Geophysical Research Letters

RESEARCH LETTER
10.1029/2022GL099678

**Key Points:**

- Model fits are performed to the spatiotemporal observed cloudiness over all oceans, using a minimal set of predictors and parameters
- Models capture global-mean, spatial, and most of seasonal variability of cloud radiative effects
- Cloud albedo and longwave effect are captured by pressure velocity and its variance, surface temperature, and lower tropospheric stability

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

G. Datseris,
george.datseris@mpimet.mpg.de

**Citation:**

Datseris, G., Blanco, J., Hadas, O., Bony, S., Caballero, R., Kaspi, Y., & Stevens, B. (2022). Minimal recipes for global cloudiness. *Geophysical Research Letters*, *49*, e2022GL099678. https://doi.org/10.1029/2022GL099678

Received 20 MAY 2022
Accepted 13 OCT 2022

**Author Contributions:**

**Conceptualization:** George Datseris, Bjorn Stevens
**Formal analysis:** George Datseris
**Funding acquisition:** Bjorn Stevens
**Supervision:** Bjorn Stevens
**Writing – original draft:** George Datseris, Bjorn Stevens
**Writing – review & editing:** George Datseris, Joaquin Blanco, Or Hadas, Sadrine Bony, Rodrigo Caballero, Yohai Kaspi, Bjorn Stevens



# Minimal Recipes for Global Cloudiness

**George Datseris[1], Joaquin Blanco[2], Or Hadas[3], Sadrine Bony[4], Rodrigo Caballero[2], Yohai Kaspi[3], and Bjorn Stevens[1]**

[1]Max Planck Institute for Meteorology, Hamburg, Germany, [2]Department of Meteorology, Stockholm University, Stockholm, Sweden, [3]Department of Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot, Israel, [4]Sorbonne University, LMD/IPSL, CNRS, Paris, France

**Abstract** Clouds are primary modulators of Earth's energy balance. It is thus important to understand the links connecting variabilities in cloudiness to variabilities in other state variables of the climate system, and also describe how these links would change in a changing climate. A conceptual model of global cloudiness can help elucidate these points. In this work we derive simple representations of cloudiness, that can be useful in creating a theory of global cloudiness. These representations illustrate how both spatial and temporal variability of cloudiness can be expressed in terms of basic state variables. Specifically, cloud albedo is captured by a nonlinear combination of pressure velocity and a measure of the low-level stability, and cloud longwave effect is captured by surface temperature, pressure velocity, and standard deviation of pressure velocity. We conclude with a short discussion on the usefulness of this work in the context of global warming response studies.

**Plain Language Summary** Clouds are important for Earth's climate, because they affect a large portion of the planet's energy balance, and hence its mean temperature. To better understand how the interplay between cloudiness and energy balance would change in a changing climate, a better theoretical understanding of how clouds are distributed over the planet, and how this connects with the state variables of the climate system such as temperature and wind speed, is required. As theoretical understanding is currently limited, in this work we explore the possibility of very simply representing the spatiotemporal distribution of clouds over the whole planet. We believe that these simple representations advance the field in the direction of a conceptual theory of global cloudiness and its impact on the energy balance. We show that the impact of cloudiness on both solar and terrestrial radiation balance can be captured well globally with only a few predictive fields, like surface temperature or vertical wind speed, combined simply and using only three tunable parameters, and without using any supplementary information such as the particular season or location on the planet.

## 1. Introduction

Clouds are one of the most fascinating, important, and complex components of Earth's climate system (Siebesma et al., 2020). Despite their importance, theoretical understanding of what controls planetary-wide cloudiness is largely absent. For example, while we have a good understanding of how clouds form and interact with radiation (Cotton et al., 2014; Houze, 2014; Siebesma et al., 2020), it is difficult to use these theories to make claims about global cloudiness. Earth System Models (ESMs) and other bottom-up approaches do couple simple models of cloud formation to the global circulation. However, so far they have not been proven effective in constraining global cloudiness variability (Sherwood et al., 2020; Zelinka et al., 2020). This makes it difficult to transparently establish links between variability in global cloudiness and Earth's energy balance, or how such a link would change in a changing climate.

Conceptual models could be useful in elucidating how clouds, circulation, and energy balance, are tied together. Existing theoretical work has linked cloudiness to circulation, and most examples of such work focus on particular circulation systems, like the tropical overturning circulation (Betts & Ridgway, 1989; Pierrehumbert & Swanson, 1995), or the Walker cell (Peters & Bretherton, 2005), or individual cyclones (Carlson, 1980). What is missing is a conceptual framework that both closes the top-of-atmosphere energy budget (and hence by necessity considers the planet as a whole), but also includes clouds. A suitable candidate for such a framework would be an energy balance model (Budyko, 1969; Ghil, 1981; North & Kim, 2017; Sellers, 1969) that explicitly represents dynamic cloudiness, likely as an implicit function of circulation measures or other state variables.

In this work we derive simple representations, or "recipes," for global cloudiness, which are simple enough to include in energy balance models, to then help link variations in the energy budget and state variables of such models to variations in cloudiness and vice-versa. These representations therefore need to capture all main features of cloudiness, which are the global mean value, mean seasonal cycle, coarse spatial variability, and the difference between the shortwave and longwave impact of cloudiness. To derive them, we will use a quantitative top-down approach, where global cloudiness is directly decomposed into contributions from several simpler spatiotemporal fields. These fields are the "ingredients" of the recipe, which we refer to simply as predictors (in the sense of statistical predictors). A model useful in theoretical work is one that can explain the most with the least amount of information. In this spirit, the main objective of the present study is to derive informative representations of cloudiness based on a minimum number of predictors.

Similar top-down approaches have been utilized in the context of the empirical cloud controlling factors framework (CCFF) (Stevens & Brenguier, 2009). For tropical low clouds there are several studies summarized in the review by Klein et al. (2017), and see also Myers et al. (2021) for ESMs versus observations. Attention has also been given to the midlatitude cloudiness (a summary of existing work on extratropical cloud controlling factors can be found in Kelleher and Grise (2019) and see also Grise and Kelleher (2021) for ESMs vs. observations). Our approach differs from past empirical approaches in that we fit absolute cloudiness, not anomalies, and our fits are not conditioned on space and time.

Section 2 describes how we define cloudiness, which predictors to consider, how to fit predictor models on observed cloudiness, and how to judge the quality of the fits. Then, Section 3 presents the main analysis and results on how well the models capture cloud albedo and cloud longwave radiative effect. A summary and discussion of potential impact for sensitivity studies concludes the paper in Section 4.

## 2. Fitting Global Cloudiness

### 2.1. Quantifying Cloudiness

To fit any model, a definition of cloudiness that is both quantitatively precise but also energetically meaningful is required. For the shortwave part, we use the energetically consistent effective cloud albedo (in the following, just "cloud albedo"), $C$, estimated using the approach of Datseris and Stevens (2021). In summary, $C$ is defined as

$$C = f\frac{\sqrt{3}(1-g)\tau}{2+\sqrt{3}(1-g)\tau} \tag{1}$$

with $f$, $\tau$, $g$ the cloud fraction, cloud optical depth, and the asymmetry factor of the cloud particle phase function. $C$ as defined above is further normalized to be energetically consistent. $C$ provides a better way to quantify shortwave impact of cloudiness than the shortwave cloud radiative effect (CRE) or just the cloud fraction, because it disentangles co-variabilities with surface properties, it does not directly share the variability of insolation (Ceppi and Nowack (2021) use a cloud albedo for similar reason), while being energetically consistent (Datseris & Stevens, 2021). For the longwave part the CRE (here denoted by $L$) is a good representation of the radiative impact of clouds. From $L$, a cloud effective emissivity can be constructed which could be added to an energy balance model directly similarly to the albedo. Both $C$, $L$ are derived from monthly mean CERES EBAF data (Loeb et al., 2018) using 20 years of measurements (2001/01–2020/12).

### 2.2. Predictors Considered

The predictors considered in this study, listed below, are obtained from the reanalysis of meteorological data (ERA5, Hersbach et al. (2020)) using the same 20 year period as CERES. We initially considered 10 predictors due to their frequent appearance in related studies and their physical connections to clouds. Four, pressure velocity $\omega_{500}$, surface wind speed $V_{\mathrm{sfc}}$, standard deviation of $\omega_{500}$ in hourly timescales $\omega_{\mathrm{std}}$, and the fraction of updrafts in a month $\omega_{\mathrm{up}}$ are dynamic. Four more are thermodynamic, and measures of temperature, like surface specific humidity and temperature $q_{\mathrm{sfc}}$, SST, and to some extent $q_{700}$ and $q_{\mathrm{tot}}$ (total column water vapor). Two more measures targeted the lower tropospheric stratification, namely estimated inversion strength EIS and the estimated cloud top entrainment index ECTEI. Of these $q_{\mathrm{sfc}}$ was discarded as it did not differ substantially from other predictors, and to our knowledge $\omega_{\mathrm{up}}$ is used for the first time in this context.

Regarding the use of many of these predictors in past studies $\omega_{500}$ is known to be important for both shortwave and longwave CREs (Bony et al., 1997, 2004; Norris & Iacobellis, 2005; Norris & Weaver, 2001), and EIS, $V_{sfc}$, SST, $q_{700}$ have been used to fit cloud cover anomalies in a variety of regimes, see for example, Klein et al. (2017); Kelleher and Grise (2019) and references therein for a more detailed discussion. ECTEI, which is much less common than the rest of the predictors, was introduced by Kawai et al. (2017) as an improvement over EIS that includes a contribution from $q_{700}$ (and hence further motivates including $q_{700}$ as a predictor). Both $q_{700}$, $q_{tot}$ are a proxy for the moisture of an atmospheric column, and expected to be relevant when fitting $L$. In our analysis however, $q_{700}$ gives consistently better fits when used in place of $q_{tot}$, keeping all other aspects fixed (not shown) and hence $q_{tot}$ was also discarded.

Unlike the other predictors, $\omega_{up}$ is useful because it is bounded in [0%, 100%], like $C$, and given that we fit absolute values instead of anomalies, it does not penalize the fits with negative values (that exist in $\omega_{500}$). It can also be used as a statistical weight to distinguish between regions of large scale subsidence, see for example, Bony et al. (1997). $\omega_{500}$, which can be thought of as a simple quantifier of storminess, has been shown to be a useful predictor of cloudiness by Norris and Iacobellis (2005) due to the nonlinear connection between vertical motion and cloud generation. Another argument favoring $\omega_{std}$ is that it relates cloudiness with the moisture of the air column better than $\omega_{500}$, see Section 3.3.

### 2.3. Fitting Process

Let $Y$ be a measure of cloudiness ($C$ or $L$ from Section 2.1) and $X_i$ be some predictor fields, for $i = 1, \ldots, n$,. $Y$, $X_i$ are global spatiotemporal fields. We assume that with sufficient accuracy we can write

$$Y \approx M = f(X_1, \ldots, X_n; p_1, \ldots, p_m) \overset{\text{e.g.}}{=} p_1 X_1 + p_2 X_2 + p_3 X_1 X_2 \tag{2}$$

with $p_j$, for $j = 1, \ldots, m$ some parameters to be estimated (all $p_j \in \mathbb{R}$). In the following we call $f$ the "cloud fitting function" and $M$ the "model." Naturally, different forms for $f$ and/or sets of predictors will yield a better fit for $C$ or $L$ respectively, as each captures different aspects of cloudiness. Given a specific form for $f$, and a set of predictors $X_i$, the parameters $p_j$ of the model are estimated via a standardized nonlinear least square optimization (Levenberg, 1944; Marquardt, 1963). The minimization objective is the squared distance between $Y$ derived from CERES observations, and $M$ produced by Equation 2. Details on the data pre-processing before doing the fits are provided in the Supporting Information.

This approach of fitting models with free parameters to observed data is similar to the CCFF, but there three important differences with previously published CCFF studies. First, we fit absolute cloudiness, not anomalies, and hence the mean value of $Y$, and its seasonal cycle, must be captured by the fit. The importance of capturing the mean value and mean seasonal cycle is further enforced by the fact that the inter-annual variability of cloudiness is small in decadal timescales (Stevens & Schwartz, 2012; Stephens et al., 2015), and hence the mean and mean seasonal cycle captures the majority of the signal. Second, to capture the mean $f$ is allowed to be nonlinear. Were it restricted to a linear function it would not be naturally bounded, which strikes us as unphysical given that important aspects of cloudiness, such as albedo, are bounded. Another argument behind allowing nonlinear functions is that, generally speaking, a theory of cloudiness should be able to predict cloudiness over a broad range of different climatic states, not just small deviations from a reference climate (which justifies using a linear framework). Third, we fit across all available space and time without conditioning on space or time. One fit is applied to the cloudiness field globally. The motivation for this is to more closely reflect physical laws governing climate, which should not depend on location. Hence, if indeed cloudiness is a functional of other climate state variables, this functional must be unique and not depend on space or time.

### 2.4. Quantitatively Measuring Fit Quality

To quantify fit quality with an objective measure that is independent of what predictors are used, we chose the normalized root mean square error, defined as

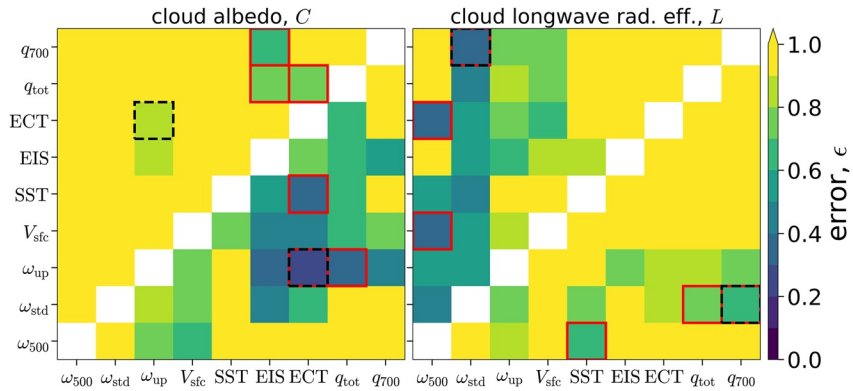


**Figure 1.** Error in temporally and zonally mean cloudiness (lower-right triangle of heatmap), and error in mean seasonal cycle (upper-left triangle of heatmap), as a function of which predictors of the $x$ and $y$ axis combine into a linear model $f = p_1X_1 + p_2X_2$ for fitting cloud albedo (left plot) or longwave cloud radiative effect (right plot). Red outline highlights the three combinations with the lowest error in each category, while black dashed outline highlights the combination with lowest error overall (by multiplying the two errors). It is possible that $e > 1$ because we are fitting without intercept. ECT stands for ECTEI (estimated clout top entrainment index).

$$\epsilon(Y, M) = \sqrt{\frac{\sum_n (Y_n - M_n)^2}{\sum_n \left(Y_n - \bar{Y}\right)^2}} \tag{3}$$

with $Y$, $M$ as in Equation 2, $\bar{Y}$ the mean of $Y$ and $n$ enumerates the data points. This error measure is used routinely in for example, spatiotemporal timeseries prediction (Isensee et al., 2019), and is agnostic of the values of $Y$, $M$ that can compare fit quality across different ways of fitting. If $\epsilon > 1$ the mean value of $Y$ is a better model than $M$ (equivalently, the variance of the observations is smaller than the mean square error between fit and observations). There are several ways to compute $\epsilon$: on full spatiotemporal data, on zonally and temporally averaged data, or on the seasonal cycles of tropics (0°–30°) and midlatitudes (30°–70°). Each measure highlights a different aspect of fit quality and all measures were taken into account when deciding the best fits.

## 3. Results and Discussion

In this section we present the "best" fits for cloud albedo $C$ and longwave CRE $L$. The "best" fits are the most minimal fits, that are in accord with physical reasoning, while also providing low values for $\epsilon$. Only requirement that the error $\epsilon$ is small is objective. The other requirements have at least a partly subjective nature. Additionally, fits that use simpler predictors, that can be more straightforwardly represented in a conceptual framework, are preferred: if two fits have approximately equal error $\epsilon$, but one uses a simpler predictor (e.g., surface temperature SST vs. atmospheric specific humidity $q_{700}$), the first fit is "better".

### 3.1. Two Predictor Linear Model

A simple, yet non-trivial model for the cloud fitting function $f$ is one that combines two predictors and two free parameters in a linear manner: $f = p_1X_1 + p_2X_2$. Even if this model does not yield a good fit for cloudiness, it is advantageous to start with it nevertheless. There are 36 combinations of all possible predictors of Section 2.2. Examining each of these possibilities already highlights which predictors are worth a closer look and for which measure of cloudiness. Figure 1 showcases two different error measures (error in temporally and zonally mean cloudiness, and median of errors in seasonal cycle of cloudiness), and how these errors depend on which predictors are used for the linear fit.

The majority of combinations result in low fit quality ($\epsilon \geq 0.9$). Nevertheless, Figure 1 reveals some useful information. For $C$, a measure of the inversion strength is necessary for a decent fit and the combination of $\omega_{up}$ and ECTEI result in the best case scenario. For $L$, the most important predictor seems to be $\omega_{std}$, which gives decent

fits in both space and time for a wide selection of second predictors (while $\omega_{500}$ gives decent fits only in time). A second important predictor for $L$ seems to be $q_{700}$ or SST.

### 3.2. Best Fit for Cloud Albedo $C$

While it is already clear in the literature that $\omega_{500}$ is an important predictor for shortwave impact of clouds (Section 2.2), the fact that $\omega_{up}$ performs so much better in a linear model hints that the bounded nature of albedo, $C \in [0\%, 100\%]$, is important. Negative predictor values yield low fit quality and also penalize fitting well positive values. One way to counter this would be to use $\omega_{up}$ as a probability weight multiplying other predictors. An alternative would be to use appropriate nonlinear functions of the more basic $\omega_{500}$. Regardless of this choice, a measure of the inversion strength must also be included in the model, as it is necessary to capture the important contribution of low clouds (here ECTEI).

A model that satisfies all these requirements, and achieves the best fit, is

$$C = 50p_1\left(\tanh\left(p_2\omega_{500} + p_3\mathrm{ECTEI}\right) + 1\right) \tag{4}$$

where we used the nonlinear function $x \to 50\,(\tanh(x) + 1)$ to map predictors to [0%, 100%]. The results of the fit (i.e., estimating the parameters $p_1, p_2, p_3$ that give least square error between Equation 4 and the observed CERES $C$) are in Figure 2. The model fit captures most main features of cloud albedo and achieves $\epsilon = 0.54$ over the full space and time, $\epsilon = 0.19$ in the zonal and temporal average, and $\epsilon = 0.65$ in seasonal cycle. The shortwave CRE (which in our study is simply the multiplication of $C$ with the insolation $I$, and then averaging), is 57.1 W/m$^2$ in CERES and 57.45 W/m$^2$ when using the model fit. The main deficiencies of the model fit are about capturing temporal variability in the extratropics. Time correlation there is notably small, and extratropical seasonal cycles of $C$ have much larger mean error than tropical cycles ($\epsilon = 0.8$ vs. 0.37). These deficiencies however could also stem from the strong temporal variability of the high reflectivity of underlying ice affecting CERES measurements. The fit also performs poorly near the coasts overall (and particularly over the western coasts of Africa and the Americas) where there are more clouds in the model than observed. This may be because the model does not capture the coastal clearing associated with strong subsidence, that often suppresses the depth of the boundary layer and formation of cloudiness in situations that would otherwise be favorable for cloud formation (Dadashazar et al., 2020; Zhang et al., 2009).

The inclusion of the parameter $p_1$ is necessary, because in observations cloud albedo does not saturate to 100%, but to much lower values (see Figure 2). We also note that using EIS instead of ECTEI in the model decreases fit quality significantly, because, while EIS and ECTEI both capture subtropical low cloud albedo well, only ECTEI captures well the low clouds in the midlatitudes. Thus, as suggested by Kawai et al. (2017), ECTEI is indeed an improvement over EIS. Another reason for favoring ECTEI is that it encodes the effects of humidity changes which are known to be important for the transition from stratiform to broken low cloud regimes (e.g., Bretherton and Wyant (1997)).

Adding more predictors increases fit quality only slightly. for example, adding a factor $p_4 V_{sfc}$ inside the tanh function decreases time and zonal mean error to $\epsilon = 0.18$ from $\epsilon = 0.19$ and seasonal cycle error to $\epsilon = 0.6$ from $\epsilon = 0.65$, as well as captures hemispheric asymmetries in $C$ slightly better. That the decrease in error is so small gives confidence that the basic physics governing cloud albedo are already captured by Equation 4. Further fine-tuning of the model risks over-fitting, and even then the additional level of detail aims for a level of precision that seems over precise given the intended purpose of the model.

The middle row of Figure 2 provides some insights on the contribution of each predictor. Both ECTEI and $\omega_{500}$ contribute to midlatitude cloud albedo, but ECTEI slightly more so. In the tropics $\omega_{500}$ contributes the albedo of the convective regimes (ITCZ, Maritime Continent), and ECTEI the albedo of the low stratocumulus decks (subsidence regions). ECTEI is in some sense a more important predictor than $\omega_{500}$, because if we set explicitly $p_2 = 0$ in Equation 4, we get lower error of $\epsilon = 0.7$ in full space and time, versus the error of $\epsilon = 0.9$ we would get if we set explicitly $p_3 = 0$ instead. Alternative models to Equation 4 can give similar results using $\omega_{up}$ instead of $\omega_{500}$. For example, using $f = p_1\omega_{up} + p_2\mathrm{ECTEI}\,(1 - \omega_{up})$ provides similar, but slightly worse, fit quality with $\epsilon = 0.57$ over full space and time and $\epsilon = 0.23$ over time and zonal mean. However, $\omega_{500}$ is a simpler predictor than $\omega_{up}$, and hence a model with $\omega_{500}$ is more minimal (and thus, "better"). The same could be said for EIS instead of

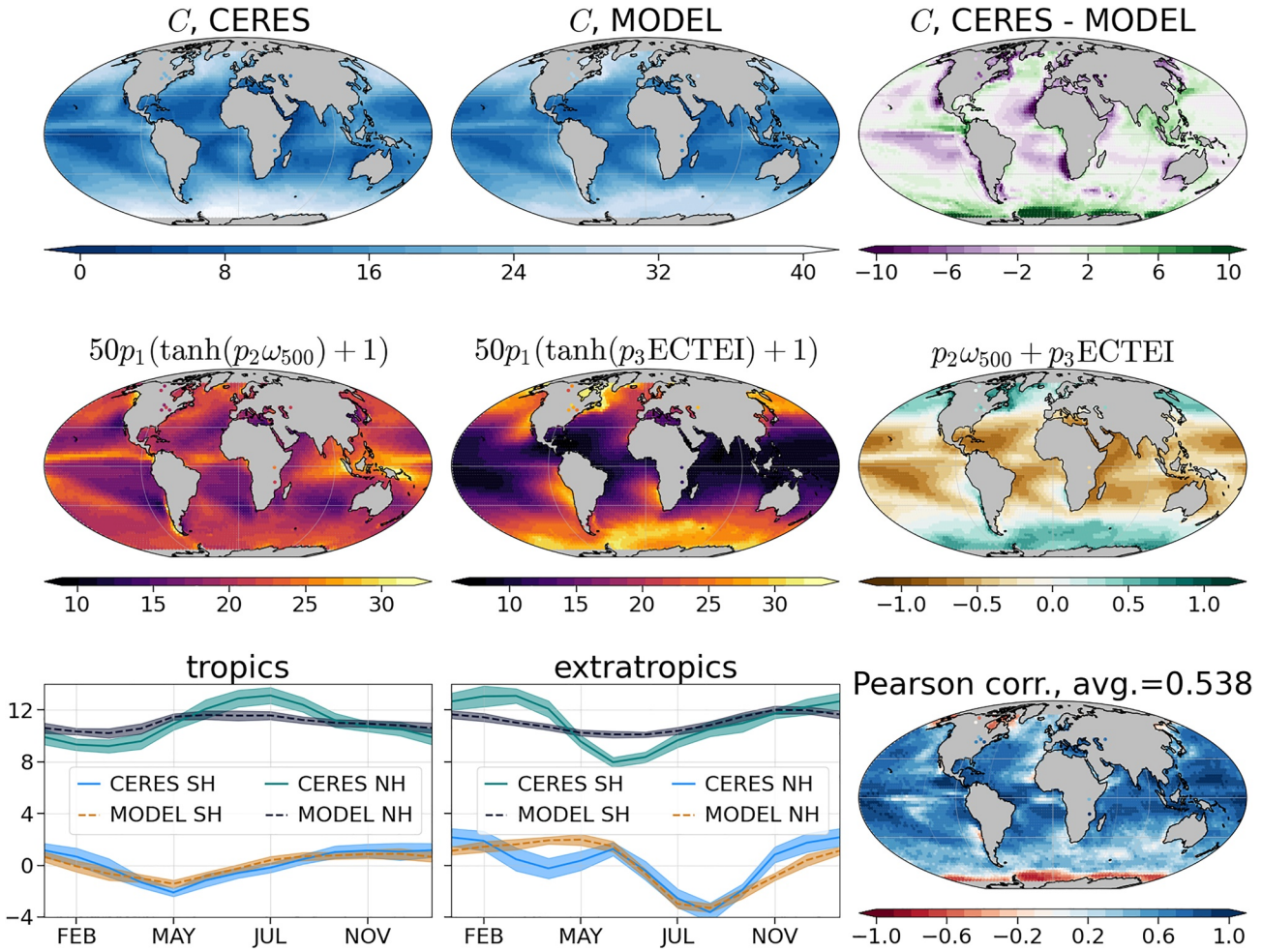


**Figure 2.** Results of fitting cloud albedo $C$ (units of %) with the simple model of Equation 4. First row are time-averaged maps. See also Figure 4 for a zonally averaged version. Second row are the contributions of different terms in the model. Third row shows how well the model captures temporal variability. First two panels are the mean seasonal cycles (with semi-transparent bands noting the standard deviation around each month) in the tropics (0–30°) and extratropics (30–70°). The mean value of all cycles has been subtracted, and some cycles are offset for visual clarity. The subtracted mean values of the cycles are presented in a table in the Supporting Information. The bottom right panel is a map of the Pearson linear correlation coefficient between the timeseries of the model and CERES data at each grid point. Units of $\omega_{500}$ in Pa/s and ECTEI in K, and $p_1 = 0.4$, $p_2 = 6.87$, $p_3 = 0.08$. We multiply $\omega_{500}$ with $-1$ before any analysis so that $\omega_{500} > 0$ means updrafts.

ECTEI, so that in some sense our study provides a form of global validation for studies of cloud amount based on these simpler predictors (see e.g., Bony et al. (2020)).

### 3.3. Best Fit for Longwave Cloud Radiative Effect *L*

Fitting $L$ is more complex for at least two reasons. First, the longwave effect of a cloud depends strongly on the infrared opacity, and hence moisture content, of the atmospheric column overshadowed by the cloud. Moisture content though is, at least partly, controlled by temperature. Warm and humid atmospheres are already almost opaque to longwave radiation, and hence the presence of a cloud would make little difference. In contrast, in a cold and dry atmosphere a cloud would bring a lot of extra absorption of outgoing longwave radiation and hence large $L$. Second, cloud height largely determines its effective emissivity (as cloud height sets its temperature), but does not have a significant effect on cloud albedo (keeping all other factors fixed).

These considerations likely explain why we were not able to find a model that had as good of a fit for $L$ as it had for $C$ when restricting the model to using at most two predictors (although it should be noted that ECTEI is a predictor including information from more than one state variables). After an analysis of several different linear and nonlinear combinations, the "best" model we could construct was of the form

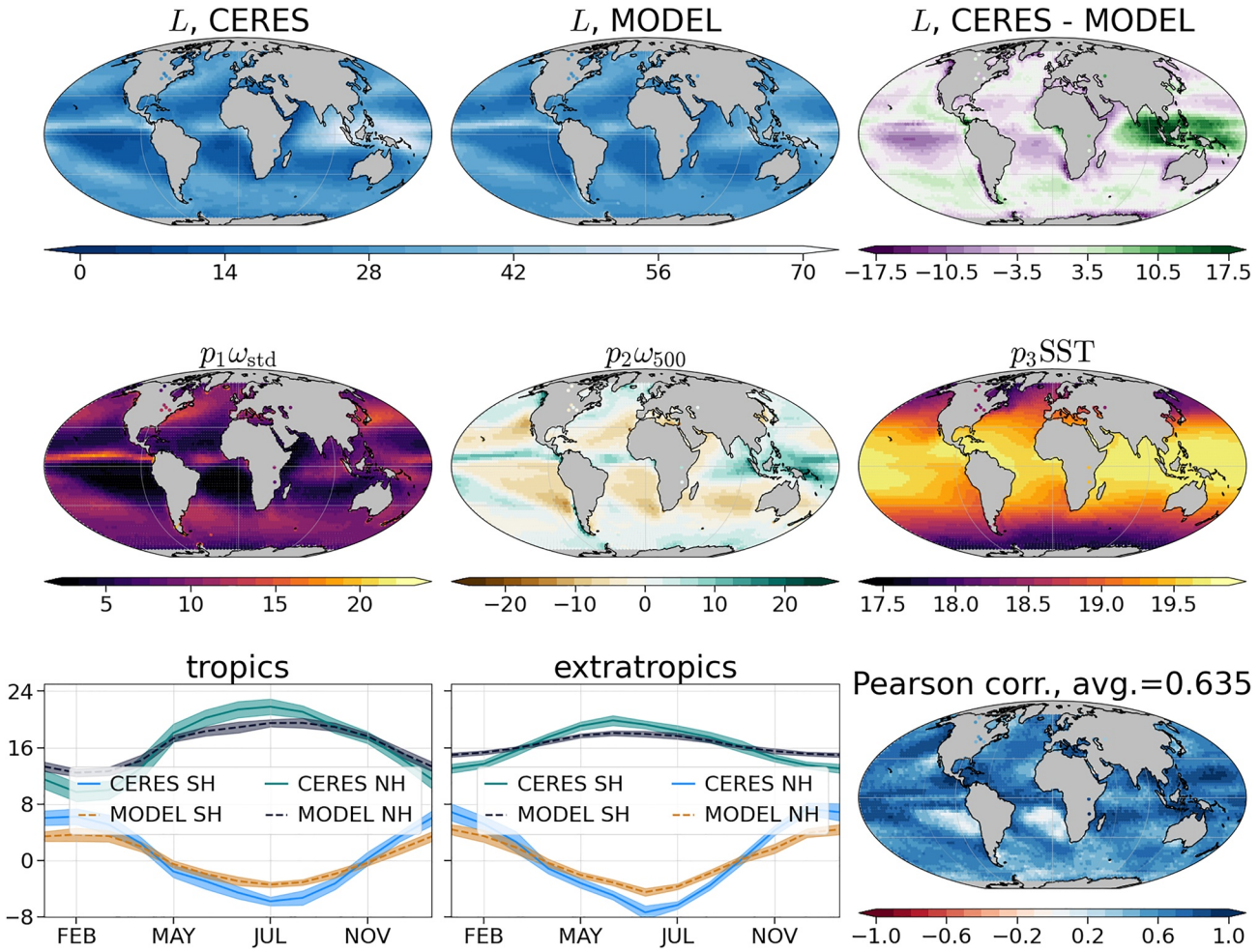


**Figure 3.** As in Figure 2 but now for longwave cloud radiative effect $L$. Units of $L$ in W/m$^2$, $\omega_{500}$, $\omega_{std}$ in Pa/s, SST in K, and $p_1 = 42.68$, $p_2 = 208.9$, $p_3 = 0.06558$.

$$L = p_1\omega_{std} + p_2\omega_{500} + p_3\text{SST} \quad (5)$$

(notice how Equation 5 has 0 intercept by force). The results of the fit are in Figure 3. The fit captures all main features of $L$. The fit errors are $\epsilon = 0.63$ over full space and time, $\epsilon = 0.46$ in time and zonal mean and $\epsilon = 0.41$ in mean seasonal cycle. The mean LCRE is 27.27 W/m$^2$ in CERES and 27.30 W/m$^2$ in our model fit. A notable deficiency of the fit is that $L$ of the Maritime Continent is significantly underestimated (this can be attributed almost entirely to the difference in magnitude of $\omega_{std}$ between the ITCZ and the Maritime Continent, see middle row of Figure 3). Spatial variability is captured worse for $L$ versus $C$, but temporal variability is captured better. A factor that contributes to this is that the temporal variability of $L$ is much simpler than it is for $C$ (e.g., relative power of 12-month periodic component is much larger in $L$ timeseries, leading to simpler seasonal cycle temporal structure).

We now give some physical justification for the choice of predictors. Monthly mean $\omega_{500}$ is a proxy for cloud height (persistent updrafts and with larger magnitude should result in higher clouds). The surface temperature SST is a proxy for the emissivity of the air column without a cloud, because the potential total moisture content of atmospheric columns is, as a first approximation, a monotonically increasing function of temperature. Using $q_{700}$ instead of SST captures spatial variability worse but improves the representation of temporal variability. Given that SST is a more basic predictor than $q_{700}$, and is directly represented in conceptual energy balance models, SST is preferred. Furthermore, and as was the case with $C$, adding more predictors, or additional nonlinear terms of existing predictors such as a factor $p_4\omega_{std}$SST, increases fit quality but only slightly.

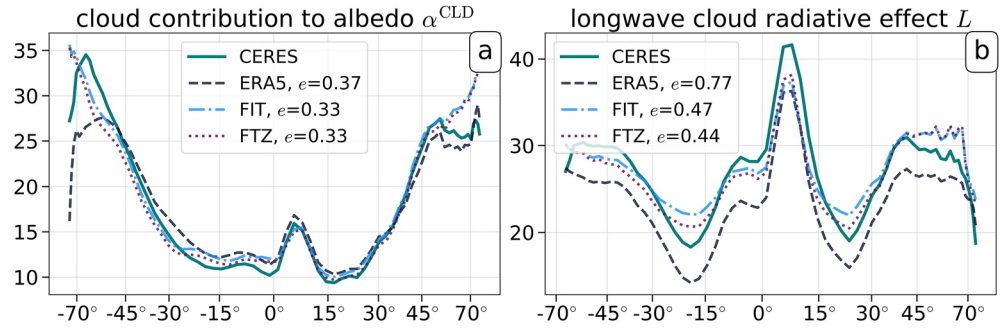


**Figure 4.** Temporally and zonally averaged data (and their errors $e$, Equation 3, vs. the CERES curve) of CERES, our model fits, and direct ERA5 output for (a) the cloud contribution to atmospheric albedo $\alpha^{CLD}$ and (b) the longwave cloud radiative effect $L$. In (a), "FIT" is a fit over temporally averaged data (no time information can be used), and "FTZ" is a fit over temporally and zonally averaged data. In (b), "FIT" is over all space and time, and "FTZ" is as before.

Interestingly, $\omega_{std}$ is the most important predictor for $L$. Even though $\omega_{500}$ captures a broader range of values (~40 vs. the ~30 of $\omega_{std}$), absence of $\omega_{std}$ significantly lowers fit quality in all combinations of cloud fitting functions $f$ and predictors we tested, even when including $\omega_{500}$ in all of them. The spatial structure of $\omega_{std}$ is the most similar to the spatial structure of $L$, with the main difference being that for $\omega_{std}$ the peak values in the tropics and extratropics have equal magnitude, while for $L$ the tropics peak values have 33% more magnitude. Hence, some other predictor must lower the extratropical magnitude of $\omega_{std}$, and here this role is fulfilled by SST in Equation 5 (or $q_{700}$, if one uses it instead of SST).

A physical connection between $\omega_{std}$ and $L$ can be thought of as follows: persistent updrafts, that are captured by $\omega_{500}$, lead to a moist atmosphere and hence weak $L$, mostly irrespectively of cloud height. On the other hand, consistent pumping of air up and down (high $\omega_{std}$, but almost zero $\omega_{500}$) would leave the atmosphere dry (for at least half the time), but the formed clouds would linger longer above the dry atmosphere and have a disproportionately strong effect, yielding high $L$. In the midlatitudes both $L$ and $\omega_{std}$ have their latitudinal maximum in the middle of the Ferrel cell (40–45°), where $\omega_{500} \approx 0$. Of course, monthly mean $\omega_{500} \approx 0$, but in the hourly timescale there is a lot of vertical motion, as captured by the high values of $\omega_{std}$. This reflects the fact that the center of the Ferrell cell coincides with the center of the midlatitude storm tracks. In the tropics, $\omega_{500}$ and $\omega_{std}$ have little differences in their latitudinal structure.

### 3.4. Comparison With ERA5 and Reduced Data

For obtaining reference values of the errors we report here, we also compare the outcome of our analysis with using direct ERA5 radiation output to measure $C$ or $L$. Calculating $L$ is straightforward, however, we cannot compute the energetically consistent effective cloud albedo from ERA5, because it requires cloud optical depth, a field not exported by ERA5. Instead, we can compute the cloud contribution to atmospheric albedo $\alpha^{CLD}$ (specifically, Equation 3 from Datseris and Stevens (2021)), which has only small differences with $C$. $\alpha^{CLD}$ also has the downside of not having a time dimension due to absence of sunlight for large portions of the data (Datseris & Stevens, 2021).

We also present fits and their errors for fitting reduced data directly, specifically temporally and zonally averaged data. Fitting reduced data increases fit quality, because this case neglects higher-order effects that contribute to for example, zonal or temporal structure. If, however, the fit quality increases only slightly, that gives confidence that the basic connections captured by our models are indeed the most important ones and hence also dominate full spatiotemporal variability. The results are in Figure 4.

Two conclusions can readily be drawn: (a) our fits have smaller error $\epsilon$ than does the cloudiness inferred from ERA5 radiation output, (b) fitting the simplified version of temporally and zonally averaged data increases fit quality only slightly, further validating the fit quality. Additionally, the best parameters of the fits change little when doing the zonal-only fit (e.g., for $C$, parameters become $p_1 = 0.4$, $p_2 = 8.25$, $p_3 = 0.077$ vs. those reported in Figure 2). This means that the contribution of each predictor does not change fundamentally in the reduced

version, giving us even more confidence that the simple models of Equations 4 and 5 capture the basic physics well.

### 3.5. Fit Stability

The exact parameter values $p_j$ in Equations 4 and 5 have been derived from fitting on current climate and their values may change for different climates. How much $p_j$ depends on climate state quantifies the "stability" of a fit; the less they change, the more accurately the fit captures the basic physics. Thankfully, the variability of $p_j$ is rather small. For example, we performed the fit for each hemisphere individually (recall that for $C$ the fit over the whole globe yielded $p = \{0.4, 6.87, 0.08\}$). As far as circulation patterns and cloudiness distributions are concerned, the two hemispheres have significant differences. Nevertheless, fitting them individually yields {0.38, 7, 0.07} for north and {0.41, 6.89, 0.086} for south. In the Supporting Information we perform a much more detailed stability analysis for $p_j$ that leads to similar conclusions: that Equations 4 and 5 capture the basic physical connections that govern global cloudiness.

## 4. Conclusions

The main goal of this work was to identify ways one can accurately represent observed global cloudiness using as few and as simple components as possible. We have shown that the combination of pressure velocity $\omega_{500}$ and a measure of temperature inversion ECTEI are enough to capture almost all main features of cloud albedo, with the exception of the extratropical seasonal cycles, while surface temperature SST, standard deviation of hourly pressure velocity $\omega_{\text{std}}$, and $\omega_{500}$, capture all main features of longwave CRE. Even though we only fitted over the ocean here, in fact the fits do not perform much worse when considering the whole planet without adding more information to the cloud fitting functions $f$ (not shown). We also note that the predictors used in the presented models were favored because of their simplicity, but also because they can be potentially connected with parameters that might be inferred from simple energy balance models, for example, quantities such as the equator-to-pole temperature gradient, as outlined in the Supporting Information. This may allow incorporating cloudiness in energy balance models, and idea we are pursuing in ongoing research.

Besides providing a starting point for a theory of global cloudiness, our models are also useful in complementing other approaches to better quantify the response of cloudiness to a change in the climate system. Studies such as Ceppi and Nowack (2021); Myers et al. (2021), and in fact most CCFF studies, perform multilinear regressions to optimize tens of thousands of free of parameters (at least as many as each spatial grid point of the used data set). Hence, they may obtain predictive power that is higher than our models. However, our models are useful for providing a baseline to compare against that is based on concrete physical requirements. Furthermore, exactly because our models do not depend on location, we believe they are better suited to answer in which areas of the globe would circulation (or temperature) changes impact global cloudiness more, or less.

## Data Availability Statement

The data sets used were monthly mean CERES EBAF (Doelling et al., 2013; Kato et al., 2018; Loeb et al., 2018; Rutan et al., 2015) for surface & top of the atmosphere radiation fields, and cloud properties, monthly mean ERA5 (Hersbach et al., 2020) for temperature, pressure, humidity, and hourly mean ERA5 for pressure velocity. The code we used is available online (Datseris, 2022). It uses the Julia language (Bezanson et al., 2017), and the packages: GLM.jl, LsqFit.jl, ClimateBase.jl, and DrWatson (Datseris et al., 2020). Figures were produced with the matplotlib library (Hunter, 2007). The code can also be used to fit any arbitrary spatiotemporal field with any combination of functional forms and predictor fields.

**Acknowledgments**
We thank Hauke Schmidt and Aiko Voigt for helpful discussions. G.D. acknowledges support from the CONSTRAIN project (EU Horizon 2020, 820829). Open access funding enabled and organized by Projekt DEAL.

## References

Betts, A. K., & Ridgway, W. (1989). Climatic equilibrium of the atmospheric convective boundary layer over a tropical ocean. *Journal of the Atmospheric Sciences*, *46*(17), 2621–2641. https://doi.org/10.1175/1520-0469(1989)046<2621:ceotac>2.0.co;2

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Bony, S., Dufresne, J.-L., Treut, H. L., Morcrette, J.-J., & Senior, C. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, *22*(2–3), 71–86. https://doi.org/10.1007/s00382-003-0369-6

Bony, S., Lau, K.-M., & Sud, Y. C. (1997). Sea surface temperature and large-scale circulation influences on tropical greenhouse effect and cloud radiative forcing. *Journal of Climate*, *10*(8), 2055–2077. https://doi.org/10.1175/1520-0442(1997)010<2055:SSTALS>2.0.CO;2

Bony, S., Semie, A., Kramer, R. J., Soden, B., Tompkins, A. M., & Emanuel, K. A. (2020). Observed modulation of the tropical radiation budget by deep convective organization and lower-tropospheric stability. *AGU Advances*, *1*(3). https://doi.org/10.1029/2019av000155

Bretherton, C. S., & Wyant, M. C. (1997). Moisture transport, lower-tropospheric stability, and decoupling of cloud-topped boundary layers. *Journal of the Atmospheric Sciences*, *54*(1), 148–167. https://doi.org/10.1175/1520-0469(1997)054<0148:mtltsa>2.0.co;2

Budyko, M. I. (1969). The effect of solar radiation variations on the climate of the Earth. *Tellus*, *21*(5), 611–619. https://doi.org/10.3402/tellusa.v21i5.10109

Carlson, T. N. (1980). Airflow through midlatitude cyclones and the comma cloud pattern. *Monthly Weather Review*, *108*(10), 1498–1509. https://doi.org/10.1175/1520-0493(1980)108<1498:ATMCAT>2.0.CO;2

Ceppi, P., & Nowack, P. (2021). Observational evidence that cloud feedback amplifies global warming. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(30), e2026290118. https://doi.org/10.1073/pnas.2026290118

Cotton, W. R., Bryan, G., & Van Den Heever, S. C. (2014). *Storm and cloud dynamics* (2nd ed.). Academic Press.

Dadashazar, H., Crosbie, E., Majdi, M. S., Panahi, M., Moghaddam, M. A., Behrangi, A., et al. (2020). Stratocumulus cloud clearings: Statistics from satellites, reanalysis models, and airborne measurements. *Atmospheric Chemistry and Physics*, *20*(8), 4637–4665. https://doi.org/10.5194/acp-20-4637-2020

Datseris, G. (2022). Code for our paper "minimal recipes for global cloudiness". *Zenodo*. https://doi.org/10.5281/zenodo.7063184

Datseris, G., Isensee, J., Pech, S., & Gál, T. (2020). Drwatson: The perfect sidekick for your scientific inquiries. *Journal of Open Source Software*, *5*(54), 2673. https://doi.org/10.21105/joss.02673

Datseris, G., & Stevens, B. (2021). Earth's albedo and its symmetry. *AGU Advances*, *2*(3). https://doi.org/10.1029/2021av000440

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., et al. (2013). Geostationary enhanced temporal interpolation for CERES flux products. *Journal of Atmospheric and Oceanic Technology*, *30*(6), 1072–1090. https://doi.org/10.1175/jtech-d-12-00136.1

Ghil, M. (1981). Energy-balance models: An introduction. In *Climatic variations and variability: Facts and theories* (pp. 461–480). Springer Netherlands. https://doi.org/10.1007/978-94-009-8514-8_27

Grise, K. M., & Kelleher, M. K. (2021). Midlatitude cloud radiative effect sensitivity to cloud controlling factors in observations and models: Relationship with southern hemisphere jet shifts and climate sensitivity. *Journal of Climate*, *34*(14), 5869–5886. https://doi.org/10.1175/JCLI-D-20-0986.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Houze, R. A., Jr. (2014). *Cloud dynamics* (2nd ed.). Academic Press.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Isensee, J., Datseris, G., & Parlitz, U. (2019). Predicting spatio-temporal time series using dimension reduced local states. *Journal of Nonlinear Science*, *30*(3), 713–735. https://doi.org/10.1007/s00332-019-09588-7

Kato, S., Rose, F. G., Rutan, D. A., Thorsen, T. J., Loeb, N. G., Doelling, D. R., et al. (2018). Surface irradiances of edition 4.0 clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) data product. *Journal of Climate*, *31*(11), 4501–4527. https://doi.org/10.1175/jcli-d-17-0523.1

Kawai, H., Koshiro, T., & Webb, M. J. (2017). Interpretation of factors controlling low cloud cover and low cloud feedback using a unified predictive index. *Journal of Climate*, *30*(22), 9119–9131. https://doi.org/10.1175/jcli-d-16-0825.1

Kelleher, M. K., & Grise, K. M. (2019). Examining Southern Ocean cloud controlling factors on daily time scales and their connections to midlatitude weather systems. *Journal of Climate*, *32*(16), 5145–5160. https://doi.org/10.1175/JCLI-D-18-0840.1

Klein, S. A., Hall, A., Norris, J. R., & Pincus, R. (2017). Low-cloud feedbacks from cloud-controlling factors: A review. *Surveys in Geophysics*, *38*(6), 1307–1329. https://doi.org/10.1007/s10712-017-9433-3

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, *2*(2), 164–168. https://doi.org/10.1090/qam/10666

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., et al. (2018). Clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918. https://doi.org/10.1175/jcli-d-17-0208.1

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, *11*(2), 431–441. https://doi.org/10.1137/0111030

Myers, T. A., Scott, R. C., Zelinka, M. D., Klein, S. A., Norris, J. R., & Caldwell, P. M. (2021). Observational constraints on low cloud feedback reduce uncertainty of climate sensitivity. *Nature Climate Change*, *11*(6), 501–507. https://doi.org/10.1038/s41558-021-01039-0

Norris, J. R., & Iacobellis, S. F. (2005). North Pacific cloud feedbacks inferred from synoptic-scale dynamic and thermodynamic relationships. *Journal of Climate*, *18*(22), 4862–4878. https://doi.org/10.1175/jcli3558.1

Norris, J. R., & Weaver, C. P. (2001). Improved techniques for evaluating GCM cloudiness applied to the NCAR CCM3. *Journal of Climate*, *14*(12), 2540–2550. https://doi.org/10.1175/1520-0442(2001)014<2540:ITFEGC>2.0.CO;2

North, G., & Kim, K.-Y. (2017). *Energy balance climate models*. Wiley-VCH.

Peters, M. E., & Bretherton, C. S. (2005). A simplified model of the Walker circulation with an interactive ocean mixed layer and cloud-radiative feedbacks. *Journal of Climate*, *18*(20), 4216–4234. https://doi.org/10.1175/JCLI3534.1

Pierrehumbert, R. T., & Swanson, K. L. (1995). Baroclinic instability. *Annual Review of Fluid Mechanics*, *27*(1), 419–467. https://doi.org/10.1146/annurev.fluid.27.1.419

Rutan, D. A., Kato, S., Doelling, D. R., Rose, F. G., Nguyen, L. T., Caldwell, T. E., & Loeb, N. G. (2015). CERES synoptic product: Methodology and validation of surface radiant flux. *Journal of Atmospheric and Oceanic Technology*, *32*(6), 1121–1143. https://doi.org/10.1175/jtech-d-14-00165.1

Sellers, W. D. (1969). A global climatic model based on the energy balance of the Earth-atmosphere system. *Journal of Applied Meteorology*, *8*(3), 392–400. https://doi.org/10.1175/1520-0450(1969)008<0392:AGCMBO>2.0.CO;2

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4). https://doi.org/10.1029/2019rg000678

Siebesma, A. P., Bony, S., Jakob, C., & Stevens, B. (Eds.) (2020)., *Clouds and climate*. Cambridge University Press. https://doi.org/10.1017/9781107447738

Stephens, G. L., O'Brien, D., Webster, P. J., Pilewski, P., Kato, S., & Li, J.-l. (2015). The albedo of Earth. *Reviews of Geophysics*, *53*(1), 141–163. https://doi.org/10.1002/2014RG000449

Stevens, B., & Brenguier, J.-L. (2009). Cloud-controlling factors: Low clouds. In J. Heintzenberg & R. J. Charlson (Eds.), *Clouds in the perturbed climate system*. The MIT Press. chap. 8. https://doi.org/10.7551/mitpress/9780262012874.001.0001

Stevens, B., & Schwartz, S. E. (2012). Observing and modeling Earth's energy flows. *Surveys in Geophysics*, *33*(3–4), 779–816. https://doi.org/10.1007/s10712-012-9184-0

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1). https://doi.org/10.1029/2019gl085782

Zhang, Y., Stevens, B., Medeiros, B., & Ghil, M. (2009). Low-cloud fraction, lower-tropospheric stability, and large-scale divergence. *Journal of Climate*, *22*(18), 4827–4844. https://doi.org/10.1175/2009jcli2891.1

## References From the Supporting Information

Brueck, M., Nuijens, L., & Stevens, B. (2015). On the seasonal and synoptic time-scale variability of the North Atlantic trade wind region and its low-level clouds. *Journal of the Atmospheric Sciences*, *72*(4), 1428–1446. https://doi.org/10.1175/jas-d-14-0054.1

Charney, J. G. (1947). The dynamics of long waves in a baroclinic westerly current. *Journal of the Atmospheric Sciences*, *4*(5), 136–162. https://doi.org/10.1175/1520-0469(1947)004<0136:TDOLWI>2.0.CO;2

Eady, E. T. (1949). Long waves and cyclone waves. *Tellus*, *1*(3), 33–52. https://doi.org/10.1111/j.2153-3490.1949.tb01265.x

Held, I. M., & Hou, A. Y. (1980). Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *Journal of the Atmospheric Sciences*, *37*(3), 515–533. https://doi.org/10.1175/1520-0469(1980)037<0515:nascia>2.0.co;2

Myers, T. A., & Norris, J. R. (2016). Reducing the uncertainty in subtropical cloud feedback. *Geophysical Research Letters*, *43*(5), 2144–2148. https://doi.org/10.1002/2015gl067416

Qu, X., Hall, A., Klein, S. A., & DeAngelis, A. M. (2015). Positive tropical marine low-cloud cover feedback inferred from cloud-controlling factors. *Geophysical Research Letters*, *42*(18), 7767–7775. https://doi.org/10.1002/2015gl065627

Sobel, A. H., Nilsson, J., & Polvani, L. M. (2001). The weak temperature gradient approximation and balanced tropical moisture waves. *Journal of the Atmospheric Sciences*, *58*(23), 3650–3665. https://doi.org/10.1175/1520-0469(2001)058<3650:TWTGAA>2.0.CO;2

Wood, R., & Bretherton, C. S. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *Journal of Climate*, *19*(24), 6425–6432. https://doi.org/10.1175/JCLI3988.1