. We also provide an explanation for the inconsistency of models with large positive feedbacks: in these models, the observed tight coupling between convective mixing and cloudiness is absent; instead, C is primarily controlled thermodynamically by $\mathcal{R}$, which exaggerates variability in C and feedbacks to warming. By not representing the variability in mesoscale circulations, the models miss an important process regulating trade cumulus clouds. Future research should focus on better understanding the processes controlling these mesoscale circulations, and how they might change in a warmer climate.

References

[1] Bony, S. & Dufresne, J.-L. Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.* **32**, L20806 (2005).

[2] Vial, J., Dufresne, J. L. & Bony, S. On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Clim. Dyn.* 3339–3362 (2013).

[3] Zelinka, M. D. *et al.* Causes of higher climate sensitivity in cmip6 models. *Geophys. Res. Lett.* **47**, e2019GL085782 (2020).

[4] Forster, P. *et al.* *The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity*, book section 7 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021). URL `www.ipcc.ch/report/ar6/wg1`.

[5] Vial, J., Bony, S., Dufresne, J.-L. & Roehrig, R. Coupling between lower-tropospheric convective mixing and low-level clouds: Physical mechanisms and dependence on convection scheme. *J. Adv. Model. Earth Syst.* **8**, 1892–1911 (2016).

[6] Brient, F. *et al.* Shallowness of tropical low clouds as a predictor of climate models' response to warming. *Clim. Dyn.* **47**, 433–449 (2016).

[7] Sherwood, S. C., Bony, S. & Dufresne, J.-L. Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature* **505**, 37–42 (2014).

[8] Bony, S. *et al.* Eurec4a: A field campaign to elucidate the couplings between clouds, convection and circulation. *Surv. Geophys.* **38**, 1529 (2017).

[9] Stevens, B. *et al.* Eurec^{4}a. *Earth Syst. Sci. Data* **13**, 4067–4119 (2021).

[10] Stevens, B., Sherwood, S. C., Bony, S. & Webb, M. J. Prospects for narrowing bounds on earth's equilibrium climate sensitivity. *Earth's Future* **4**, 512–522 (2016).

[11] Sherwood, S. C. *et al.* An assessment of earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics* **58**, e2019RG000678 (2020).

[12] Malkus, J. S. On the structure of the trade wind moist layer. *Papers in physical oceanography and meteorology* **13**, 1–47 (1958).

[13] Nuijens, L., Serikov, I., Hirsch, L., Lonitz, K. & Stevens, B. The distribution and variability of low-level cloud in the north atlantic trades. *Quart. J. Roy. Meteor. Soc.* **140**, 2364–2374 (2014).

[14] Rieck, M., Nuijens, L. & Stevens, B. Marine Boundary Layer Cloud Feedbacks in a Constant Relative Humidity Atmosphere. *J. Atmos. Sci.* **69**, 2538–2550 (2012).

[15] Stevens, B. *et al.* The Barbados Cloud Observatory: Anchoring Investigations of Clouds and Circulation on the Edge of the ITCZ. *Bull. Amer. Meteor. Soc.* **97**, 787–801 (2016).

[16] Dawe, J. T. & Austin, P. H. Statistical analysis of an les shallow cumulus cloud ensemble using a cloud tracking algorithm. *Atmos. Chem. Phys.* **12**, 1101–1119 (2012).

[17] Lamer, K., Kollias, P. & Nuijens, L. Observations of the variability of shallow trade wind cumulus cloudiness and mass flux. *J. Geophys. Res. Atmos.* **120**, 6161–6178 (2015).

[18] Klingebiel, M., Konow, H. & Stevens, B. Measuring shallow convective mass flux profiles in the trade wind region. *J. Atmos. Sci.* **78**, 3205 – 3214 (2021).

[19] Siebesma, A. P. *et al.* A Large Eddy Simulation Intercomparison Study of Shallow Cumulus Convection. *J. Atmos. Sci.* **60**, 1201–1219 (2003).

[20] Vogel, R., Bony, S. & Stevens, B. Estimating the shallow convective mass flux from the subcloud-layer mass budget. *J. Atmos. Sci.* **77**, 1559 – 1574 (2020).

[21] Vial, J., Bony, S., Stevens, B. & Vogel, R. Mechanisms and model diversity of trade-wind shallow cumulus cloud feedbacks: A review. *Surv. Geophys.* (2017).

[22] Albright, A. L., Bony, S., Stevens, B. & Vogel, R. Observed subcloud layer moisture and heat budgets in the trades. *J. Atmos. Sci.* (2022).

[23] Bony, S. & Stevens, B. Measuring area-averaged vertical motions with dropsondes. *J. Atmos. Sci.* **76**, 767–783 (2019).

[24] George, G., Stevens, B., Bony, S., Klingebiel, M. & Vogel, R. Observed impact of mesoscale vertical motion on cloudiness. *J. Atmos. Sci.* **78**, 2413 – 2427 (2021).

[25] Narenpitak, P., Kazil, J., Yamaguchi, T., Quinn, P. & Feingold, G. From sugar to flowers: A transition of shallow cumulus organization during atomic. *J. Adv. Model. Earth Syst.* **13**, e2021MS002619 (2021).

[26] Medeiros, B. & Nuijens, L. Clouds at barbados are representative of clouds across the trade wind regions in observations and climate models. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3062–E3070 (2016).

[27] Schulz, H. C^{3}ontext: a common consensus on convective organization during the eurec^{4}a experiment. *Earth Syst. Sci. Data* **14**, 1233–1256 (2022).

[28] George, G. *et al.* Joanne: Joint dropsonde observations of the atmosphere in tropical north atlantic meso-scale environments. *Earth Syst. Sci. Data* **13**, 5253–5272 (2021).

[29] Konow, H. *et al.* Eurec^{4}a's *HALO*. *Earth Syst. Sci. Data* **13**, 5545–5563 (2021).

[30] Vial, J. *et al.* A new look at the daily cycle of trade wind cumuli. *J. Adv. Model. Earth Syst.* **11**, 3148–3166 (2019).

[31] Bony, S. *et al.* Eurec^{4}a observations from the safire atr42 aircraft. *Earth Syst. Sci. Data* **14**, 2021–2064 (2022).

[32] Bretherton, C. S. & Blossey, P. N. Understanding mesoscale aggregation of shallow cumulus convection using large-eddy simulation. *J. Adv. Model. Earth Syst.* **9**, 2798–2821 (2017).

[33] George, G. *Observations of meso-scale circulation and its relationship with cloudiness in the Tropics*. Ph.D. thesis, Universität Hamburg, Hamburg (2021).

[34] Webb, M. J. *et al.* The cloud feedback model intercomparison project (cfmip) contribution to cmip6. *Geosci. Model Dev.* **10**, 359–384 (2017).

[35] Nuijens, L., Medeiros, B., Sandu, I. & Ahlgrimm, M. The behavior of trade-wind cloudiness in observations and models: The major cloud components and their variability. *J. Adv. Model. Earth Syst.* **7**, 600–616 (2015).

[36] Nam, C., Bony, S., Dufresne, J.-L. & Chepfer, H. The too few, too bright tropical low-cloud problem in cmip5 models. *Geophys. Res. Lett.* **39**, n/a–n/a (2012).

[37] Cesana, G. V. & Del Genio, A. D. Observational constraint on cloud feedbacks suggests moderate climate sensitivity. *Nature Climate Change* **11**, 213–218 (2021).

[38] Naumann, A. K., Stevens, B., Hohenegger, C. & Mellado, J. P. A conceptual model of a shallow circulation induced by prescribed low-level radiative cooling. *J. Atmos. Sci.* **74**, 3129 – 3144 (2017).

[39] Lambaerts, J., Lapeyre, G., Plougonven, R. & Klein, P. Atmospheric response to sea surface temperature mesoscale structures. *J. Geophys. Res. Atmos.* **118**, 9611–9621 (2013).

[40] Myers, T. A. *et al.* Observational constraints on low cloud feedback reduce uncertainty of climate sensitivity. *Nature Climate Change* **11**, 501–507 (2021).

[41] Vogel, R., Nuijens, L. & Stevens, B. The role of precipitation and spatial organization in the response of trade-wind clouds to warming. *J. Adv. Model. Earth Syst.* **8**, 843–862 (2016).

[42] Radtke, J., Mauritsen, T. & Hohenegger, C. Shallow cumulus cloud feedback in large eddy simulations – bridging the gap to storm-resolving models. *Atmos. Chem. Phys.* **21**, 3275–3288 (2021).

[43] Mieslinger, T. *et al.* Optically thin clouds in the trades. *Atmos. Chem. Phys.* **22**, 6879–6898 (2022).

Fig. 1 | Illustration of two mechanisms for the coupling of mixing and cloudiness. a, The *mixing-dessication mechanism* contends that E increases in response to an increase in M, which leads to a reduction in $\mathcal{R}$ and cloud-base cloudiness C, and a relationship $C \propto \mathcal{R} \propto M^{\beta}$ with $\beta < 0$. **b**, The *mesoscale motion control* of cloudiness instead suggests that M is equally controlled by both E and W, such that M is uncorrelated to $\mathcal{R}$ and $\beta > 0$.

Fig. 2 | Timeseries of mixing and cloudiness during EUREC^{4}A. Measurements of **a**, M, **b**, E and W, **c**, C, and **d**, $\mathcal{R}$, with filled symbols representing the 3 h-scale and open symbols the 1 h-scale. The vertical bars in **a-c** show the estimation uncertainty at the 3 h-scale (see Methods Sec. 'Uncertainty estimation'). The $\mathcal{R}$ in **d** is shown for both the HALO (blue) and ATR (green) aircraft, with the *x* markers representing the data points that are excluded in the correlations due to inconsistent sampling of the mesoscale cloud patterns between the two aircraft. The campaign mean$\pm 1\sigma$ is shown on the left side of each panel.

Fig. 3 | Relationships among M, $\mathcal{R}$, and C. The relationships between **a**, the observed C and the reconstructed $\widehat{C}$ from the regression $\widehat{C} = a_0 + a_{\mathrm{M}}\widetilde{M} + a_{\mathcal{R}}\widetilde{\mathcal{R}}$, **b**, M and $\mathcal{R}$, **c**, M and C, and **d**, $\mathcal{R}$ and C are shown at the 3 h-scale. The error bars represent the estimation uncertainty for M and C, and the sampling uncertainty for $\mathcal{R}$ (see Methods). The dotted line in **a** is the 1:1 line. The size of the markers in **b** represents C. The shading in **c** represents the scaling for $C \propto 2M/w^{*}$ using the mean$\pm 2\sigma$ of the velocity scale w^{*}. The grey *x* markers represent data that are excluded in the correlations due to inconsistent sampling between the two aircraft (see Fig. 2d and Methods).

Fig. 4| Relationships in climate models and link to trade cumulus feedback. a, Mean$\pm\sigma$/2 of M and C, **b**, correlation coefficients r between M and C ($r_{\mathrm{M,C}}$) and $\mathcal{R}$ and C ($r_{\mathcal{R}\mathrm{,C}}$), and **c**, ratio of the standardized multiple linear regression coefficients $a_{\mathrm{M}}/a_{\mathcal{R}}$ and the thermodynamic component of the trade cumulus radiative feedback. The models are colored in bins of feedback strength. Open symbols refer to models with frequent stratocumulus (defined as having $\mathcal{R}$>94% more than 15% of the time, see Extended Data Fig. 8a). The grey shading represents the 25th to 75th quartile and the grey bars the 95%-CI of bootstrapped observational values. For plotting purposes, **a** shows the mean $\bar{M}$-30 for IPSL-CM6A, and **c** shows the ratio $a_{\mathrm{M}}/a_{\mathcal{R}}$+3 for BCC-CSM2. In **c**, the upper end of the observational 95%-CI (at 6.75) is cropped.

Methods

EUREC^{4}A field campaign

We use data from the EUREC^{4}A field campaign, which took place in January and February 2020 and was anchored in Barbados[8,9]. We focus on measurements made by the HALO[29] and ATR aircraft[31], which flew coordinated patterns in the ca. 220 km diameter EUREC^{4}A circle centered at 13.3°N, 57.7°W. The HALO aircraft flew three circles at 10.2 km altitude in 200 min (ca. 60 min per circle plus 15 min break between circles) and launched dropsondes every 30° of heading (ca. 12 sondes per circle) to characterize the large-scale environment[28]. At the same time, the ATR aircraft flew 2–3 50 min rectangle-patterns inside the circle near cloud base and measured the cloud fraction with horizontally-staring cloud radar and backscatter lidar, and with several in-situ probes and sensors[31]. Observations from the Barbados Cloud Observatory (BCO)[15], and the R/V Meteor[44] provide additional context at the western and eastern boundaries of the EUREC^{4}A circle.

A typical flight day of HALO comprised two sets of three consecutive circles lasting about 3 h and comprising 30-36 sondes (sometimes defined as circling[9,22,29]). The 3 h-circle sets are separated by a 1.5 h break to refuel the ATR. The circle patterns were flown from January 22 to February 15 with different starting times between 04:30 and 19:30 local time (LT) to sample the diurnal cycle. Four additional single dropsonde circles are also used, three of which were flown by the P3 aircraft[45] during nighttime (starting at 00:15 LT on February 9 and 10, and at 01:30 LT on February 11). In total, the dataset comprises 73 circles (1 h-scale) and 24 sets of three consecutive circles (3 h-scale), for which 16 have coincident ATR data. We assume that HALO and ATR sample comparable conditions on the 3 h-scale. This is confirmed by the similar cloud-base $\mathcal{R}$ of the aircraft during most flights (Fig. 2d), except for the first 3 h-circle set on February 2 and the second 3 h-circle set on February 7 and 13 where the spatial scale of the cloud organization was larger than the scale of the domain sampled by the ATR. These three 3 h-circle sets are marked in the figures and excluded from the calculated correlations.

The spatial scale of the observations represents the lower end of Orlanski's[46] meso-α scale and is comparable in size to a climate model grid box. The 200–300 km scale is the relevant scale of the cloud processes for a trade cumulus ensemble and also the scale that convective parameterizations target. It lies in between the $\mathcal{O}(1\,\mathrm{km})$ scale of individual clouds and the synoptic scale of $\mathcal{O}(1000\,\mathrm{km})$, and is associated with the emergence of the prominent trade cumulus cloud organization patterns[47]. As the airmasses are advected by about 30 km per hour (at the campaign-mean wind speed of $\sim$9 m s^{-1} at 1 km height), the spatial sampling of the 220 km diameter circle does not differ substantially between the 1 h and 3 h timescales, which motivates our nomenclature focus on the time rather than space scale. Using the measurements, model and reanalysis data we would not expect our results to change substantially if the analysis domain were increased or reduced by a factor of two or more (see Methods subsection 'Mass flux estimation' for a discussion of the scale sensitivity of the results).

The Barbados region was chosen as location of EUREC^{4}A because shallow trade cumulus clouds are the dominant cloud type in the area during winter[13]. Furthermore, clouds in the Barbados region are similar to clouds across the trade-wind regions in both observations and models[26]. The mean meteorological conditions during the EUREC^{4}A campaign, as sampled by the dropsondes, also correspond well to the average January-February conditions from 12 years of data from the ERA-Interim reanalysis[48] (their Fig. 5), albeit with a 10% larger 850 hPa relative humidity during EUREC^{4}A (the EUREC^{4}A dropsondes also have an $\sim$8 % larger relative humidity compared to the 2013-2022 average in ERA5, not shown). Also, all the four prominent patterns of mesoscale cloud organization[47] were present during the campaign[27]. The conclusions drawn from the EUREC^{4}A data are thus relevant across the tropics and for climate timescales.

Observations

For estimating the cloud-base mass flux M, $\mathcal{R}$, and many other variables, we use dropsonde data from the JOANNE dataset[28], namely Level-3 (gridded quality-checked sondes) and Level-4 (circle products) vertical profiles of thermodynamic quantities, wind, and mesoscale vertical velocity, W. The HALO dropsondes are corrected for a dry bias by multiplying the relative humidity with 1.06[28].

For the cloud-base cloud fraction C, we use the BASTALIAS lidar-radar synergy product[31], which includes both cloud and drizzle (but not rain) and constitutes an upper bound on C. We also test the relationships for three additional estimates of C:

- the non-drizzling cloud product from the radar-lidar synergy (C_{only}), which excludes drizzle and constitutes a lower bound on C
- in situ estimates from a microphysical probe defined based on thresholds of liquid water content plus particle size (C_{pma})
- in situ high-frequency (25 Hz) humidity sensor, with cloud defined as relative humidity $\geq$98% (C_{turb})

The in situ sensors measure the along-track C, while the lidar-radar synergy samples clouds inside the rectangle at a distance up to 8 km from the aircraft[31]. Despite pronounced differences in the measurement principles and sampling, Fig. 18 of ref[31] demonstrates the internal consistency and robustness among the independent C estimates. The ATR turbulence measurements also include measurements of vertical up- and downdraft velocities[49], from which an in-cloud mass flux M_{turb} is computed by multiplying C_{turb} with the in-cloud vertical velocity.

Additional HALO aircraft measurements used are total projected cloud cover (CC) estimates from the differential absorption lidar WALES, the hyperspectral imager specMACS, and the cloud radar HAMP[29]. From these cloud masks we derive the CC along the 1 h circle. For specMACS and HAMP, the cloud detection is ambiguous and we consider both the *probably cloudy* and the *most likely cloudy* flags in our CC estimates.

We also use ceilometer and cloud radar data from the BCO and the R/V Meteor to test the robustness of the sub-cloud layer height definition (not shown). Radar cloud fraction profiles are obtained by correcting the hydrometeor fraction profiles with ceilometer data during periods of rain (see ref[30] for a description of the correction applied). The BCO cloud radar data also demonstrates that missing the level of maximum cloud-base cloud fraction in 3 h averages by e.g. 60 m does not affect the variability of C (correlations of r=0.99 and r=0.93 with the maximum C when 60 m above and below the peak level, respectively), and only marginally affects its magnitude (18% and 33% smaller relative to the maximum C for being 60 m above or below the peak level, respectively). So only if the ATR flight level deviated from the height of maximum cloudiness in ways that co-varied with M would we expect such a height difference to influence our analysis. As the ATR aircraft usually flew a bit above h (Extended Data Fig. 3a), and because it sampled much more clouds in 3 h compared to the stationary BCO, a potential influence of missing the peak level is deemed not to bias our findings.

Surface buoyancy flux

To estimate the surface buoyancy flux ($\overline{w'\theta_{\mathrm{v}}'}\big|_{\mathrm{s}}$, needed to compute M), we use dropsonde humidity, temperature, and wind data at 20 m height, and apply the Coupled Ocean-Atmosphere Response Experiment (COARE) bulk flux algorithm version 3.6[50,51]. For the sea-surface temperature (SST), we extrapolate the 2 m depth SST of the R/V Meteor (thermosalinograph primary backboard temperature), or alternatively from the AutoNaut

Caravela[52], to the dropsonde location based on a fixed zonal and meridional SST gradient of -0.14 K/degree. A gradient of -0.14 K/degree corresponds to the median zonal and meridional gradient (-0.145 K/degree and -0.135 K/degree, respectively) across the EUREC^{4}A circle over the period from January 19 to February 15 in the ERA5 reanalysis[53] and in two satellite SST products (from the Advanced Baseline Imager on board the Geostationary Operational Environmental Satellite, GOES-16 ABI, and the Collecte Localisation Satellites, CLS).

The sonde-derived surface buoyancy flux on the 3 h-scale compares favorably to bulk fluxes from the R/V Meteor mast, with a correlation coefficient r=0.83 and a mean offset of 0.1% relative to R/V Meteor. The sonde-derived flux has a comparable magnitude to the flux measured at the R/V Ron Brown[54] further upstream, and is also well-correlated (r=0.81) with ERA5. The ERA5 fluxes, however, overestimate the surface buoyancy flux compared to the sonde-derived flux by 25%, which is mostly due to the overestimation of the sensible heat flux by 64% relative to the observations (9.8 W m^{-2} and 6.0 W m^{-2} for ERA5 and dropsondes, respectively). A strong overestimation of the sensible heat flux compared to buoy measurements in the region is also present in the predecessor ERA-interim reanalysis[55]. Overall, the good correspondence of our sonde-derived surface buoyancy flux with the independent data lends credibility to our estimation procedure. The sonde-derived surface buoyancy flux is also used to compute the Deardorff sub-cloud layer vertical velocity scale $w^* = (h\,\frac{g}{\theta_\mathrm{v}}\overline{w'\theta'_\mathrm{v}}\big|_\mathrm{s})^{1/3}$ shown in Fig. 3c, where g is the gravitational acceleration.

Mass flux estimation

Vogel et al.[20] developed a method to estimate the shallow-convective mass flux at the sub-cloud layer top as a residual of the sub-cloud layer mass budget, and tested it in real-case large-eddy simulations (LES) over the tropical Atlantic. Here the method is applied to

EUREC^{4}A observations, in parallel with Albright et al.[22] who close the sub-cloud layer moisture and heat budgets and provide an independent constraint on the entrainment rate E. Except for the surface-buoyancy flux estimate (see the previous section), all data for the budgets come entirely from the dropsondes.

Eq. 1 expresses the budget of the sub-cloud layer height h per unit area and constant density. $\frac{\partial h}{\partial t}$ represents the temporal fluctuation of h and $V_h \cdot \nabla h$ its horizontal advection, E is the entrainment rate, W the mesoscale vertical velocity (positive upwards), and M the convective mass flux at h.

The sub-cloud layer height h is defined as the height where the virtual potential temperature (θ_v) first exceeds its density-weighted mean from 100 m up to h by a fixed threshold $\epsilon = 0.2\,\mathrm{K}$[22,56]. Extended Data Fig. 3a confirms that our h is usually close to the ATR flight altitude, and h is also well within the range of independent BCO and R/V Meteor observations of the maximum radar cloud-base cloud fraction and the peak frequency of the first ceilometer cloud-base height (not shown). This confirms that our h agrees well with the level of maximum near-base cloud fraction, which was set as the target height for the ATR flight level and thus for evaluating the mass budget[31].

The entrainment rate E represents the deepening of h due to small-scale mixing at the sub-cloud layer top. We use a modified version of the classical flux-jump model[57,58] that accounts for the finite thickness of the transition layer, the ∼150 m thick stable layer separating the mixed layer from the cloud layer (see ref[22] for details). The buoyancy flux at h is modeled as a fixed fraction A_e of the surface buoyancy flux, $\overline{w'\theta'_\mathrm{v}}\big|_\mathrm{s}$, where A_e is the effective entrainment efficiency. The buoyancy-jump at the sub-cloud layer top is computed as $\Delta\theta_\mathrm{v} = \Delta\theta + 0.61(\overline{\theta}\Delta q + \overline{q}\Delta\theta)$, with $\Delta\theta = C_\theta(\theta_{\mathrm{h}+} - \overline{\theta})$ and $\Delta q = C_\mathrm{q}(q_{\mathrm{h}+} - \overline{q})$. q is the specific humidity, C_q and C_θ are scaling coefficients accounting for uncertainty in the depth over which the jumps are computed, the subscript h+ refers to the value of q or

θ above h (computed as the average from h to $h + 100\,\text{m}$), and $\overline{q}$ and $\overline{\theta}$ are averages from 50 m to the mixed-layer top (defined as the height of maximum relative humidity below 900 m). Finally, E is computed as

$$E = \frac{A_\text{e}\,\overline{w'\theta_\text{v}'}\big|_\text{s}}{\Delta\theta_\text{v}} \tag{S1}$$

The uncertain parameters A_e, C_q and C_θ are estimated through a joint Bayesian inversion to close the moisture and heat budgets by ref[22], yielding maximum-likelihood estimates of $A_\text{e} = 0.43 \pm 0.06$ (mean$\pm 1\sigma$), $C_\text{q} = 1.26 \pm 0.34$, and $C_\theta = 1.15 \pm 0.31$.

The mesoscale vertical velocity W at h is computed by vertically integrating the divergence of the horizontal wind field measured by the dropsondes[23] from the surface up to h. W is at the lower end of the meso-α scale of ref[46], what climate modelers often associate with the *large-scale*. The terms h, E and W are computed at the 1 h-scale of a single circle and then aggregated to the 3 h-scale (three circles).

The temporal fluctuation of h is estimated as the linear regression slope of h computed from the 30-36 soundings available per 3 h-circle set. Similarly, the horizontal advection of h is estimated as the sum of the linear regressions of the eastward ($\partial h/\partial x$) and northward ($\partial h/\partial y$) gradients of the individual h, multiplied by the wind speed at the 3 h-mean h. Both $\partial h/\partial t$ and $V_h \cdot \nabla h$ are only available on the 3 h-scale.

The default M shown in the paper is the equilibrium mass flux $M = E + W$, which reproduces well the mass flux diagnosed directly from cloud-core area fraction and vertical velocity in LES[20]. This equilibrium M is also available on the 1 h-scale of an individual circle. Taking into account $\partial h/\partial t$ and $V_h \cdot \nabla h$ in the mass flux estimate leads to $M' = M - \frac{\partial h}{\partial t} - V_h \cdot \nabla h$, which shows very similar characteristics compared to M (Extended Data Fig. 3). This is mainly because both the advection ($-1.3 \pm 2.7\,\text{mm}\,\text{s}^{-1}$) and temporal fluctuation ($0.5 \pm 6.8\,\text{mm}\,\text{s}^{-1}$) terms are on average about zero, and the advection term is also nearly invariant. The inclusion of advection and $\frac{\partial h}{\partial t}$ in M' slightly enhances variability on the diurnal timescale (Extended Data Fig. 1a).

Cold pools formed by evaporating precipitation destroy the structure of the sub-cloud layer and make the estimation of h less robust. We thus exclude soundings that fall into cold pools in the analysis using the criterion of $h < 400\,\mathrm{m}$ developed by ref[56] based on the EUREC^{4}A soundings. The influence of these and other assumptions on the magnitude and variability of M are discussed in the Methods subsection 'Robustness of observational estimates'. Also note that our M is defined as the (mass) specific mass flux and has units of velocity. It differs from the more familiar mass flux (in units of $\mathrm{kg\,m^{-2}\,s^{-1}}$) by the air density, which is usually assumed to be constant[18,59], and which is justified here given the small density variations across the measurements (mean$\pm\sigma$ of $1.104\pm0.0077\,\mathrm{kg\,m^{-3}}$, i.e. less than 0.7% of the mean).

While the 1 h-scale variability of M can be substantial (e.g., 2$^{\mathrm{nd}}$ 3 h-circle sets on Jan 26 and Feb 13, Fig. 2), the median estimation uncertainty is only 20% at the 3 h-scale (see section below). Also, M has a similar magnitude and reassuring correlation (r=0.77) to the independent M_{turb} estimate from in-situ turbulence measurements on the ATR aircraft (Extended Data Fig. 2d).

The mass budget terms show different degrees of scale sensitivity (see also discussion in ref[20]). Extended Data Fig. 2c and 4a show that the correlation between W and M is slightly larger at the 1h-scale compared to the 3h-scale ($r_{W,M\ \mathrm{3h}}$=0.60 and $r_{W,M\ \mathrm{1h}}$=0.67), while they are essentially the same for E and M ($r_{E,M\ \mathrm{3h}}$=0.54 and $r_{E,M\ \mathrm{1h}}$=0.55). The scale sensitivity of the W variance is in line with radiosonde data from the EUREC^{4}A ship array, which show that the divergence amplitudes at equivalent radii of 100–300 km scale inversely with radius[60] (as in ERA5 and ICON, consistent with ref[23]). In ERA5, the scale sensitivity of the surface buoyancy flux, which contributes most to variability in

E (Extended Data Fig. 4b), is much smaller compared to the scale sensitivity of W (not shown). This is likely because variability in the surface buoyancy flux is mostly controlled by the surface wind speed (Extended Data Fig. 4h) and radiative cooling[61], both of which are large-scale. The surface wind speed has autocorrelation coefficients of 0.74 for a two day and 0.48 for an eight day lag (Fig. 3d of ref[22]). Although weaker compared to the synoptic variability, the surface wind also has a distinct diurnal cycle[62,63], which causes a diurnal cycle of the surface buoyancy flux (Extended Data Fig. 1c and ref[20]). Some of the diurnal variability in E is thus lost for longer temporal averaging. Also, the variability in the temporal fluctuation and horizontal advection of h (eq. 1) decreases on larger scales[20]. In summary, M variability decreases on larger averaging scales. The scale sensitivity of W is larger compared to E, such that the contribution of W to M variability tends to become smaller compared to the contribution of E on much larger scales.

As noted above, E describes the net effect of local processes and must be inferred from the statistics of other quantities (i.e., the mean sub-cloud layer growth rate, or the dilution of sub-cloud layer properties). This raises the question if the E estimate itself might depend on the mesoscale environment and therefore introduce spurious co-variabilities between M, W, and C. The Bayesian estimation of the uncertain parameter estimates A_e, C_q, and C_θ is a priori independent of M and W. Also, the synoptic variability during EUREC^{4}A can be well explained by keeping them constant[22]. Ref[22] also explored to what extent other factors correlated with residuals in their Bayesian fits and found no evidence of a systematic effect of other factors, including windspeed and shear[64]. As discussed above, the variability in E tends to be less scale-sensitive than W, and mostly controlled by larger-scale factors like the surface wind speed (through the surface buoyancy flux, Extended Data Fig. 4b,h). Furthermore, E and W are anticorrelated ($\mathrm{r}_{E,W}$=-0.35, Extended Data Fig. 4g). So both statistically from the anticorrelation and physically through the scale argument, we believe that our parameterization of E does not induce spurious co-variability.

Uncertainty estimation

For the M, $\mathcal{R}$, and C estimates, we distinguish two sources of uncertainty: sampling uncertainty and estimation (or retrieval) uncertainty. For all terms, the sampling uncertainty is computed at the 3 h-scale as the standard error, SE = $\sigma/\sqrt{n}$, of the three individual 1 h circle values (each representing $\sim$50 min of flight or up to 12 sondes), where σ is the standard deviation and n the number of circles.

The estimation uncertainty is computed differently for every term according to the underlying assumptions and choices. For $\mathcal{R}$, the manufacturer stated uncertainty (i.e., repeatability) is 2% and some additional uncertainty stems from the correction of the dry bias of the HALO dropsondes (see ref[28]). Because this uncertainty is the same for all data points, the estimation uncertainty of $\mathcal{R}$ is not shown in the figures. For C, the estimation uncertainty is computed for every 3 h-circle set as the SE of the four different estimates of C, namely C itself, C_{only}, C_{turb}, and C_{pma}. The uncertainty estimate therefore represents uncertainty in measurement principles and spatial sampling[31]. Additional uncertainties of the individual C estimates (e.g. due to the choice of thresholds) are neglected, as sensitivity tests suggest they are smaller than the uncertainty among the different C estimates[31].

For W, the advection term $V_h \cdot \nabla h$, and the temporal fluctuation $\partial h/\partial t$, the estimation uncertainty is taken as the SE of the respective regression used to compute the term. Because $\partial h/\partial t$ is computed from individual sondes, it contains both temporal and spatial variability of h on the 3 h-scale and its SE is inflated.

The estimation uncertainty of the surface buoyancy flux is a combination of uncertainty in the underlying SSTs and in the COARE bulk flux algorithm. We estimate the uncertainty in the underlying SSTs by computing the SE of five different versions of the flux (three with different fixed SST gradients (the default median value and the median$\pm IQR$, i.e. -0.14 K/degree, -0.21 K/degree, and -0.07 K/degree), one with a temporally varying gradient (not shown), and one with a different baseline SST (from the AutoNaut Caravela[52] instead of the R/V Meteor)), and add a 5% uncertainty of the COARE algorithm given in ref[50] as the 1 h–uncertainty in the 0-10 $\mathrm{m\,s^{-1}}$ wind speed range. For A_e and $\Delta\theta_\mathrm{v}$, we use the relative uncertainties of the Bayesian inversion as the estimation uncertainty (i.e. $\sigma(A_\mathrm{e})/A_\mathrm{e}$ for A_e, and the average of $\sigma(C_\mathrm{q})/C_\mathrm{q}$ and $\sigma(C_\theta)/C_\theta$ for $\Delta\theta_\mathrm{v}$).

Uncertainties in the three individual terms of E are propagated by adding their fractional uncertainties in quadrature to yield the estimation uncertainty of E. In the same spirit, the estimation uncertainty is propagated from the 1 h-scale to the 3 h-scale, and from the individual terms of Eq. 1 to M.

The uncertainties of the correlations and the multiple linear regression are estimated with bootstrapping (10,000 repetitions). We communicate these uncertainties by mentioning the 25th and 75th quartiles in the text, and by displaying both the quartiles and the 2.5% and 97.5% quantiles (representing the 95% confidence interval, CI) in Fig. 4 and Extended Data Fig. 8. Apart from the uncertainty quantification described here, we assess the robustness of the M and C observations to several other choices and assumptions in the Methods subsection 'Robustness of observational estimates'.

Other mixing indicators

Other proxies for lower-tropospheric mixing were used in previous studies[5,7,65] that can be estimated from the dropsonde data and compared to the variability in C. Here we compute the boundary-layer vertical advection (BVA) diagnostic from ref[65] defined as $\mathrm{BVA} = \int_0^{Z_\mathrm{min}} W(z)\frac{\partial MSE}{\partial z}\rho dz$, where MSE is the moist static energy, Z_min the level of minimum MSE that marks the top of the trade-wind layer (on average at 2900 m), and ρ the density. Note that a lower (more negative) BVA value indicates stronger mixing.

Ref[65] found a pronounced positive relationship between changes in BVA and changes in C from a series of single-column model experiments with the IPSL-CM5A model, which is characterized by a strong positive low-cloud feedback and the presence of the mixing-desiccation mechanism (Fig. 4). Extended Data Fig. 4f shows a pronounced negative correlation between BVA and M in the EUREC^{4}A data, indicating good agreement in their complementary definitions of mixing. Smaller BVA (stronger mixing) is also associated with larger C (not shown), which is at odds with the IPSL-CM5A model. The absolute correlation between BVA and C (r=0.34), however, is considerably smaller than the correlation between M and C (r=0.72).

General circulation models

The cloud fraction, net mass flux (upward and downward), and relative humidity at cloud base are calculated for 10 Coupled Model Intercomparison Project (CMIP) models:

- 4 from the fifth phase, CMIP5[66]: CanAM4[67], MPI-ESM-LR[68], IPSL-CM5A-LR[69], HadGEM2-A[70], and
- 6 from the sixth phase, CMIP6[71]: BCC-CSM2-MR[72], CNRM-CM6-1[73], IPSL-CM6A-LR[74], MIROC6[75], MRI-ESM2-0[76], HadGEM3-GC31-LL[77],

using the sub-hourly vertical profiles at selected sites (named *cfSites* in CMIP5 and CF_{subhr} in CMIP6) provided by CFMIP[34]. Note that the M from the models is not computed using Eq. 1, but is defined according to the respective convective parameterization scheme of the models (see references above). We use the atmosphere-only *amip* configuration from 1979-2008, selecting data from December, January, February, and March to be broadly consistent with the winter conditions sampled during EUREC^{4}A. For each model, between 2–6 sites are available in the north Atlantic trades between 60-50°W and 12-16°N, namely the BOMEX, NTAS, EUREC^{4}A, BCO, and RICO sites. All profiles with clouds above

600 hPa (about 4.2 km) are dropped to ensure a focus on shallow convection. We verified that in terms of the large-scale environment, the cfSites fall into the climatological trade cumulus regime as defined by ref[40].

The cloud base level is defined as the level of maximum cloud fraction between 870-970 hPa (between about 400-1300 m). If the maximum cloud fraction is smaller than 0.25% for a given profile, the cloud-base level is taken at the climatological level of maximum cloud fraction. The hourly cloud-base data are aggregated to a 3 h-timescale, which corresponds to the 3 h-scale of the EUREC^{4}A data, as well as a monthly timescale. The values computed are insensitive to (a) averaging across the sites before aggregating to the 3 h-timescale, (b) removing the site near the Northwest Tropical Atlantic Station buoy upstream of the EUREC^{4}A circle (near 51°W and 15°N), (c) focusing only on January and February, and (d) excluding nighttime values outside the hours sampled during EUREC^{4}A (not shown).

We use the thermodynamic component of the change in the cloud radiative effect at the top of the atmosphere (ΔCRE) with warming under given dynamical conditions to quantify the strength of the trade cumulus radiative feedback. Ref[2] showed that the ΔCRE with warming is a good approximation of the cloud feedback computed with radiative kernels[78]. The CRE is defined as the difference between all-sky (all, including clouds) and clear-sky (clr, clouds assumed to be transparent to radiation) net downward radiative fluxes, CRE = $R_{\text{all}} - R_{\text{clr}} = (LW_{\text{clr}} - LW_{\text{all}}) + (SW_{\text{all}} - SW_{\text{clr}})$ = CRE_{LW} + CRE_{SW}, with R being the total radiative flux, and LW and SW its longwave and shortwave components. The radiative fluxes are defined positive downward. The ΔCRE with warming is then simply the difference in CRE between the warmer *amip4K* (4 K uniform increase in SST) and the *amip* (control) simulations, normalized by the 4 K temperature difference (i.e., $\Delta\text{CRE}/\Delta\text{T}_\text{s}$ = ($\text{CRE}_{\text{amip4K}}$ - CRE_{amip}) / 4K). To restrict the feedback estimation to the trade cumulus regime, we focus on ocean-only grid points between 35°S to 35°N, and use the regime partitioning of ref[40] with trade cumulus regimes in each simulation (*amip* or *amip4K*) defined as having a climatological annual mean estimated inversion strength smaller than 1 K and a vertical velocity at 700 hPa between 0–15 $\mathrm{hPa\,d^{-1}}$.

Robustness of observational estimates

Applying the mass budget formulation to the EUREC[4]A dropsonde data involves several choices for definitions and thresholds. These choices are guided by constraints from independent data and from closure of the moisture and heat budgets in ref[22], which provides justification for the default configuration described in the Methods subsection 'Mass flux estimation'. Nevertheless, it is important to assess and understand the sensitivity of the mass budget estimates and the key relationships to different estimation choices.

We focus first on the influence of different definitions of the sub-cloud layer height h and the entrainment rate E on the mean and standard deviation (σ) of M and E, the respective correlations of M with E and W, and the correlation and mean difference to the independent M_{turb} estimate from turbulence measurements onboard the ATR aircraft (see Extended Data Fig. 2). For the h definition, we compare our default h to an alternative definition, *h.parcel*, which defines h as the level of neutral buoyancy of a surface-lifted parcel (with density-weighted θ_{v} averaged from 30–80 m) plus 0.2 K θ_{v}-excess. Using *h.parcel* leads to a 16 m shallower mean h compared to the default. The slightly shallower h decreases $\Delta\theta_{\mathrm{v}}$ (the denominator of the E formulation in Eq. S1) from 0.36 K to 0.34 K, which slightly increases E and M by $\sim 1.5\,\mathrm{mm\,s^{-1}}$. While W is unaffected by this small change in h, the resulting M has a slightly reduced correlation to the independent M_{turb} compared to the default M (r=0.69 vs. r=0.77). The same chain of arguments holds for increasing and decreasing the threshold ϵ in the h definition by $\pm 0.05\,\mathrm{K}$. With $\epsilon = 0.25\,\mathrm{K}$ instead of 0.2 K (case *h.eps=0.25*), h increases by 31 m, and through the larger $\Delta\theta_{\mathrm{v}}$ decreases E and M by ∼3.3 mm s^{-1}. Due to the presence of a thin transition layer[22], the response to $\epsilon \pm 0.05\,\mathrm{K}$ is nonlinear and a reduction of ϵ to $0.15\,\mathrm{K}$ (*h.eps=0.15*) leads to a disproportionately smaller $\Delta\theta_{\mathrm{v}}$ and ∼6 mm s^{-1} larger E and M. The 35 m shallower mean h with $\epsilon = 0.15\,\mathrm{K}$ also strongly increases σ_{E}, which increases the correlation between E and M at the expense of a decreased correlation between the unaffected W and M (Extended Data Fig. 2c).

The next set of choices regard the entrainment rate estimate. We test the influence of the different surface buoyancy flux estimates from ERA5 and R/V Meteor. As the ERA5 flux is 25% larger than the other fluxes, we scale it to have same mean as the dropsonde-derived flux (case *sbf=ERA5.sc*). For *sbf=ERA5.sc*, the variability in E and M are substantially larger compared to the default dropsonde flux, increasing their correlation. For case *sbf=Meteor*, the differences to the default estimates is smaller (Extended Data Fig. 2a,b) and the correlation with M_{turb} slightly larger than in the other configurations. The estimates are also unaffected by changing the three coefficients A_{e}, C_{q}, and C_{θ} estimated by Bayesian inversion in ref[22] to close the moisture and energy budgets during EUREC^{4}A when cold pool soundings (defined as having $h < 400\,\mathrm{m}$ following ref[56]) are excluded (*diffEpars*). We further compare four different ways of computing $\Delta\theta_{\mathrm{v}}$. Computing the value at h+ as averages from h to $h + 50$ or $h + 150\,\mathrm{m}$ (instead of to $h + 100\,\mathrm{m}$) has a similar (but more linear) influence as increasing $\epsilon \pm 0.05\,\mathrm{K}$ (see discussion above). Using two different heights for averaging θ_{v} across the mixed layer (up to h in *tvbar=h* and up to the level where q first falls below its mean by a threshold of 0.3 g kg^{-1} in *tvbar=qgrad*) hardly influences the estimates.

Lastly, we show the influence of computing the mass budget including the cold pool soundings for two sets of surface buoyancy flux estimates, case *withCP* for the default dropsonde-derived flux and *withCP_sbf=ERA5.sc* for the scaled ERA5 flux. In both cases, the mean and σ of both M and E are increased when cold pools are included (matching the mean E of ref[22] who included cold pools). However, especially for the default surface fluxes (case *withCP*), the correlation with M_{turb} is strongly reduced.

Extended Data Fig. 2a,d also show the influence of selected choices on the total mass flux M', which includes the contribution of the temporal fluctuation and horizontal advection of h. Because these additional terms are on average nearly zero (Extended Data Fig. 3c), their inclusion does not affect $\overline{M}$. σ_{M} instead increases by $\sim$1.5 mm s^{-1} due to the pronounced variability in the temporal fluctuation term. As this term is not very robust, we use the more reliable equilibrium M as our best estimate. The equilibrium M is also robust at the 1 h-scale of an individual circle (case *1h-scale*).

Overall, Extended Data Fig. 2 makes us very confident in the robustness of our mass budget estimates because they only show a modest sensitivity to the various choices, and because we can explain these sensitivities physically. Also, the independent ATR M_{turb} estimates (Extended Data Fig. 2d) and the additional constraints on E from our complementary analyses of the moisture and heat budgets in ref[22] (dashed lines in Extended Data Fig. 2b) lend further credibility to our default estimation choices.

Next, we focus on the sensitivity of the key relationships between M, C, and $\mathcal{R}$ to a selected set of plausible estimation choices of M and the different C estimates from the ATR aircraft. Extended Data Fig. 5a shows that the positive correlation between M and C is significant for all parameter choices, and both the equilibrium M and total M'. Furthermore, also the negligible correlation between M and $\mathcal{R}$ is very robust.

Extended Data Fig. 5b further confirms that the default M also has strong correlations with the three independent estimates of C from the ATR aircraft. The same is true for the other estimation choices of M, with a small overall range of correlations of $0.52 < \mathrm{r}_{\mathrm{M,C}} < 0.73$. Correlations between C and $\mathcal{R}$ are more variable between the different C

estimates and range from $0.12 < \mathrm{r}_{\mathcal{R},\mathrm{C}} < 0.63$. It's not surprising that the C_{only} estimate that neglects contributions from drizzle has the strongest correlation with $\mathcal{R}$, as it mostly features passive clouds that are more affected by ambient humidity than the more active clouds that also include drizzle. Note that there is also a slight dependency of r($\mathcal{R}$,C) on the M estimates, as the cases *h.parcel* and *h.eps=0.25* result in different h and thus different heights where $\mathcal{R}$ is evaluated.

The bottom panels of Extended Data Fig. 5 also confirm the robustness of the correlation coefficient of the multiple linear regression $\widehat{C} = a_0 + a_{\mathrm{M}}\widetilde{M} + a_{\mathcal{R}}\widetilde{\mathcal{R}}$ and the ratio of the standardized regression coefficients $a_{\mathrm{M}}/a_{\mathcal{R}}$ to the M estimation choices (Extended Data Fig. 5c) and the different C estimates (Extended Data Fig. 5d). There is no configuration with $a_{\mathrm{M}}/a_{\mathcal{R}} < 1$, indicating that C is always more strongly coupled to M than to $\mathcal{R}$ in the observations. Slightly larger values of $a_{\mathrm{M}}/a_{\mathcal{R}}$ and smaller correlations are evident for the total M'.

Also, the standard deviation of C (σ_C) is very similar for the different C estimates that include drizzle (between 2.1–3.7%, with 3.1% being the σ_C of the default BASTALIAS lidar-radar synergy product), and only slightly lower for the C_{only} estimate (1.6%) when using the full sample. Variability is slightly reduced in the smaller sample that overlaps with the HALO flights, because it excludes two night flights with larger cloudiness and two flights in dry environments with very small cloudiness (σ_C of 1.7–2.4% for the C estimates that include drizzle).

Overall, Extended Data Fig. 5 thus demonstrates the insensitivity of the observed relationships to a wide range of configurations. We therefore conclude that the relationships between mixing and cloudiness observed during EUREC^{4}A are very robust.

Data availability

All data used in this study are published in the EUREC^{4}A database of AERIS (https://eurec4a.aeris-data.fr/, last access: 28 July 2022). We use v2.0.0 of the JOANNE dropsonde data[28] (https://doi.org/10.25326/246). The specific ATR datasets[31] used are the BASTALIAS product (https://doi.org/10.25326/316), the turbulence measurements[49] (https://doi.org/10.25326/128), and the PMA CloudComposite dataset (https://doi.org/10.25326/237). The specific HALO datasets[29] used are cloud masks derived from WALES cloud-top height estimates (https://doi.org/10.25326/216), HAMP cloud radar (https://doi.org/10.25326/222), and specMACS (https://doi.org/10.25326/166), and the flight segmentation product (https://doi.org/10.5281/zenodo.4900003). From the BCO[15], we used ceilometer (https://doi.org/10.25326/367) and cloud radar data (https://doi.org/10.25326/55). From the R/V Meteor[44], we used standard dship meteorological data for the EUREC^{4}A Meteor cruise M161 (retrieved from http://dship.bsh.de/, last access: 28 June 2022), surface heat fluxes (https://doi.org/10.25326/312), ceilometer measurements (https://doi.org/10.25326/53), and cloud radar data (v1.1, https://doi.org/10.25326/164). We further used data from AutoNaut Caravela[52] (https://doi.org/10.25326/366), and 10 min air-sea flux data (v1.3, https://doi.org/10.25921/etxb-ht19) from the R/V Brown[54]. Also, we used CLS Daily High Resolution Sea Surface Temperature maps (retrievable through the AERIS operational center https://observations.ipsl.fr/aeris/eurec4a-data/SATELLITES/CLS/SST/, last access: 28 June 2022, or directly from https://datastore.cls.fr/catalogues/sea-surface-temperature-infra-red-high-resolution-daily), GOES-16 ABI SSTs from the ABI_G16-STAR-L3C-v2.7 product (https://doi.org/10.25921/rtf0-q898), and ERA5[53] reanalysis data. The CMIP5 and

CMIP6 climate model output are available for download at https://esgf-node.llnl.gov.

Code availability

The scripts used for the analyses and other supporting information that may be useful for reproducing this study can be obtained from https://doi.org/10.5281/zenodo.7032765.

References

[44] Mohr, W. *et al.* Eurec4a campaign, cruise no. m161, 17 jan 2020 - 03 mar 2020, bridgetown *barbados* - ponta delgada *portugal*, meteor-berichte (2020). URL `https://www.tib.eu/de/suchen/id/awi%3A9cd8fcd5182aa44edcd25c7d688ba9c78`

[45] Pincus, R. *et al.* Observations from the noaa p-3 aircraft during atomic. *Earth Syst. Sci. Data* **13**, 3281–3296 (2021).

[46] Orlanski, I. A rational subdivision of scales for atmospheric processes. *Bull. Amer. Meteor. Soc.* **56**, 527–530 (1975).

[47] Stevens, B. *et al.* Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds. *Quart. J. Roy. Meteor. Soc.* **146**, 141–152 (2020).

[48] Brueck, M., Nuijens, L. & Stevens, B. On the seasonal and synoptic time-scale variability of the north atlantic trade wind region and its low-level clouds. *J. Atmos. Sci.* **72**, 1428–1446 (2015).

[49] Brilouet, P.-E. *et al.* The eurec^{4}a turbulence dataset derived from the safire atr 42 aircraft. *Earth Syst. Sci. Data* **13**, 3379–3398 (2021).

[50] Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A. & Edson, J. B. Bulk parameterization of air-sea fluxes: Updates and verification for the coare algorithm. *J. Climate* **16**, 571 – 591 (2003).

[51] Edson, J. B. *et al.* On the exchange of momentum over the open ocean. *J. Phys. Oceanogr.* **43**, 1589 – 1610 (2013).

[52] Siddle, E., Heywood, K. J., Webber, B. G. M. & Bromley, P. First measurements of ocean and atmosphere in the tropical north atlantic using caravela, a novel uncrewed surface vessel. *Weather* **76**, 200–204 (2021).

[53] Hersbach, H. *et al.* The era5 global reanalysis. *Quart. J. Roy. Meteor. Soc.* **146**, 1999–2049 (2020).

[54] Quinn, P. K. *et al.* Measurements from the rv *Ronald H. Brown* and related platforms as part of the atlantic tradewind ocean-atmosphere mesoscale interaction campaign (atomic). *Earth Syst. Sci. Data* **13**, 1759–1790 (2021).

[55] Bigorre, S. P. & Plueddemann, A. J. The annual cycle of air-sea fluxes in the northwest tropical atlantic. *Frontiers in Marine Science* **7** (2021).

[56] Touzé-Peiffer, L., Vogel, R. & Rochetin, N. Cold pools observed during eurec4a: detection and characterization from atmospheric soundings. *J. Appl. Meteor. Climatol.* (2022).

[57] Lilly, D. K. Models of cloud-topped mixed layers under a strong inversion. *Quart. J. Roy. Meteor. Soc.* **94**, 292–309 (1968).

[58] Stull, R. B. The energetics of entrainment across a density interface. *J. Atmos. Sci.* **33**, 1260–1267 (1976).

[59] Ghate, V. P., Miller, M. A. & DiPretore, L. Vertical velocity structure of marine boundary layer trade wind cumulus clouds. *J. Geophys. Res. Atmos.* **116** (2011).

[60] Stephan, C. C. & Mariaccia, A. The signature of the tropospheric gravity wave background in observed mesoscale motion. *Weather Clim. Dynam.* **2**, 359–372 (2021).

[61] Naumann, A. K., Stevens, B. & Hohenegger, C. A moist conceptual model for the boundary layer structure and radiatively driven shallow circulations in the trades. *J. Atmos. Sci.* **76**, 1289 – 1306 (2019).

[62] Vial, J., Vogel, R. & Schulz, H. On the daily cycle of mesoscale cloud organization in the winter trades. *Quart. J. Roy. Meteor. Soc.* **147**, 2850–2873 (2021).

[63] Savazzi, A. C. M., Nuijens, L., Sandu, I., George, G. & Bechtold, P. The representation of winds in the lower troposphere in ecmwf forecasts and reanalyses during the eurec4a field campaign. *Atmos. Chem. Phys. Disc.* **2022**, 1–29 (2022).

[64] Canut, G., Couvreux, F., Lothon, M., Pino, D. & Saïd, F. Observations and large-eddy simulations of entrainment in the sheared sahelian boundary layer. *Boundary Layer Meteorol.* **142**, 79–101 (2012).

[65] Brient, F. & Bony, S. Interpretation of the positive low-cloud feedback predicted by a climate model under global warming. *Clim. Dyn.* **40**, 2415–2431 (2013).

[66] Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of cmip5 and the experiment design. *Bull. Amer. Meteor. Soc.* **93**, 485 – 498 (2012).

[67] von Salzen, K. *et al.* The canadian fourth generation atmospheric global climate model (canam4). part i: Representation of physical processes. *Atmosphere-Ocean* **51**, 104–125 (2013).

[68] Stevens, B. *et al.* Atmospheric component of the mpi-m earth system model: Echam6. *J. Adv. Model. Earth Syst.* **5**, 146–172 (2013).

[69] Hourdin, F. *et al.* Impact of the lmdz atmospheric grid configuration on the climate and sensitivity of the ipsl-cm5a coupled model. *Clim. Dyn.* **40**, 2167–2192 (2013).

[70] Martin, T. H. D. T. G. M. *et al.* The hadgem2 family of met office unified model climate configurations. *Geosci. Model Dev.* **4**, 723–757 (2011).

[71] Eyring, V. *et al.* Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

[72] Wu, T. *et al.* The beijing climate center climate system model (bcc-csm): the main progress from cmip5 to cmip6. *Geosci. Model Dev.* **12**, 1573–1600 (2019).

[73] Voldoire, A. *et al.* Evaluation of cmip6 deck experiments with cnrm-cm6-1. *J. Adv. Model. Earth Syst.* **11**, 2177–2213 (2019).

[74] Hourdin, F. *et al.* Lmdz6a: The atmospheric component of the ipsl climate model with improved and better tuned physics. *J. Adv. Model. Earth Syst.* **12**, e2019MS001892 (2020).

[75] Tatebe, H. *et al.* Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in miroc6. *Geosci. Model Dev.* **12**, 2727–2765 (2019).

[76] Yukimoto, S. *et al.* The meteorological research institute earth system model version 2.0, mri-esm2.0: Description and basic evaluation of the physical component. *J. Meteorol. Soc. Japan Ser. II* **advpub** (2019).

[77] Walters, D. *et al.* The met office unified model global atmosphere 6.0/6.1 and jules global land 6.0/6.1 configurations. *Geosci. Model Dev.* **10**, 1487–1520 (2017).

[78] Soden, B. J. & Held, I. M. An assessment of climate feedbacks in coupled ocean–atmosphere models. *J. Climate* **19**, 3354 – 3360 (2006).

Acknowledgments: We thank all the scientists, engineers, technicians, and pilots who made the EUREC^{4}A data collection possible (see ref[9] for the complete list). We acknowledge Julien Delanoë and Patrick Chazette for making the horizontal radar-lidar cloud measurements on board the ATR possible, Marie Lothon, Pierre-Etienne Brilouet and the SAFIRE team for making the turbulent measurements and ATR flights possible, Martin Wirth and Silke Gross for producing sub-cloud layer height retrievals from the WALES lidar onboard HALO, Chris Fairall and Elizabeth Thompson for insights regarding SSTs and surface fluxes, Elizabeth Siddle for sharing the Autonaut Caravela data, Imke Schirmacher for comparison with her R/V Meteor surface flux estimates, Johannes Röttenbacher for making R/V Meteor data available, and Cathy Hohenegger for feedback on the manuscript. We thank three reviewers for their feedback. The EUREC^{4}A field campaign was funded by multiple international organizations[9]. The French team and ATR operations were funded by an ERC Advanced Grant (EUREC4A grant no 694768) backed by some H2020 funding (CONSTRAIN grant no 820829), and the German team and HALO operations were funded by the Max Planck Society (MPG), the German Research Foundation (DFG), the German Meteorological Weather Service (DWD) and the German Aerospace Center (DLR).

Author contributions: R.V., S.B., and B.S conceived the study, which was the primary objective of the EUREC^{4}A campaign; R.V. performed the analyses and wrote the paper; J.V. processed the CFMIP output and helped define model diagnostics relevant to cloud feedbacks; A.L.A. helped with the entrainment rate estimate and the analysis of CFMIP model outputs; S.B. diagnosed the cloud fraction and mass flux estimates from ATR observations; G.G. contributed additional JOANNE dropsonde and ERA5 products, and B.S. produced Fig. 1. All authors contributed to interpreting the results and writing the paper.

Competing interests: The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Raphaela Vogel.

Peer review information We thank three anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available. **Reprints and permissions information** is available at http://www.nature.com/reprints.

Extended Data Fig. 1 | Diurnal cycles of key terms. **a**, M and M', **b**, E and W, **c**, surface buoyancy flux, and **d**, C versus local time both at the 3 h-scale (filled) and 1 h-scale (open circles). The vertical bars show the estimation uncertainty at the 3 h-scale (see Methods Sec. 'Uncertainty estimation'). The correlation coefficients given in the legends represent the correlation between the individual terms and the time at the 3 h-scale ('r') and the 1 h-scale ('r.1h' in brackets).

Extended Data Fig. 2 | Influence of estimation procedure and parameter choices on mass budget estimates. Campaign mean and standard deviation of **a**, M and **b**, E, **c**, correlation coefficients between M and E ($\mathrm{r_{M,E}}$) and M and W ($\mathrm{r_{M,W}}$), and **d**, correlation and mean difference between M and M_{turb} from ATR turbulence measurements, for different configurations of the mass budget. Open symbols in **a** and **d** show the total M'. The dashed lines in **b** show the mean and standard deviation of E from ref[22], and the zero line in **d**. See the Methods subsection 'Robustness of observational estimates' for details.

Extended Data Fig. 3 | Timeseries of other mass budget terms. Shown are **a**, h and the flight level of the ATR aircraft, **b**, the equilibrium M, the total M', and the M_{turb} from ATR turbulence measurements, **c**, the temporal fluctuation and advection terms, **d**, the surface buoyancy flux, and **e**, the $\Delta\theta_{\mathrm{v}}$. The vertical bars show the estimation uncertainty at the 3 h-scale (see Methods Sec. 'Uncertainty estimation'), and the small open circles show the 1 h-scale. The *x* markers in **a-b** indicate the data that are excluded in the correlations due to inconsistent sampling between the two aircraft. The campaign mean$\pm 1\sigma$ is shown on the left side of each panel.

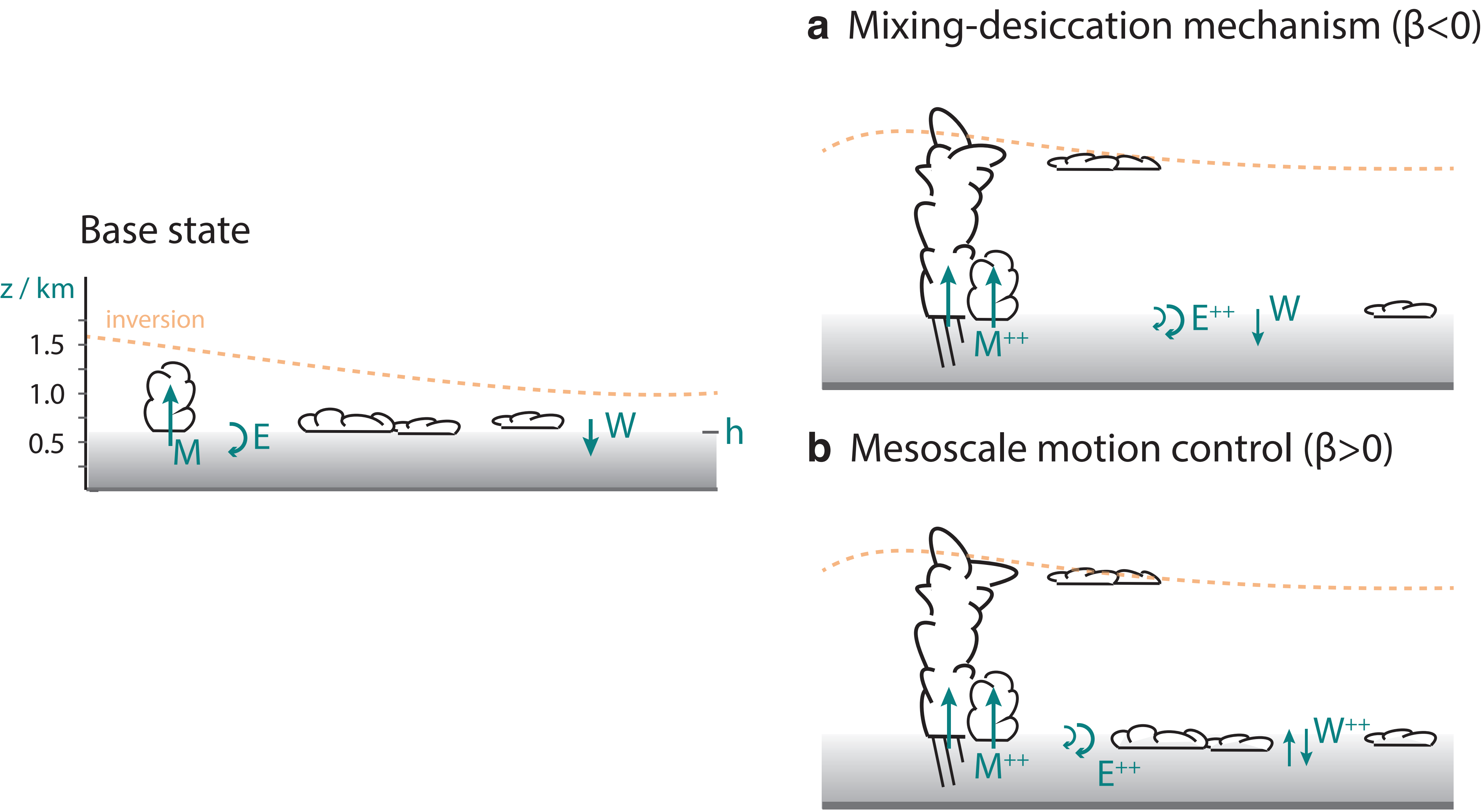
Extended Data Fig. 4 | Relationships of other key terms. **a**, E and W versus M, **b**, surface buoyancy flux versus E, **c**, $\Delta\theta_{\mathrm{v}}$ versus E, **d**, $\mathcal{R}$ versus E, **e**, $\mathcal{R}$ versus W, **f**, BVA mixing indicator versus M, **g**, E versus W, and **h**, 10 m wind speed versus surface buoyancy flux. **a-c,f-h** show both the 3 h-scale (filled) and 1 h-scale (open circles, with the corresponding correlation coefficient denoted as 'r.1h'). A dotted 1:1 line is shown in **a** and **g**. In **d-e**, the error bars represent the estimation uncertainty for E and W, and the sampling uncertainty for $\mathcal{R}$ (see Methods). The correlations in **d-e** are given both for the sample with consistent sampling among the HALO and ATR aircraft (blue points, as used for the correlations in Fig. 2&3), and for the entire sample of the HALO aircraft (including the grey points that represent the three data points marked with *x* in Fig. 3, and 8 other data points when ATR was not flying. The corresponding correlation coefficient is denoted as 'r.all').

Extended Data Fig. 5 | Influence of different M and C estimates on key relationships. Correlation coefficients r of **a**, M and C ($\mathrm{r_{M,C}}$) and M and $\mathcal{R}$ ($\mathrm{r}_{\mathrm{M},\mathcal{R}}$) and **b**, M and C ($\mathrm{r_{M,C}}$) and $\mathcal{R}$ and C ($\mathrm{r}_{\mathcal{R},\mathrm{C}}$). **c-d**, correlations of the reconstructed $\widehat{C} = a_0 + a_{\mathrm{M}}\widetilde{M} + a_{\mathcal{R}}\widetilde{\mathcal{R}}$ and the observed C ($\mathrm{r}_{\widehat{\mathrm{C}},\mathrm{C}}$), as well as the ratio of the standardized regression coefficients $a_{\mathrm{M}}/a_{\mathcal{R}}$. **a** and **c** also show the relationships for the total M' (open symbols), whereas **b** and **d** show the relationships for different estimates of C (different symbols). See details in Methods subsection 'Robustness of observational estimates'.

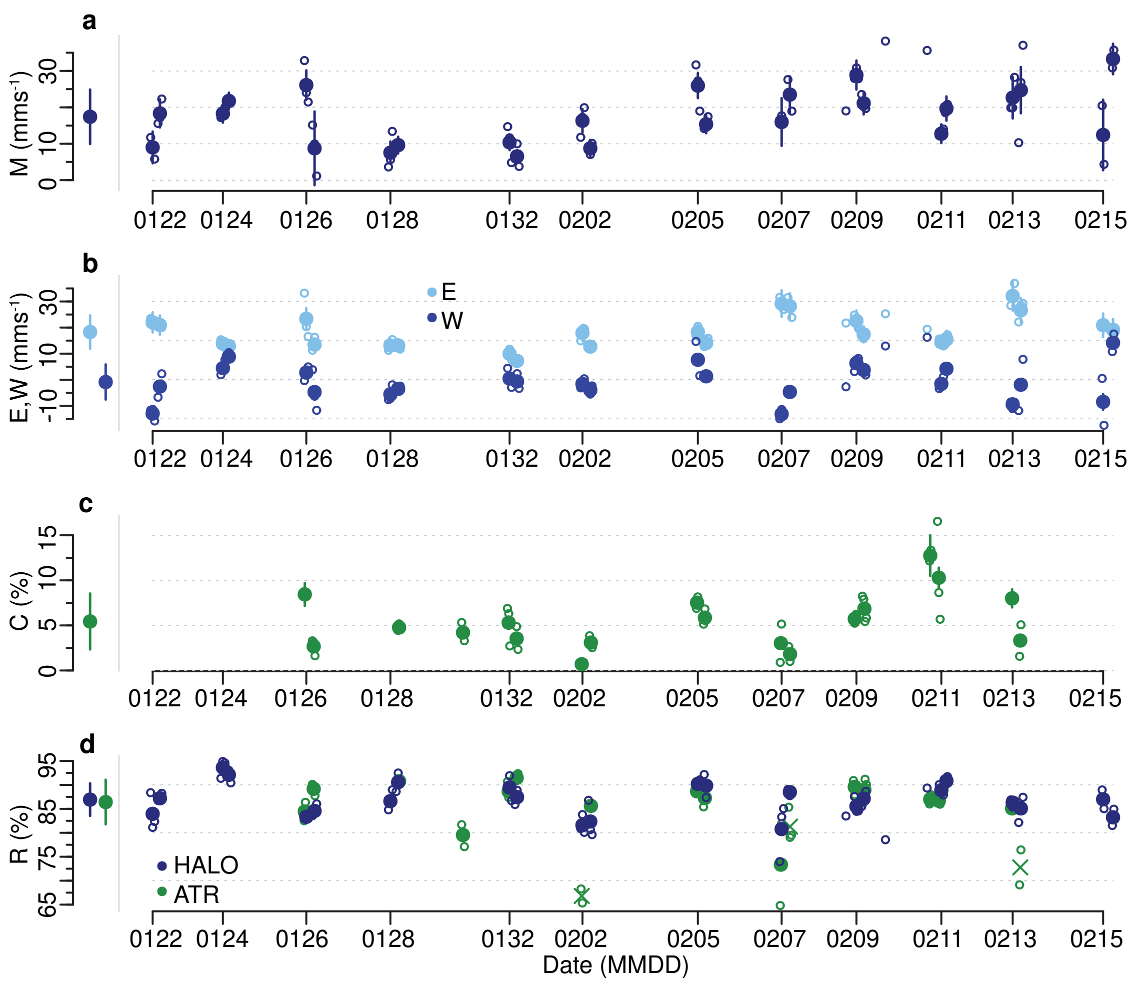
Extended Data Fig. 6 | Relationship of M with three estimates of the total projected cloud cover (CC). CC from **a**, WALES backscatter lidar, **b**, hyperspectral imager specMACS, and **c**, HAMP cloud radar on board HALO. The error bars represent the sampling uncertainty (for the CC estimates) and the estimation uncertainty (for M, see Methods Sec. 'Uncertainty estimation').

Extended Data Fig. 7 | Individual relationships of C, M and $\mathcal{R}$ for climate models. Relationships among individual 3 h C and M (1st and 3rd column) and C and $\mathcal{R}$ (2nd and 4th column) for all ten climate models. The red and blue points represent the median and mean of the respective variables, and the red lines extend from the 25th to the 75th quartile. The grey vertical line in the $\mathcal{R}$ panels shows the 94% $\mathcal{R}$-threshold.

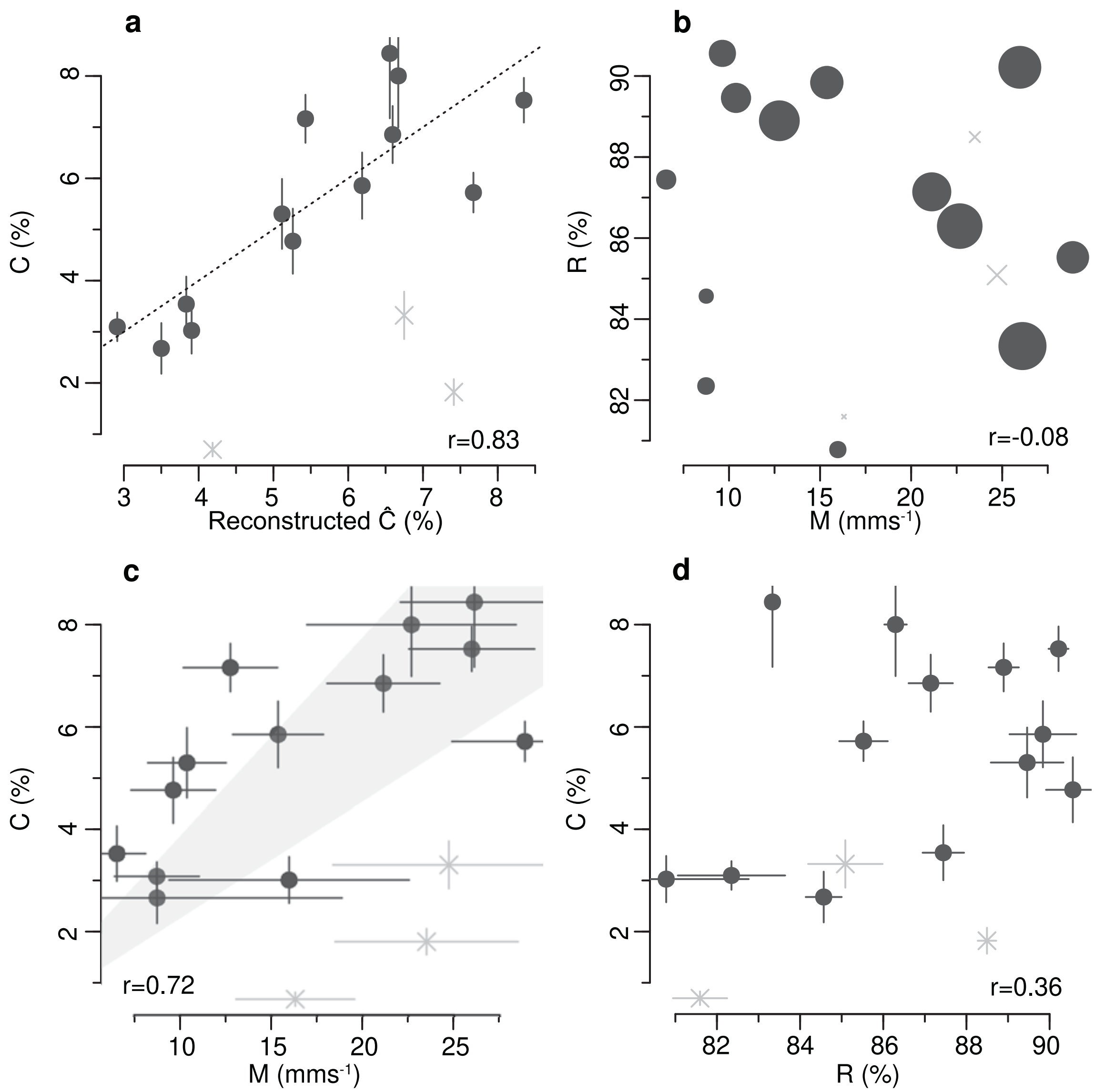
Extended Data Fig. 8 | Comparison of other variables and relationships in climate models against the EUREC^{4}A data. **a**, mean $\mathcal{R}$ and fraction of *stratocumulus-like* conditions with $\mathcal{R}$>94%, **b**, standard deviation of $\mathcal{R}$ and W ($\sigma_{\mathcal{R}}$ and σ_{W}), **c**, r^2 of multiple linear regression $\widehat{C} = a_0 + a_{\mathrm{M}}\widetilde{M} + a_{\mathcal{R}}\widetilde{\mathcal{R}}$ and correlation coefficient of M and $\mathcal{R}$, **d**, standard deviation of C (σ_{C}) and thermodynamic component of the cloud feedback ΔCRE/ΔT$_{\mathrm{s}}$, as well as the 3 h and monthly correlations of **e**, M and C, and **f**, $\mathcal{R}$ and C. **e-f** also show the inter-model correlation coefficients of the respective variables and the 1:1 line (dotted). As in Fig. 4, the models are colored in bins of feedback strength, and open symbols indicate models with frequent stratocumulus (defined as having $\mathcal{R}$>94% more than 15% of the time). The observational uncertainty range is shown in grey, with the shading representing the 25th to 75th quartile and the grey bars the 95%-CI of bootstrapped values. HadGEM2-A is not shown in **b** due to the absence of W output.



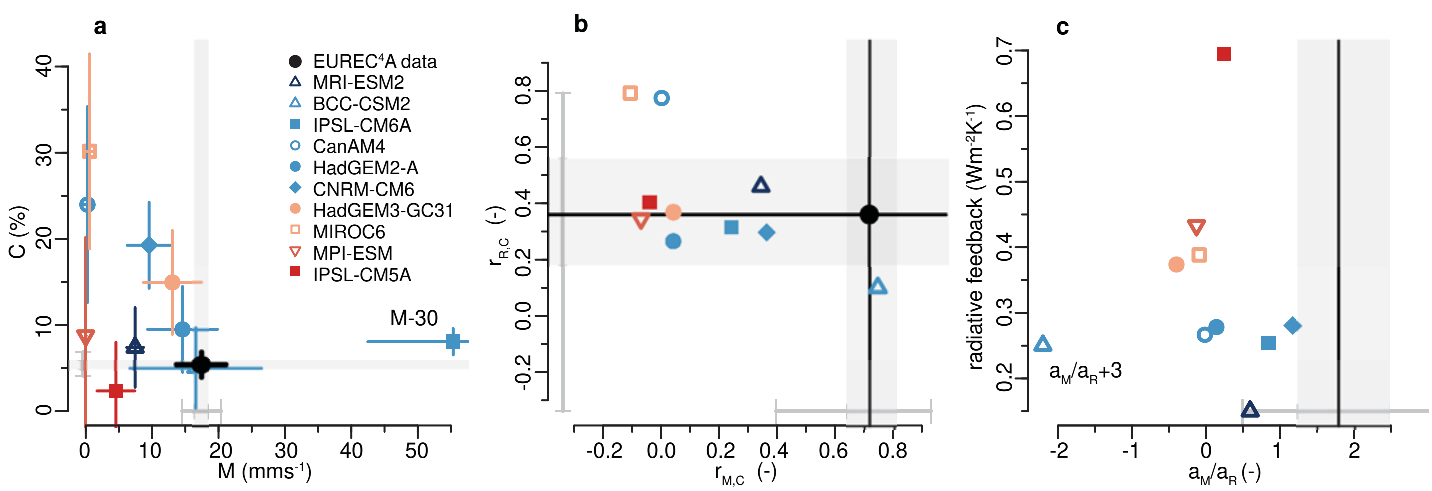
Base state
z / km
inversion
1.5
1.0
0.5
M
E
W
h
a Mixing-desiccation mechanism (β<0)
M++
E++
W
b Mesoscale motion control (β>0)
M++
E++
W++



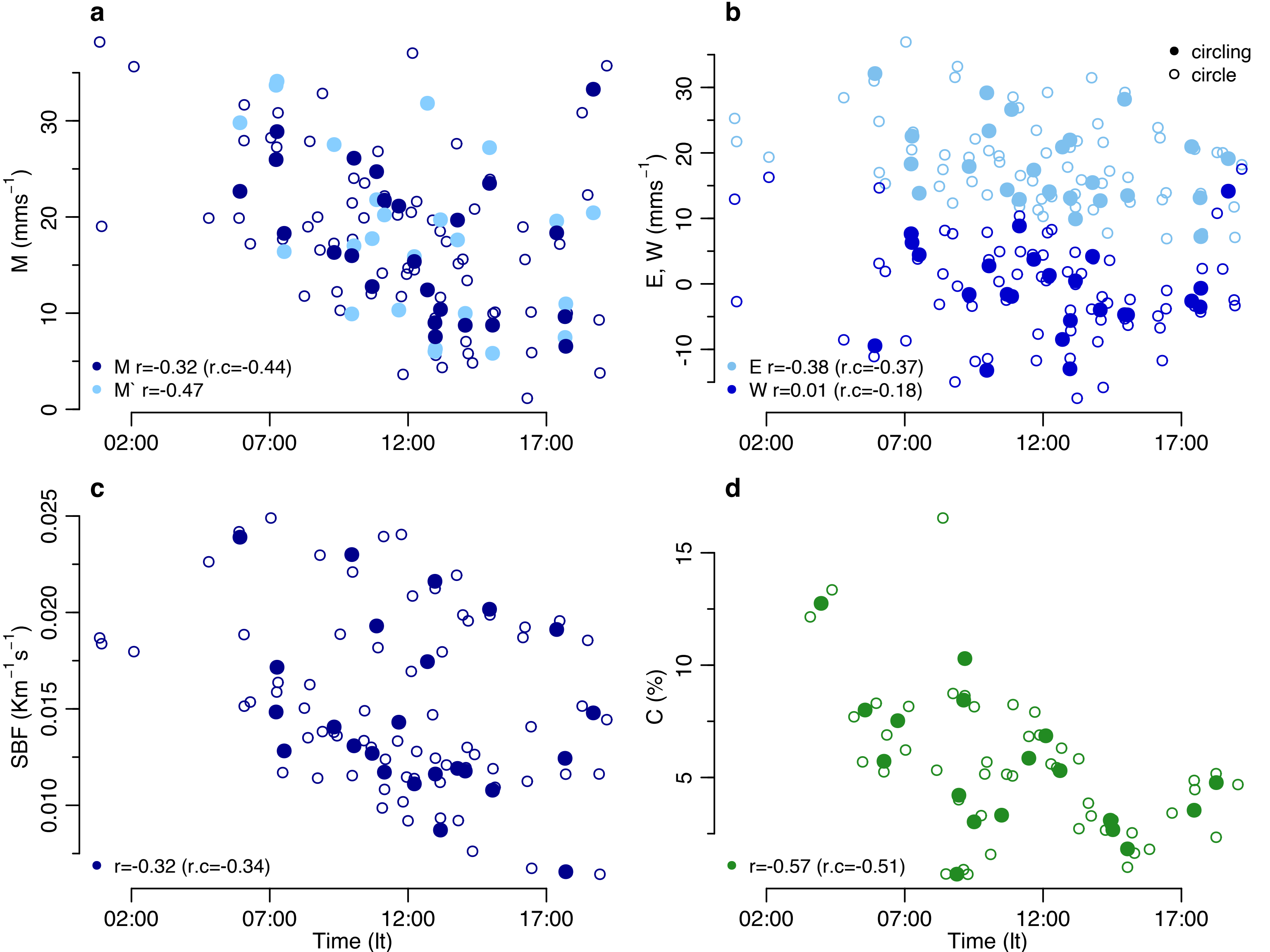
a
M (mms⁻¹)
b
E,W (mms⁻¹)
E
W
c
C (%)
d
R (%)
HALO
ATR
0122
0124
0126
0128
0132
0202
0205
0207
0209
0211
0213
0215
Date (MMDD)



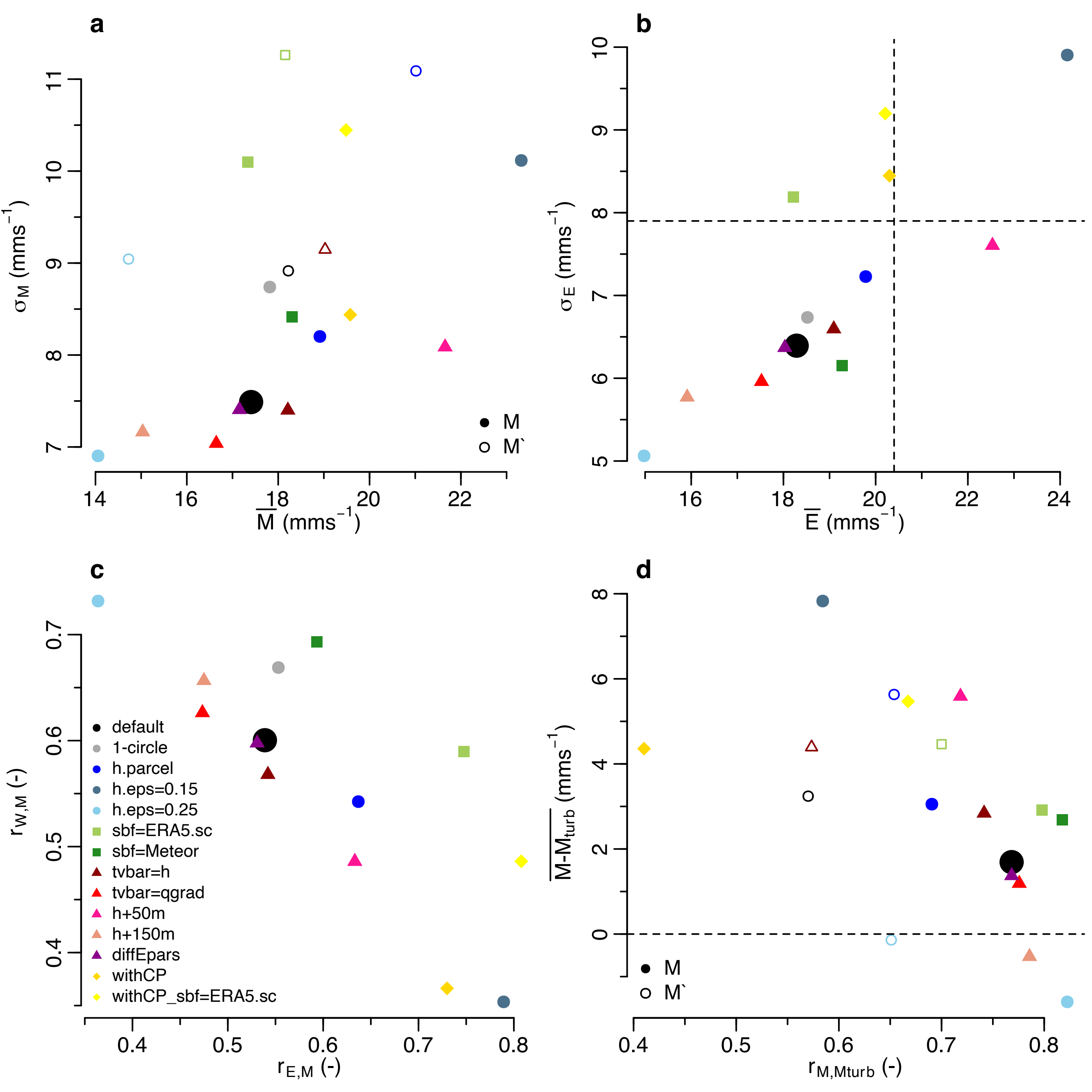
a
r=0.83
C (%)
Reconstructed Ĉ (%)
b
r=-0.08
R (%)
M (mms-1)
c
r=0.72
C (%)
M (mms-1)
d
r=0.36
C (%)
R (%)



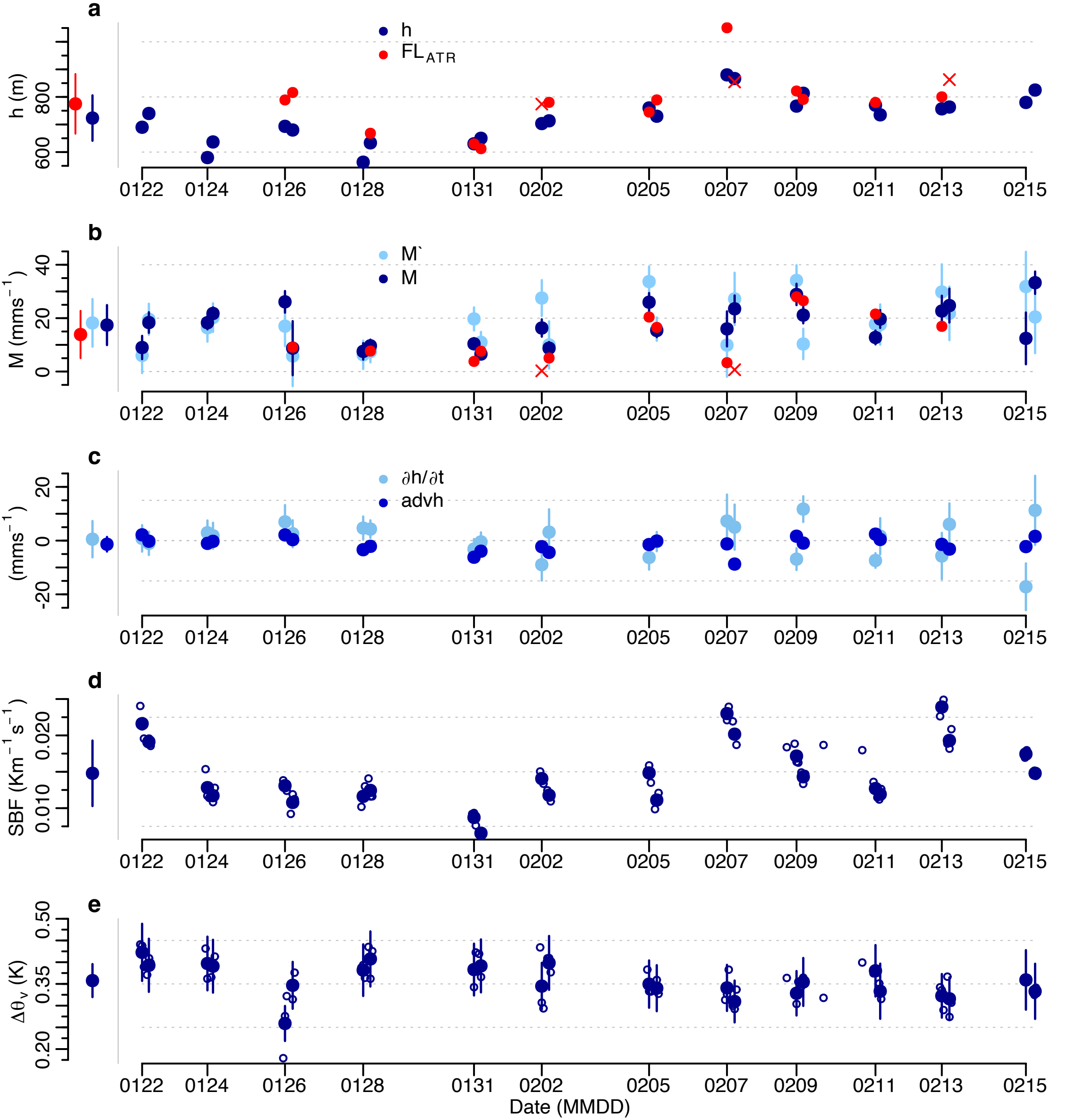
a
b
c
EUREC4A data
MRI-ESM2
BCC-CSM2
IPSL-CM6A
CanAM4
HadGEM2-A
CNRM-CM6
HadGEM3-GC31
MIROC6
MPI-ESM
IPSL-CM5A
M-30
C (%)
M (mms-1)
r_R,C (-)
r_M,C (-)
radiative feedback (Wm-2K-1)
a_M/a_R+3
a_M/a_R (-)

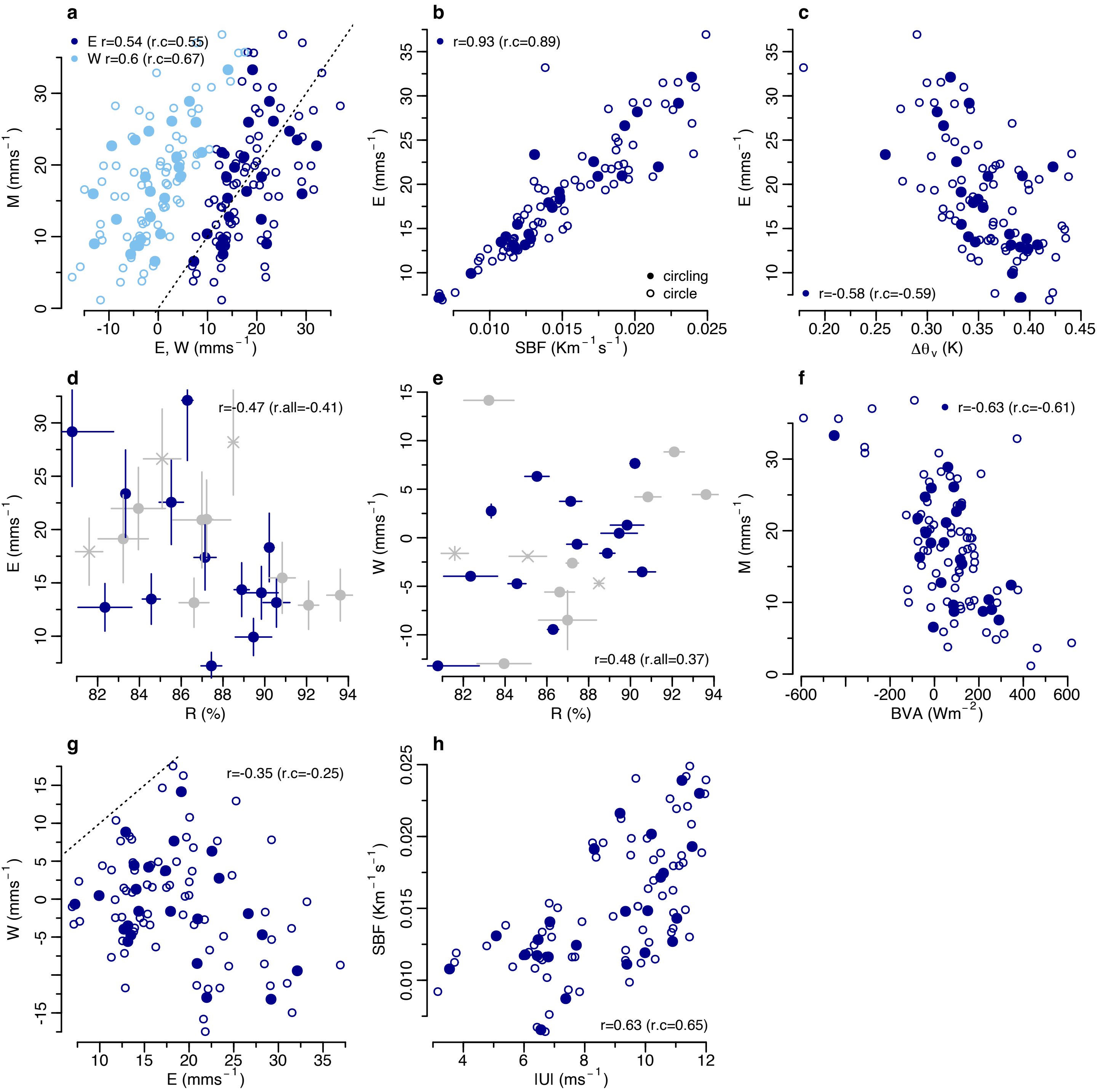


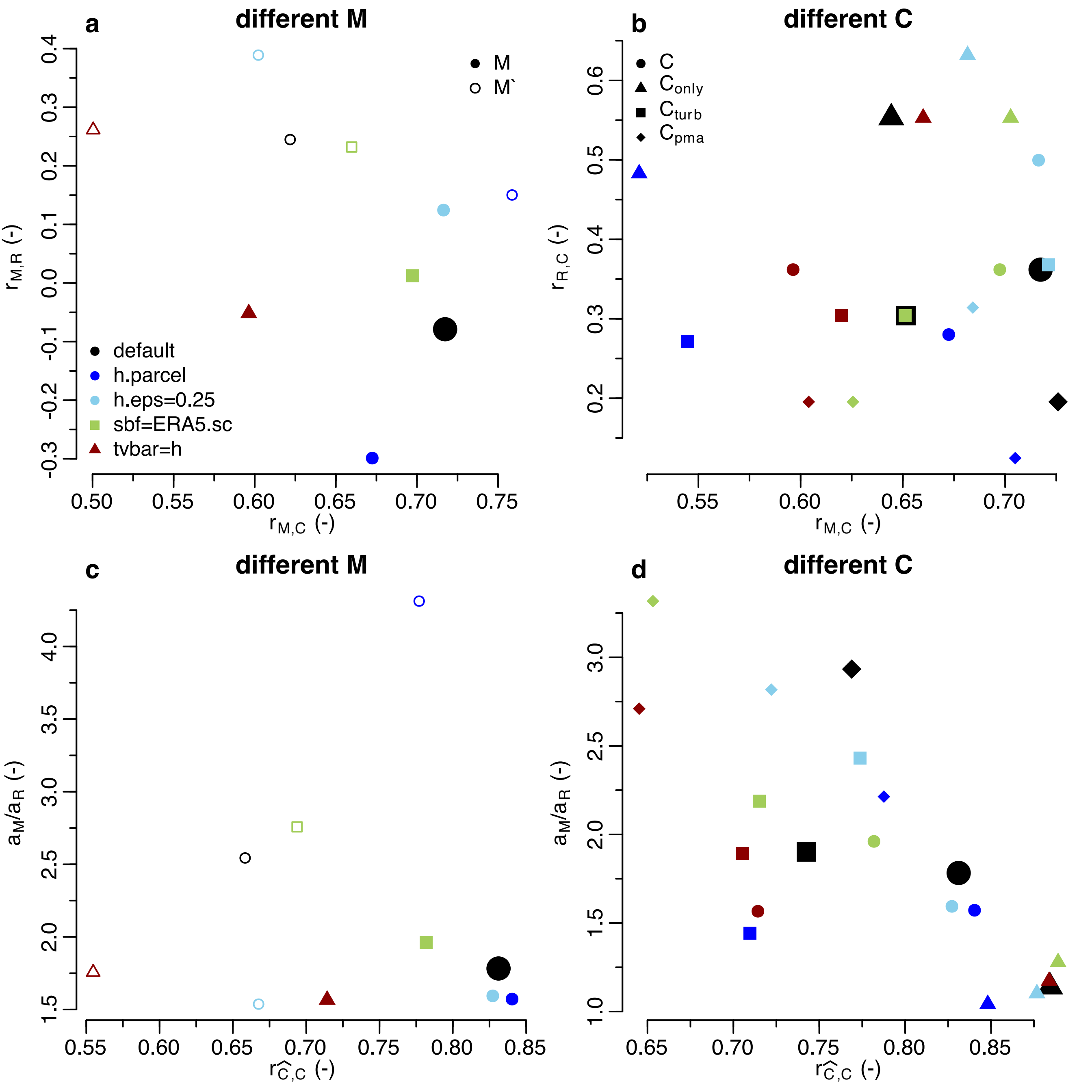
a
b
c
d
circling
circle
M (mms⁻¹)
E, W (mms⁻¹)
SBF (Km⁻¹s⁻¹)
C (%)
M r=-0.32 (r.c=-0.44)
M` r=-0.47
E r=-0.38 (r.c=-0.37)
W r=0.01 (r.c=-0.18)
r=-0.32 (r.c=-0.34)
r=-0.57 (r.c=-0.51)
02:00
07:00
12:00
17:00
Time (lt)

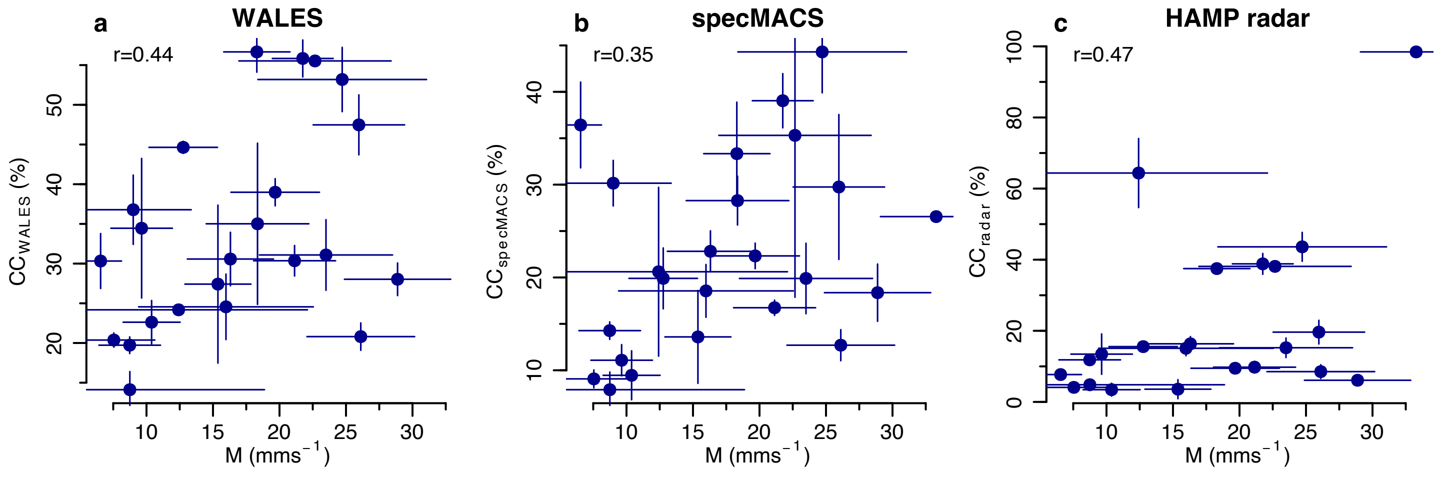


a
b
c
d
σ_M (mms^{-1})
$\overline{M}$ (mms^{-1})
σ_E (mms^{-1})
$\overline{E}$ (mms^{-1})
$r_{W,M}$ (-)
$r_{E,M}$ (-)
$\overline{M\text{-}M_{turb}}$ (mms^{-1})
$r_{M,Mturb}$ (-)
M
M`
default
1-circle
h.parcel
h.eps=0.15
h.eps=0.25
sbf=ERA5.sc
sbf=Meteor
tvbar=h
tvbar=qgrad
h+50m
h+150m
diffEpars
withCP
withCP_sbf=ERA5.sc

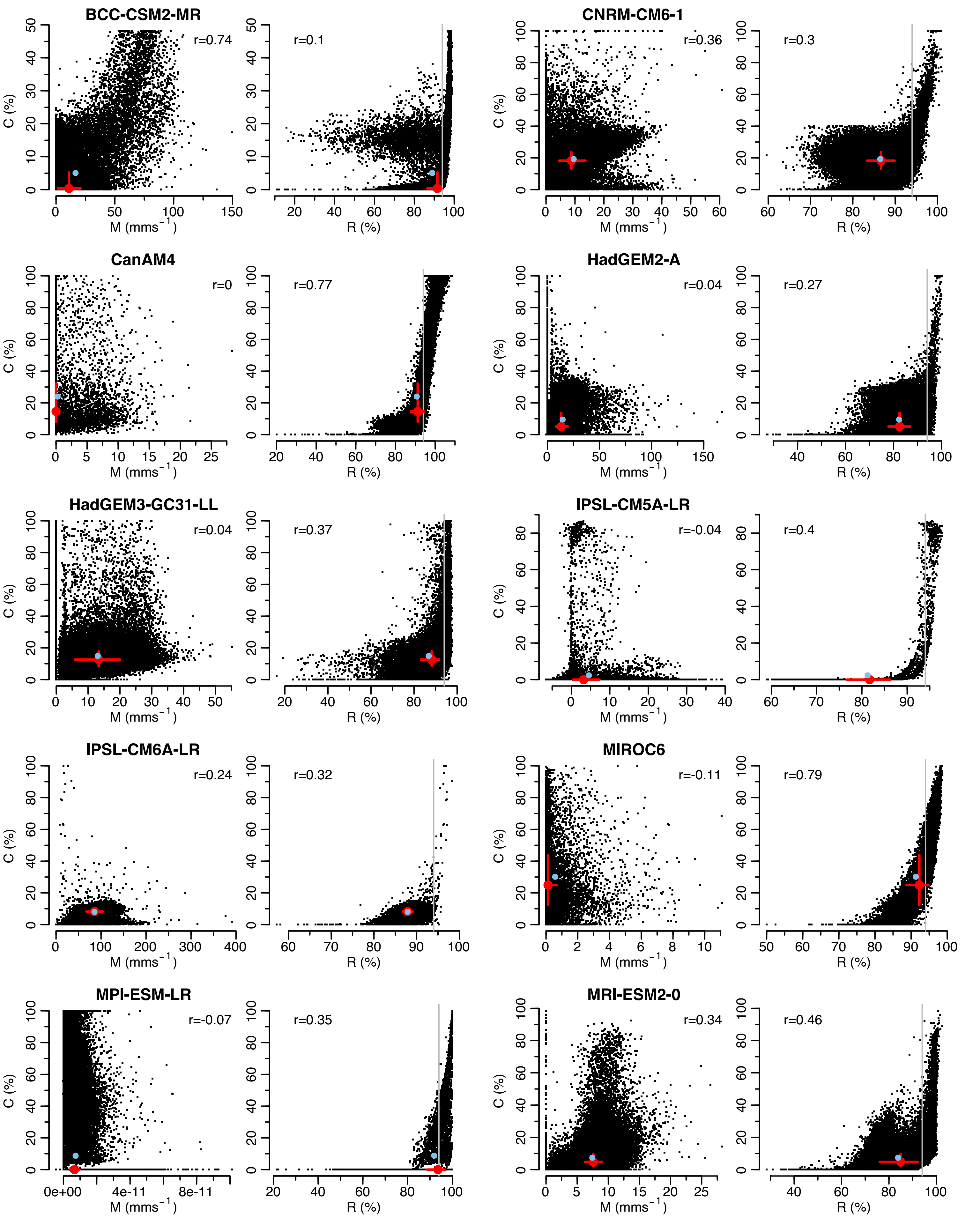

a
h
FL_{ATR}
h (m)
600
800
b
M`
M
M (mms^{-1})
0
20
40
c
∂h/∂t
advh
(mms^{-1})
-20
0
20
d
SBF (Km^{-1}s^{-1})
0.010
0.020
e
$\Delta\theta_v$ (K)
0.20
0.35
0.50
0122
0124
0126
0128
0131
0202
0205
0207
0209
0211
0213
0215
Date (MMDD)


a
E r=0.54 (r.c=0.55)
W r=0.6 (r.c=0.67)
M (mms^{-1})
E, W (mms^{-1})
b
r=0.93 (r.c=0.89)
circling
circle
E (mms^{-1})
SBF (Km^{-1}s^{-1})
c
r=-0.58 (r.c=-0.59)
E (mms^{-1})
$\Delta\theta_v$ (K)
d
r=-0.47 (r.all=-0.41)
E (mms^{-1})
R (%)
e
r=0.48 (r.all=0.37)
W (mms^{-1})
R (%)
f
r=-0.63 (r.c=-0.61)
M (mms^{-1})
BVA (Wm^{-2})
g
r=-0.35 (r.c=-0.25)
W (mms^{-1})
E (mms^{-1})
h
r=0.63 (r.c=0.65)
SBF (Km^{-1}s^{-1})
|U| (ms^{-1})

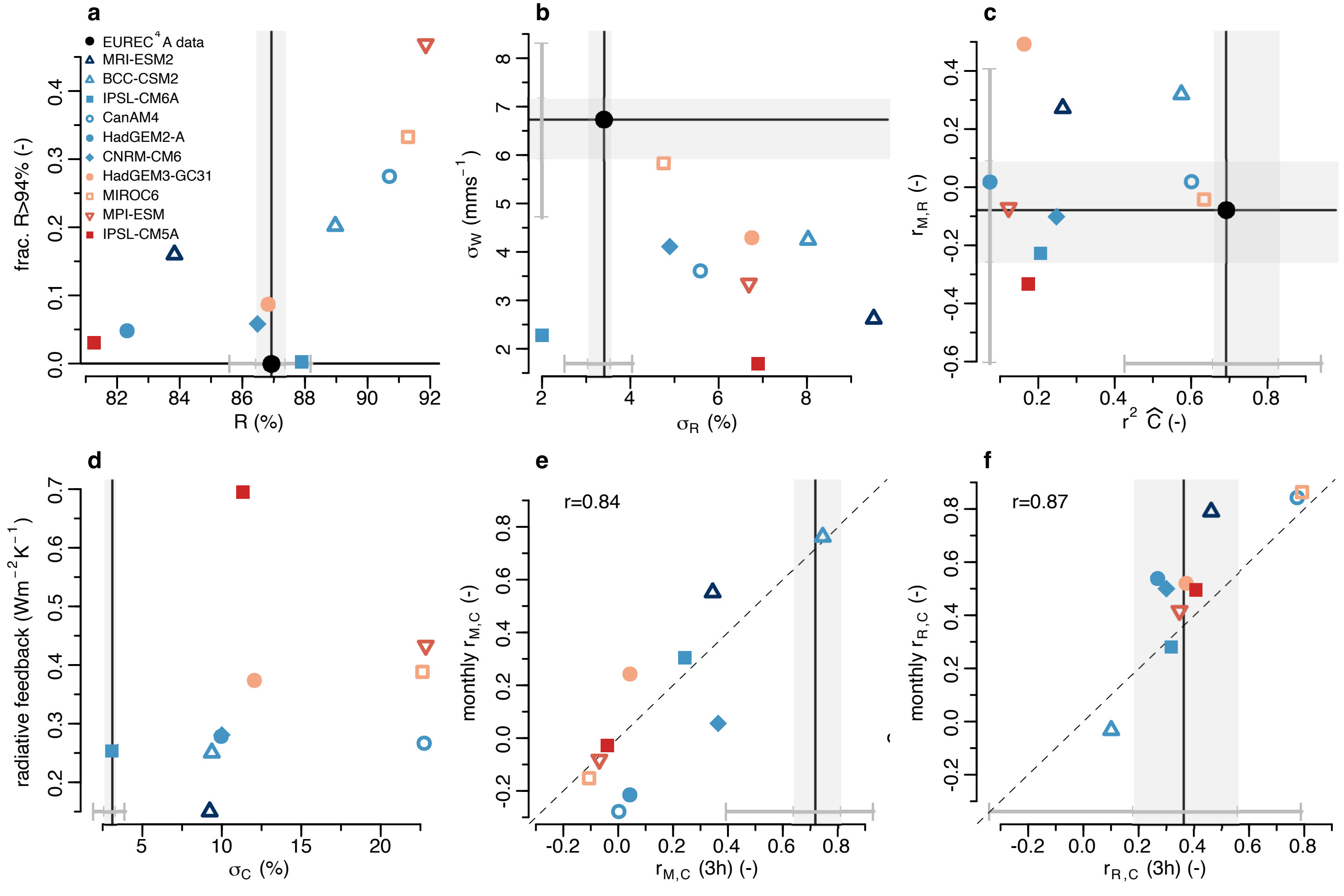

a
different M
b
different C
c
d
M
M`
default
h.parcel
h.eps=0.25
sbf=ERA5.sc
tvbar=h
C
C only
C turb
C pma
r M,R (-)
r M,C (-)
r R,C (-)
a M/a R (-)
r Ĉ,C (-)



a
WALES
r=0.44
CC_{WALES} (%)
M (mms^{-1})
b
specMACS
r=0.35
$CC_{specMACS}$ (%)
M (mms^{-1})
c
HAMP radar
r=0.47
CC_{radar} (%)
M (mms^{-1})



BCC-CSM2-MR
r=0.74
r=0.1
CNRM-CM6-1
r=0.36
r=0.3
CanAM4
r=0
r=0.77
HadGEM2-A
r=0.04
r=0.27
HadGEM3-GC31-LL
r=0.04
r=0.37
IPSL-CM5A-LR
r=-0.04
r=0.4
IPSL-CM6A-LR
r=0.24
r=0.32
MIROC6
r=-0.11
r=0.79
MPI-ESM-LR
r=-0.07
r=0.35
MRI-ESM2-0
r=0.34
r=0.46
C (%)
M (mms^{-1})
R (%)



a
EUREC^{4}A data
MRI-ESM2
BCC-CSM2
IPSL-CM6A
CanAM4
HadGEM2-A
CNRM-CM6
HadGEM3-GC31
MIROC6
MPI-ESM
IPSL-CM5A
frac. R>94% (-)
R (%)
b
σ_W (mms^{-1})
σ_R (%)
c
$r_{M,R}$ (-)
r^2 $\widehat{C}$ (-)
d
radiative feedback (Wm^{-2}K^{-1})
σ_C (%)
e
r=0.84
monthly $r_{M,C}$ (-)
$r_{M,C}$ (3h) (-)
f
r=0.87
monthly $r_{R,C}$ (-)
$r_{R,C}$ (3h) (-)