Geophysical Research Letters

RESEARCH LETTER
10.1029/2024GL110124

Multifractal Analysis for Evaluating the Representation of Clouds in Global Kilometer-Scale Models

Lilli J. Freischem[1], Philipp Weiss[1], Hannah M. Christensen[1], and Philip Stier[1]

[1]Department of Physics, University of Oxford, Oxford, UK

Key Points:

- Quantifiable, structural evaluation metrics such as multifractal analysis should be used to evaluate and improve km-scale models
- Multifractal analysis finds that deep convection in the ICON model is not organized enough leading to smaller fractal parameters
- The model's bias toward smaller fractal parameters can be attributed to clouds simulated over the ocean

Supporting Information:

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

Correspondence to:

L. J. Freischem,
lilli.freischem@physics.ox.ac.uk

Citation:

Freischem, L. J., Weiss, P., Christensen, H. M., & Stier, P. (2024). Multifractal analysis for evaluating the representation of clouds in global kilometer-scale models. *Geophysical Research Letters*, *51*, e2024GL110124. https://doi.org/10.1029/2024GL110124

Received 5 MAY 2024
Accepted 27 SEP 2024

Author Contributions:

Conceptualization: Lilli J. Freischem, Philipp Weiss, Hannah M. Christensen, Philip Stier
Formal analysis: Lilli J. Freischem
Methodology: Lilli J. Freischem, Philipp Weiss, Hannah M. Christensen, Philip Stier
Software: Lilli J. Freischem
Supervision: Hannah M. Christensen, Philip Stier
Validation: Lilli J. Freischem
Visualization: Lilli J. Freischem
Writing – original draft: Lilli J. Freischem
Writing – review & editing: Lilli J. Freischem, Philipp Weiss, Hannah M. Christensen, Philip Stier

Abstract Clouds are one of the largest sources of uncertainty in climate predictions. Global km-scale models need to simulate clouds and precipitation accurately to predict future climates. To isolate issues in their representation of clouds, models need to be thoroughly evaluated with observations. Here, we introduce multifractal analysis as a method for evaluating km-scale simulations. We apply it to outgoing longwave radiation fields to investigate structural differences between observed and simulated anvil clouds. We compute fractal parameters which compactly characterize the scaling behavior of clouds and can be compared across simulations and observations. We use this method to evaluate the nextGEMS ICON simulations via comparison with observations from the geostationary satellite GOES-16. We find that multifractal scaling exponents in the ICON model are significantly lower than in observations. We conclude that too much variability is contained in the small scales (<100 km) leading to less organized convection and smaller, isolated anvils.

Plain Language Summary In this paper, we present a new approach to evaluating state-of-the-art high-resolution climate models. We use a type of analysis that captures how a field like outgoing radiation varies between two points in space; it is called multifractal analysis. We apply multifractal analysis to snapshots of climate model simulations and satellite observations, and compare the results to evaluate the model. In contrast to traditional evaluation approaches, our method focuses on the evaluation of the spatio-temporal structure of cloud fields, exploiting previously untapped information content. Hence, it can take into account the fine details in time and space that high-resolution climate models provide. We use our method to evaluate the ICON atmospheric model. We find that the simulations does not contain enough large clusters of clouds, as found in big thunderstorms, but instead clouds are randomly distributed in space: the simulated clouds are not organized enough.

1. Introduction

Global high-resolution km-scale models simulate the global atmosphere at an unprecedented level of detail. Advances in high-performance computing have enabled running such global km-scale simulations: first for short time periods—30 days—as demonstrated by Miura et al. (2007) who succeeded in simulating the Madden-Julian oscillation with the Nonhydrostatic Icosahedral Atmospheric Model (NICAM); then for longer time periods of up to 30 years at 5 km grid spacing (Koldunov et al., 2023). By simulating the climate at kilometer-scales, models can better resolve important processes—particularly deep convection—that conventional climate models must parameterize (Hohenegger et al., 2020; Stevens et al., 2019). This significantly improves the accuracy of the simulated cloud and precipitation fields which is a key step toward reducing the uncertainty in projections of global warming (Randall et al., 2003).

Nonetheless, significant uncertainties in km-scale simulations remain due to parameterizations of remaining unresolved subgrid-scale processes (Holloway et al., 2013; Stevens et al., 2019). The representation of clouds in global climate models has been recognized for decades as one of the largest sources of uncertainty in predictions of climate variability and regional impacts of climate change (e.g., Randall et al., 2003; Williams & Webb, 2008). Km-scale models need to simulate clouds and precipitation accurately and reliably to be used to predict future climates. To pin down the remaining issues in their representation of clouds, and thereby facilitate their improvement, simulations need to be thoroughly evaluated via comparisons with observations (Stevens et al., 2019; Takasuka et al., 2024; Zadra et al., 2018).

Traditionally, climate model performance is evaluated against spatio-temporally averaged observations (Flato et al., 2014; Gleckler et al., 2008). However, the spatio-temporal structure encodes crucial information about underlying physical processes of simulated convection and the resulting cloud structures, which is lost in

aggregated evaluation (Mooers et al., 2023). Employing a spatio-temporal evaluation strategy using satellite observations, combined with in situ local retrievals, provides direct constraints on physical processes.

Palmer (2016) proposed to evaluate simulations by visually comparing them to satellite observations. Analogous to the Turing test for Artificial Intelligence, climate models pass the climatic Turing test if they cannot be distinguished from satellite images. When applied to snapshots of top-of-atmosphere outgoing longwave radiation (OLR), such a visual comparison facilitates evaluating a model's ability to accurately simulate deep convection as OLR is sensitive to deep convective cores and their associated anvil clouds. An early study with this aim was conducted by Christensen and Driver (2021) who evaluated km-scale models based on the monofractal dimension of clouds, computed from OLR. The authors showed that the fractal nature of clouds is reproduced by simulations and linked observed differences in fractal dimensions to different boundary layer schemes.

However, clouds are known to be multifractal; increasingly intense regions are distributed over increasingly sparse fractal sets (Schertzer & Lovejoy, 1989). This is not captured by monofractal analysis, which assumes that a single fractal dimension can fully characterize the cloud field: it requires binarizing the cloud field and thereby loses structural information. Multifractal analysis provides a more in-depth characterization of the spatial structure of cloud variability (Tessier et al., 1993).

Here, we introduce multifractal analysis as a method for evaluating km-scale simulations. We apply it to OLR fields to investigate structural differences between observed and simulated deep convective clouds in the tropics. We compute fractal parameters which compactly characterize OLR variability and can be compared across simulations and observations. We use this method to evaluate the cycle 3 nextGEMS ICON simulation via comparison with GOES-16 images.

This paper is structured as follows: Section 2 describes the model and satellite data sets used in this study; Section 3 explains the empirical computation of multifractal scaling behavior from OLR images; Section 4 compares the scaling behavior of OLR fields in ICON and GOES-16, first, looking at average difference in the whole tropical band visible from GOES-16 in September 2020 and second, comparing fractals in smaller areas and their diurnal cycle; finally, Section 5 summarizes our findings and discusses their significance as well as limitations.

2. Data

We evaluate high-resolution atmospheric model data from the ICON model using geostationary satellite observations from the GOES-16 satellite. We focus our analyses on outgoing longwave radiation as it is observable by (geostationary) satellites and well-suited to constrain deep convection (Ohring & Gruber, 1983).

2.1. nextGEMS ICON Model

We analyse the global ICON simulations produced for the 3rd nextGEMS cycle (Koldunov et al., 2023). The model was run in ICON-Sapphire configuration (Hohenegger et al., 2023). It is a coupled model with uniform 5 km grid spacing in atmosphere and ocean, using time steps of 40 seconds for cloud microphysics and turbulence and 12 minutes for radiation. It was initialized at 00:00 UTC on 20 January 2020 and was run for 5 years. OLR is provided at 30 min intervals and saved on the Hierarchical Equal Area isoLatitude Pixelisation (HEALPix) grid (Górski et al., 2005). We use the finest resolution available, HEALPix zoom level 10, where the 4,096 pixels on each iso-latitude ring located between 20°N and 20°S have 0.088° spacing.

2.2. GOES-16 Outgoing Longwave Radiation

We use observations from the Advanced Baseline Imager (ABI) of the first satellite of the Geostationary Operational Environmental Satellites (GOES) series, GOES-16 (GOES-R Calibration Working Group & GOES-R Series Program, 2017). The satellite was launched in 2016, became operational in December 2017 and is positioned at 75.2°W. Since 2019, it is operating in mode 6 and provides a full disk image at 2 km spatial resolution every 10 minutes (Schmit & Gunshor, 2020).

We estimate OLR from ABI narrowband infrared measurements using the OLR algorithm (see Text S1 in Supporting Information S1) developed by the Earth Radiation Budget Team of the GOES-R Algorithm Working Group (Lee et al., 2010).

2.3. Region and Time Period

We analyse the tropical band visible from the GOES-16 satellite which ranges from 20°W to 130°W (Figure S1 in Supporting Information S1). Our method requires images to have regular longitudinal pixel distances (see Section 3). We regrid GOES-16 observations to a regular 0.05° grid using nearest neighbor interpolation implemented in the SciPy package, as the same method was used to interpolate ICON data onto the HealPix grid (Virtanen et al., 2020). We tested the sensitivity of our analysis to the slight difference in resolution—GOES observations are at 0.05° and ICON simulations are at 0.088° resolution—by regridding GOES images to the HealPix grid which did not affect results: fractal parameters computed on the two grids were equal up to the second decimal point.

In our analysis, we assume constant distance between pixels. We use data between 20°N and 20°S to limit the distortion in longitudinal distance to $\cos(20°) = 0.94$ (Christensen & Driver, 2021; Pierrehumbert, 1996).

We use 1 month of data, September 2020, at hourly intervals for our analyses. This limits the computational cost of our analyses while providing enough data to compute robust multifractal statistics. We analyse a month within the first ICON simulation year to limit the impact of long-term model drifts on our findings.

3. Methods

Clouds exhibit fractal structures over a wide range of scales. However, they are not simply self-similar, or monofractal; if they followed isotropic scaling, large clouds would be completely indistinguishable from small clouds, which is not the case. For example, assume we have a roughly spherical cloud with a diameter of 1 km. Following isotropic scaling, this would imply that there are clouds with a diameter of 100 km that are 100 km high which is not realistic. Instead, atmospheric scaling is non-isotropic, due to effects such as the stratification of the atmosphere because of gravity and the differential rotation because of the Coriolis force (Schertzer & Lovejoy, 1989). The resulting scaling behavior cannot be described by a single fractal dimension: clouds are multifractal.

To analyse the multifractal behavior of deep convective clouds, we fit scaling exponents to structure functions which are directly computed from OLR fields (Davis et al., 1994; Vainshtein et al., 1994). We then parameterize the scaling exponents, using the two-parameter fit introduced by Pierrehumbert (1996), and subsequently compare parameters across simulations and observations as a new quantitative constraint on simulated cloud structures.

3.1. Structure Functions

Structure functions describe the average difference of a signal evaluated at two points in space (or time), separated by some distance r. More precisely, the qth-order structure function denotes the average of the qth power of absolute differences in some one-dimensional signal $\theta(x)$ as a function of the distance r between two points:

$$S_q(r) = \langle |\theta(x + r) - \theta(x)|^q \rangle, \tag{1}$$

where r is a finite distance in x, and $\langle \ldots \rangle$ indicates a spatial average over all x, thereby assuming statistical homogeneity.

The first order structure function $S_1(r)$ is an unweighted average of fluctuations in $\theta(x)$ at a certain radius. The ensemble average in higher order structure functions is dominated by atypically large fluctuations in the signal, which correspond to the largest values of $|\theta(x + r) - \theta(x)|$.

3.2. Multifractal Scaling

The scaling range of a signal is defined as the range where its structure functions follow a power-law. Within the scaling range, the magnitude of the structure function is proportional to the separation distance r, raised to some exponent ζ_q:

$$S_q(r) \propto r^{\zeta_q}, \tag{2}$$

such that $\log S_q(r)$ is approximately linear in $\log r$ (Davis et al., 1994). For a fractal signal, this range should exist and be the same for all orders q. The scaling range can be identified from plots of the structure functions as explained in Section 3.3.

Monofractal processes have a single fractal dimension. We can study a set of fractal objects within a real-valued image by choosing a threshold above which all points in an image form the set of monofractal objects. The image is monofractal if the resulting set of objects has the same fractal dimension independent of the chosen threshold. When we study the scaling of a fractal field via its structure functions, instead of increasing a threshold, we increase the order q of the structure function whereby we assign more weight to larger fluctuations in the signal. If the field is monofractal, large fluctuations follow the same scaling law as small fluctuations. Raising absolute differences in the signal to the power q is equivalent to raising the structure function itself to the power q, or multiplying its logarithm by q. The ζ exponents are linearly related and follow $\zeta_q = q\zeta_1$.

In the atmosphere, however, energy and matter are concentrated into smaller and smaller regions by multiplicative cascade processes. For example, kinetic energy is transferred from larger-scale eddies to progressively smaller eddies via the turbulent energy cascade, until it is dissipated into thermal energy. Hence, atmospheric fields are multifractal where increasingly intense regions are distributed over increasingly sparse fractal sets (Cho et al., 2000; Schertzer & Lovejoy, 1989). If we study a set of cloud objects by choosing an OLR threshold below which all points are considered "cloudy," we would observe decreasing fractal dimensions as the OLR threshold is increased (Figure 3, Christensen & Driver, 2021). The scaling exponents do not follow a linear relationship but lie on a concave curve.

The second order structure function $S_2(r)$ provides the same information as the power spectrum. From the Wiener-Khinchin theorem, it follows that $\beta = \zeta_2 + 1$, where β is the negative exponent of the horizontal wavenumber k power spectrum (Lovejoy, 2023). Since the ζ exponents of a monofractal field are linearly related to each other, the power spectrum can fully describe the scaling behavior of monofractal fields. In the case of multifractals, however, ζ_2 does not capture the full scaling behavior of the field. Hence, the ζ_q curve provides a more in-depth characterization of the spatial structure of cloud variability than the power spectrum alone (Tessier et al., 1993).

3.3. Computational Procedure

Structure functions are computed separately for every image in our data set, closely following the procedure outlined by Pierrehumbert (1996). Each OLR snapshot is treated as an ensemble of one-dimensional signals along a constant latitude. We compute the OLR structure function $\hat{S}_q$ for each image by computing the structure function of each one-dimensional latitudinal signal in the image according to Equation 1, where $\theta(x)$ is OLR at a fixed latitude as a function of longitude x, and averaging over all latitudes (Pierrehumbert, 1996).

We compute structure functions of order 1–10, and the corresponding ζ_q exponents. Structure functions for higher values of q suffer from sampling issues as single points will eventually dominate the ensemble average. In general, values of q beyond where ζ_q becomes linear in q, which is reached by $q = 10$ (Figure 1), do not add information and are therefore omitted (Davis et al., 1994).

In our analyses of GOES-16 data, which is on a regular 0.05° latitude-longitude grid, we start the computation at distance $r = 1$ pixel ≈ 5.5 km and compute structure functions of order 1–10. This is repeated for pixel distances up to $r = 512$ pixels ≈ 2842 km, increasing r by a factor of two at each step. We compute structure functions of ICON images following the same methodology. We use pixel differences from $r = 1$ pixel ≈ 9.8 km to $r = 291$ pixel ≈ 2839 km to obtain data points at approximately the same distance in kilometers, as the ICON data has a longitudinal grid-spacing of 0.088°.

We visually identify the range where OLR exhibits multifractal scaling which corresponds to the linear range of the structure functions in log-log space. Multifractal scaling exponents are computed according to Equation 2 by applying linear regression to the log-log plots of the structure functions within that scaling range. For visualization purposes, we normalize structure functions as

$$\hat{S}_{q,\mathrm{norm}}(r) = \hat{S}_q(r)/\hat{S}_q(1 \text{ pixel}), \tag{3}$$

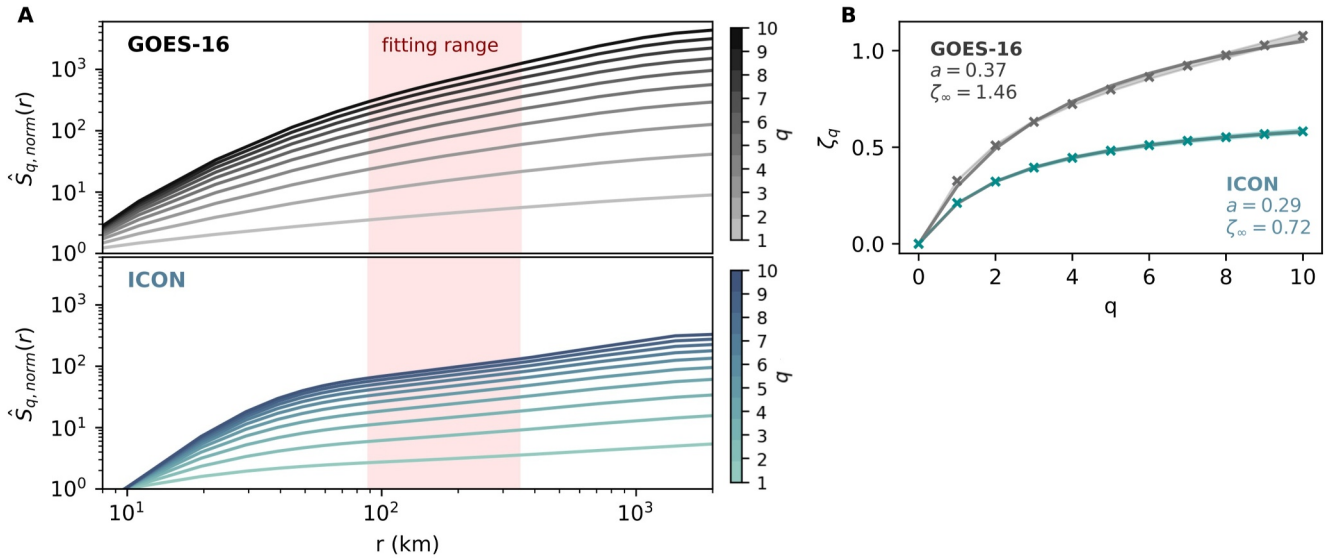

Figure 1. Comparison of average normalized structure functions and fractal parameters of outgoing longwave radiation fields in the GOES-16 satellite and the nextGEMS cycle 3 ICON model. Analysed was the region from 20°N to 20°S and 20°W to 130°W in September 2020 at hourly intervals. (a) Shows normalized structure functions $\hat{S}_{q,\mathrm{norm}}(r)$ of order $q = 1$ (the lowest line of each plot) to $q = 10$ (the highest line of each plot). The red shaded range corresponds to the fitting range used to compute the multifractal scaling exponents shown in (b) from the log-log slope of the structure functions. The scaling exponents were parameterized according to Equation 4.

using the value of $\hat{S}_q$ at the grid spacing of the data: 0.05° for GOES-16 and 0.088° for ICON. This operation shifts the structure functions downwards in the log-log plots as $\hat{S}_q(1\ \mathrm{pixel})$ is constant for each value of q. It therefore does not affect the ζ values obtained from the slope of the structure function in log-log space.

It is important to note that a straight line on the log-log plot of structure function estimates is only expected for an ensemble average (Lovejoy & Schertzer, 2006). For a single line of OLR in our images, the scaling will be broken. Even structure functions of individual images are only approximately linear. Clouds exhibit multifractal scaling for scales between 1 mm and thousands of kilometers (e.g., Lovejoy & Schertzer, 2006; Pierrehumbert, 1996). However, the linear range of $\hat{S}_q$ of individual images in our study depends on the resolution of the data, the effective resolution of the model, and the area of the region over which the structure functions are averaged.

We require a more compact representation of the scaling behavior than the ζ curves to evaluate models. For this purpose, we parameterize the calculated ζ_q exponents as

$$\zeta_q = \frac{aq}{1 + aq/\zeta_\infty}, \tag{4}$$

using the two-parameter fit introduced by Pierrehumbert (1996). The first parameter a describes the most probable behavior of the field. OLR variability over distance r typically scales like r^a. The second parameter ζ_∞ measures the degree of multifractality, or intermittency, of the field. An OLR field has higher ζ_∞ when deep convection is more aggregated, as will be illustrated in Section 4.2.

4. Results and Discussion

We first compare OLR structure functions and fractal parameters of the ICON model with those computed from GOES-16 data over the tropical band visible from the satellite. This comparison shows a bias in the ICON model toward lower fractal parameters; cloud fields in ICON do not have enough variability in the large scales.

We then study fractal parameters of sub-regions within the tropical band visible from GOES-16. This facilitates gaining a physical intuition of the interpretation of differences in fractal parameters. It also enables studying the diurnal cycle of convection via fractal parameters. This is of particular interest as general circulation models struggle to accurately simulate the diurnal cycle of convection and precipitation (Baranowski et al., 2019).

4.1. Comparison of Fractal Parameters of Cloud Fields

For a first evaluation of the ICON simulation, we compare average structure functions and fractal parameters between GOES-16 and ICON.

As shown in Figure 1a, averaged structure functions are approximately linear from 60 to 1,000 km. However, there is too much statistical variation in the large scales, and the structure functions of individual images are usually not linear beyond 400 km. Hence, we compute multifractal scaling exponents according to Equation 2 by applying linear regression to the log-log plots of the structure functions using a fitting range of $r = 90$ to 350 km. The ICON structure functions have a significant bend at $r = 40$ km; this scale corresponds to the effective model resolution (Fan et al., 2022; Klaver et al., 2020).

Finally, we compute fractal parameters a and ζ_∞ by fitting Equation 4 to the ζ curves (Figure 1b). The comparison between model and satellite shows that fractal parameters in ICON are too low—a and ζ_∞ are 0.29 and 0.72 in ICON, compared to 0.37 and 1.46 in GOES-16, respectively. The fractal parameters characterize the curve of scaling exponents ζ which correspond to the log-log slope of structure functions $S_q(r)$ (see Section 3). Hence, smaller fractal parameters indicate shallower structure functions in the ICON model compared to observations. The structure functions are shallower because $S_{q,ICON}$ is larger than $S_{q,GOES}$ for small r but approaches the same magnitude at large r: too much OLR variability in ICON is contained in the small scales. This result is consistent with the finding that km-scale models with no convective parameterization tend to have too high energies at small scales (Fan et al., 2022; Stephan et al., 2022).

4.2. Regional Studies

We study fractal parameters of smaller regions within the tropical band visible from GOES-16 (shown in Figure S1 in Supporting Information S1) to gain physical intuition of the interpretation of fractal parameters and to investigate potential explanations for the observed differences in scaling behavior.

We study the scaling of clouds over three sub-regions: the Amazon delta (15° S − 10° N, 50° W − 80° W)—the largest land area in the tropical band visible from GOES-16—and the Intertropical Convergence Zone (ITCZ) over the Pacific (5° N − 20° N, 90° W − 130° W) and the Atlantic (5° N − 20° N, 20° W − 60° W) oceans respectively. We mask out sea areas in the Amazon region (see Figure 2), and we mask out the land areas in the ITCZ regions. This enables us to study cloud scaling and evaluate model performance over land and over sea separately.

First, we consider the Amazon basin (Figure 2; see Figures S2 and S3 in Supporting Information S1 for equivalent figures for the Pacific and the Atlantic ITCZ and Text S2 in Supporting Information S1 for related discussion). For low fractal parameters, we observe "popcorn convection" over the Amazon basin. The images are relatively homogeneous: the popcorn convection is spread out evenly in the entire region. Note that for low fractal parameters, deep convection in the ICON model is more homogeneously distributed across the domain than in GOES-16 observations. In contrast, images with large fractal parameters contain larger, organized convective systems over the continent. The corresponding deep convective cloud fields are more diverse, there are some large organized systems, as well as areas of varying sizes that contain smaller anvils or are cloud-free. Some of these cloud fields look almost monofractal—the large anvils in the image look like the smaller anvils, scaled up. Overall, we can see that cloud fields with similar fractal parameters look similar as well, indicating that fractal parameters are well-suited for characterizing visual similarity between models and observations.

Cloudfields corresponding to the smallest fractal parameters appear around the same time each day in GOES-16 and ICON. This motivates an investigation of the diurnal cycle of fractal parameters. We compute the average structure function for each hour of the day to which we then fit scaling exponents and compute fractal parameters. The value for typical variability a is fairly consistent throughout the day (not shown) while intermittency ζ_∞ shows a significant diurnal cycle. This result is consistent with the study by Pierrehumbert (1996). Hence, we focus on ζ_∞ for the regional comparison.

The diurnal cycle is shown for each sub-region in Figure 3. ICON's fractal parameters match well with observations over the Amazon basin. On average, intermittency $\zeta_\infty = 1.3$ in ICON and $\zeta_\infty = 1.4$ in GOES. In addition, there is a clear diurnal cycle in ζ_∞ in the region which matches notably well in GOES-16 and ICON. As expected from Figure 2, ζ_∞ has a daily minimum around 13:00 local time in ICON and 14:00 in GOES. Noticeable is a fast

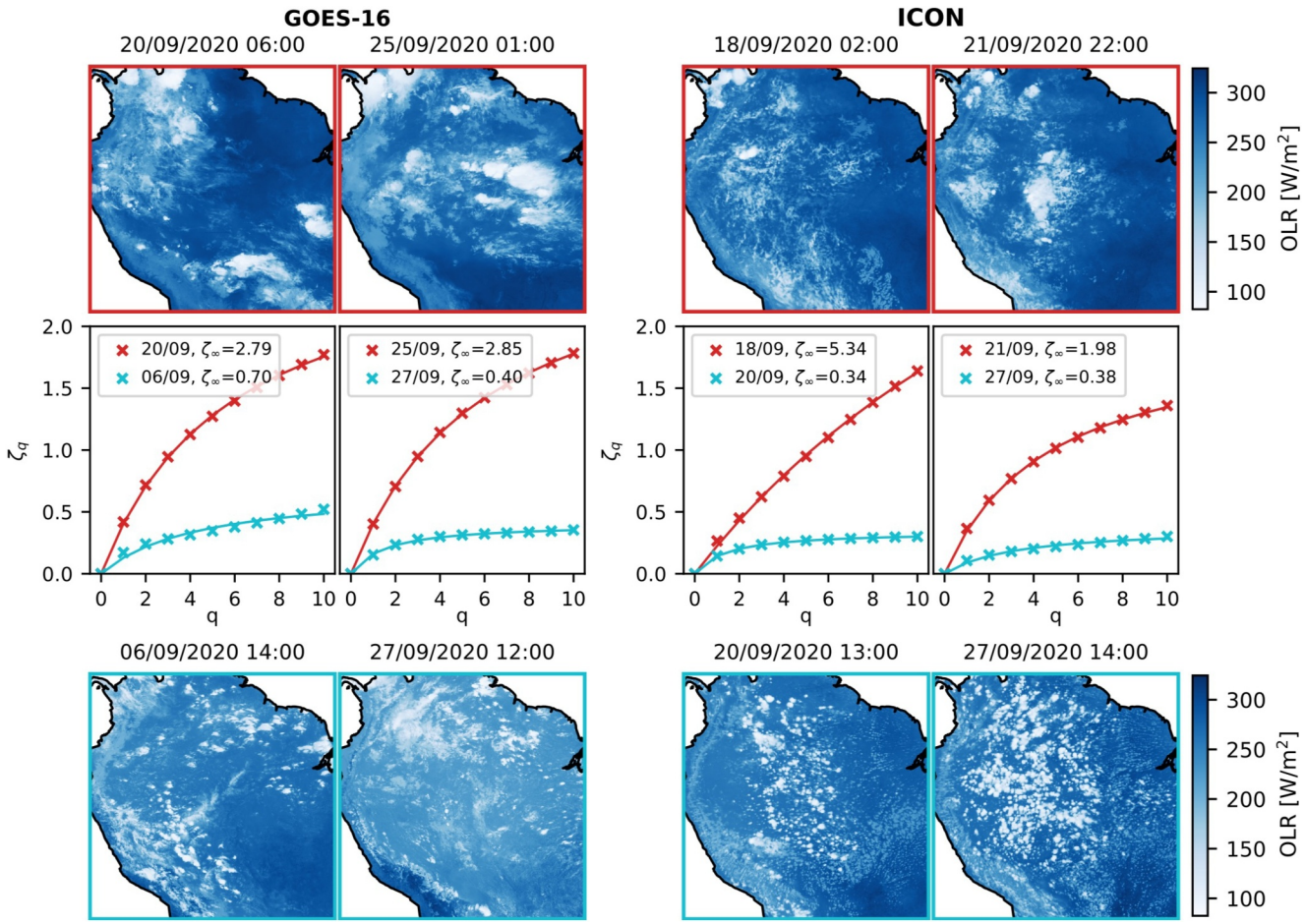

Figure 2. GOES-16 and ICON outgoing longwave radiation and corresponding fractal curves. Shown are snapshots and curves corresponding to either end of the observed range of fractal parameters, that is, two of the largest and smallest ζ_∞ values in red and blue, respectively. Times are given in local time (GMT-5).

decrease in ζ_∞ values in ICON between 10:00 and 12:00 local time whereas observations show a more gradual decline in ζ_∞ from 08:00 to 13:00 local time. The ζ_∞ values observed between 13:00 and 24:00 are lower in ICON than they are in GOES. Convection in ICON is not organized enough in the afternoon and evening.

In contrast, Figure 3 shows that fractal parameters of clouds forming in the ITCZ over the Pacific and the Atlantic ocean are significantly lower in ICON than they are in GOES-16 observations. On average, $\zeta_{\infty,\mathrm{ICON}} = 0.9$ whereas $\zeta_{\infty,\mathrm{GOES}} = 2.1$ in the two regions. These findings indicate that ICON's bias toward lower fractal parameters identified in Section 4.1 is primarily due to the behavior of clouds over the sea. The model seems unable

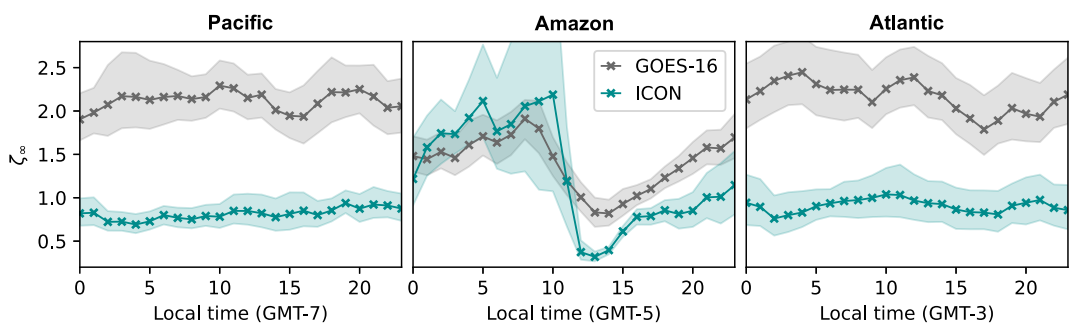

Figure 3. Diurnal cycles of ζ_∞ for the ITCZ over the Pacific, the Amazon delta, and the ITCZ over the Atlantic. The shaded regions denote the 95% confidence interval computed via bootstrapping.

to capture the organization of deep convective clouds in the ITCZ, and there are too many small anvils as can be seen in the examples in Figures S2 and S3 in Supporting Information S1.

5. Conclusions

We introduce multifractal analysis as an evaluation metric for km-scale climate models and present a case study evaluation of the representation of clouds in the nextGEMS cycle 3 ICON-Sapphire model (Koldunov et al., 2023). We compute OLR from GOES-16 ABI narrowband infrared measurements to facilitate the comparison between model and satellite. We then compute structure functions to which we fit scaling exponents and compute fractal parameters. A field is multifractal when the magnitude of its qth order structure function is proportional to the separation distance r, raised to some exponent, $S_q(r) \propto r^{\zeta_q}$, where the scaling exponents ζ_q are not linearly related. Clouds in observations are known to exhibit fractal structures over a wide range of scales. Our results show that clouds simulated by a km-scale model are multifractal for scales between 60 and 1,000 km.

In contrast to traditional approaches to climate model evaluation, multifractal analysis does not require spatially averaging simulations and instead exploits the physical information content encoded in cloud spatial structures. The significance of this property can be demonstrated by comparing the results of our regional fractal studies with average differences in OLR in the respective regions (see Figure S4 in Supporting Information S1). The average difference in OLR over the Amazon region is 14.5 W/m^s, whereas it is 0.8 W/m^s for the Pacific ITCZ and 3.9 W/m^s for the Atlantic ITCZ. In contrast, we found that fractal parameters and their diurnal cycle matched well over the Amazon region but were significantly different between GOES-16 and the ICON model over the Pacific and the Atlantic ITCZ (Figure 3). Evaluating model performance based on either average OLR difference or fractal parameters alone would give opposite results in these regions. Neither method can give the full answer and to evaluate the representation of clouds accurately, a combination of methods is needed. While it remains necessary that km-scale models perform well when evaluated with traditional aggregate metrics, these no longer provide sufficient constraints for evaluating simulations.

We showed that the fractal parameter ζ_∞ captures convective organization; low ζ_∞ values correspond to unorganized convection and high ζ_∞ values are obtained when anvils are organized into large systems. Fractal parameters can therefore be used as an index for convective organization and complement existing, object-based organization indices (Mandorli & Stubenrauch, 2023). An advantage of multifractal analysis is that it does not require binarising the OLR field. It therefore does not rely on choosing a threshold value to define cloudy and non-cloudy regions and can consider the full information contained in the OLR field.

The efficiency of our method makes it a useful tool for model intercomparison studies, which will be a topic of future work. The impact of various modeling choices, such as the implemented turbulence and microphysical parameterizations, and whether convection is parameterized or not, on the fractal parameters will be of particular interest. Initiatives such as the International Satellite Cloud Climatology Project—Next Generation (ISCCP-ng) will further facilitate access to global, cross-calibrated geostationary satellite data sets. It will therefore be perfectly suited to provide much needed observational constraints for global km-scale models.

A key limitation of our study is that by analysing outgoing longwave radiation, we focus on deep convection and anvil clouds. Shallow convection and thin cirrus clouds do not strongly influence OLR, and are therefore not investigated. However, they contribute key uncertainties in climate projections (Myers et al., 2021). Hence, a comprehensive evaluation of the representation of clouds in climate models should look beyond anvils. This could be addressed by computing narrowband radiances using satellite simulators which can directly be compared to observations, and will be the subject of future work (Bodas-Salcedo et al., 2011).

Overall, this study shows that analysing the previously unexploited spatio-temporal structure of clouds provides a promising new way to constrain global km-scale models. It can provide key insights into model performance and shed light on issues in the representation of clouds.

Data Availability Statement

NextGEMS Cycle 3 ICON data is archived by the German Climate Computing Center (DKRZ) and can be accessed via DKRZ's supercomputer Levante after registration at https://luv.dkrz.de/register/. In addition, derived data can be found at the World Data Center for Climate at https://doi.org/10.26050/WDCC/nextGEMS_

cyc3 (Koldunov et al., 2023). GOES-16 OLR data was derived from Level 1b radiance measurements which were supplied by the National Oceanic and Atmospheric Administration (NOAA) and can be downloaded at https://console.cloud.google.com/marketplace/product/noaa-public/goes (GOES-R Calibration Working Group & GOES-R Series Program, 2017). The code used for this research is openly available in a GitHub repository at https://github.com/lillif/multifractals (Freischem, 2024).

Acknowledgments
This work used supercomputing resources of the German Climate Computing Center (Deutsches Klimarechenzentrum, DKRZ) granted by its Scientific Steering Committee (WLA) under project ID 1153. We thank Dr. Hai-Tien Lee and Dr. Istvan Laszlo (NOAA) who provided an implementation of the GOES OLR retrieval algorithm, the required regression coefficients, and some GOES OLR data. Lilli J. Freischem acknowledges funding from the NERC Doctoral Training Partnership in Environmental Research Grant NE/S007474/1. Philip Stier and Philipp Weiss acknowledge funding from the Horizon 2020 projects nextGEMS under Grant 101003470 and FORCeS under Grant 821205. Hannah M. Christensen was funded by Natural Environment Research Council Grant NE/P018238/1 and through a Leverhulme Trust Research Leadership Award.

References

Baranowski, D. B., Waliser, D. E., Jiang, X., Ridout, J. A., & Flatau, M. K. (2019). Contemporary GCM fidelity in representing the diurnal cycle of precipitation over the maritime continent. *Journal of Geophysical Research: Atmospheres*, *124*(2), 747–769. https://doi.org/10.1029/2018JD029474

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S. A., et al. (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. https://doi.org/10.1175/2011BAMS2856.1

Cho, J. Y. N., Newell, R. E., & Sachse, G. W. (2000). Anomalous scaling of mesoscale tropospheric humidity fluctuations. *Geophysical Research Letters*, *27*(3), 377–380. https://doi.org/10.1029/1999GL010846

Christensen, H. M., & Driver, O. G. (2021). The fractal nature of clouds in global storm-resolving models. *Geophysical Research Letters*, *48*(23), e2021GL095746. https://doi.org/10.1029/2021GL095746

Davis, A., Marshak, A., Wiscombe, W., & Cahalan, R. (1994). Multifractal characterizations of nonstationarity and intermittency in geophysical fields: Observed, retrieved, or simulated. *Journal of Geophysical Research*, *99*, 8055–8072. https://doi.org/10.1029/94JD00219

Fan, D., Greybush, S. J., Chen, X., Lu, Y., Zhang, F., & Young, G. S. (2022). Exploring the role of deep moist convection in the wavenumber spectra of atmospheric kinetic energy and brightness temperature. *Journal of the Atmospheric Sciences*, *79*(10), 2721–2737. https://doi.org/10.1175/JAS-D-21-0285.1

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., et al. (2014). In I. Held, A. Pitman, S. Planton, & Z.-C. Zhao (Eds.), *Evaluation of climate models* (pp. 741–866). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.020

Freischem, L. J. (2024). Multifractal analysis of clouds [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.13832882

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, *113*(D6), D06104. https://doi.org/10.1029/2007JD008972

GOES-R Calibration Working Group, & GOES-R Series Program. (2017). NOAA GOES-R series advanced baseline imager (ABI) level 1b radiances [Dataset]. *NOAA National Centers for Environmental Information*. https://doi.org/10.7289/V5BV7DSR

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. (2005). HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, *622*(2), 759–771. https://doi.org/10.1086/427976

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., et al. (2023). ICON-Sapphire: Simulating the components of the Earth system and their interactions at kilometer and subkilometer scales. *Geoscientific Model Development*, *16*(2), 779–811. https://doi.org/10.5194/gmd-16-779-2023

Hohenegger, C., Kornblueh, L., Klocke, D., Becker, T., Cioni, G., Engels, J. F., et al. (2020). Climate statistics in global simulations of the atmosphere, from 80 to 2.5 km grid spacing. *Journal of the Meteorological Society of Japan Series II*, *98*(1), 73–91. https://doi.org/10.2151/jmsj.2020-005

Holloway, C. E., Woolnough, S. J., & Lister, G. M. S. (2013). The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study. Part I: Characterization of large-scale organization and propagation. *Journal of the Atmospheric Sciences*, *70*(5), 1342–1369. https://doi.org/10.1175/JAS-D-12-0227.1

Klaver, R., Haarsma, R., Vidale, P. L., & Hazeleger, W. (2020). Effective resolution in high resolution global atmospheric models for climate studies. *Atmospheric Science Letters*, *21*(4), e952. https://doi.org/10.1002/asl.952

Koldunov, N., Kölling, T., Pedruzo-Bagazgoitia, X., Rackow, T., Redler, R., Sidorenko, D., et al. (2023). nextGEMS: Output of the model development cycle 3 simulations for ICON and IFS [Dataset]. *World Data Center for Climate (WDCC) at DKRZ*. https://doi.org/10.26050/WDCC/nextGEMS_cyc3

Lee, H.-T., Laszlo, I., & Gruber, A. (2010). *ABI Earth radiation budget upward longwave radiation: TOA (Outgoing Longwave Radiation)*. NOAA NESDIS Center for Satellite Applications and Research. Retrieved from https://www.goes-r.gov/products/ATBDs/option2/RadBud_OLR_v2.0_no_color.pdf

Lovejoy, S. (2023). Review article: Scaling, dynamical regimes, and stratification. How long does weather last? How big is a cloud? *Nonlinear Processes in Geophysics*, *30*(3), 311–374. https://doi.org/10.5194/npg-30-311-2023

Lovejoy, S., & Schertzer, D. (2006). Multifractals, cloud radiances and rain. *Journal of Hydrology*, *322*(1–4), 59–88. https://doi.org/10.1016/j.jhydrol.2005.02.042

Mandorli, G., & Stubenrauch, C. J. (2023). Assessment of object-based indices to identify convective organization. *EGUsphere*, *2023*, 1–29. https://doi.org/10.5194/egusphere-2023-1985

Miura, H., Satoh, M., Nasuno, T., Noda, A. T., & Oouchi, K. (2007). A Madden-Julian Oscillation event realistically simulated by a global cloud-resolving model. *Science*, *318*(5857), 1763–1765. https://doi.org/10.1126/science.1148443

Mooers, G., Pritchard, M., Beucler, T., Srivastava, P., Mangipudi, H., Peng, L., et al. (2023). Comparing storm resolving models and climates via unsupervised machine learning. *Scientific Reports*, *13*(1), 22365. https://doi.org/10.48550/arXiv.2208.11843

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

Ohring, G., & Gruber, A. (1983). Satellite radiation observations and climate theory. *Advances in Geophysics*, 237–304. https://doi.org/10.1016/S0065-2687(08)60175-2

Palmer, T. N. (2016). A personal perspective on modelling the climate system. *Proceedings of the Royal Society*, *A*(2188), 20150772. https://doi.org/10.1098/rspa.2015.0772

Pierrehumbert, R. T. (1996). Anomalous scaling of high cloud variability in the tropical Pacific. *Geophysical Research Letters*, *23*(10), 1095–1098. https://doi.org/10.1029/96GL01121

Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*(11), 1547–1564. https://doi.org/10.1175/BAMS-84-11-1547

Schertzer, D., & Lovejoy, S. (1989). Generalised scale invariance and multiplicative processes in the atmosphere. *Pure and Applied Geophysics*, *130*(1), 57–81. https://doi.org/10.1007/BF00877737

Schmit, T. J., & Gunshor, M. M. (2020). Chapter 4—ABI imagery from the GOES-R Series. In S. J. Goodman, T. J. Schmit, J. Daniels, & R. J. Redmon (Eds.), *The GOES-R Series* (pp. 23–34). Elsevier. https://doi.org/10.1016/B978-0-12-814327-8.00004-4

Stephan, C. C., Duras, J., Harris, L., Klocke, D., Putman, W. M., Taylor, M., et al. (2022). Atmospheric energy spectra in global kilometre-scale models. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *74*, 280–299. https://doi.org/10.16993/tellusa.26

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., et al. (2019). DYAMOND: The DYnamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, *6*(1), 61. https://doi.org/10.1186/s40645-019-0304-z

Takasuka, D., Kodama, C., Suematsu, T., Ohno, T., Yamada, Y., Seiki, T., et al. (2024). How can we improve the seamless representation of climatological statistics and weather toward reliable global K-scale climate simulations? *Journal of Advances in Modeling Earth Systems*, *16*(2), e2023MS003701. https://doi.org/10.1029/2023MS003701

Tessier, Y., Lovejoy, S., & Schertzer, D. (1993). Universal multifractals: Theory and observations for rain and clouds. *Journal of Applied Meteorology*, *32*(2), 223–250. https://doi.org/10.1175/1520-0450(1993)032⟨0223:UMTAOF⟩2.0.CO;2

Vainshtein, S. I., Sreenivasan, K. R., Pierrehumbert, R. T., Kashyap, V., & Juneja, A. (1994). Scaling exponents for turbulence and other random processes and their relationships with multifractal structure. *Physical Review*, *50*(3), 1823–1835. https://doi.org/10.1103/PhysRevE.50.1823

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Williams, K., & Webb, M. (2008). A quantitative performance assessment of cloud regimes in climate models. *Climate Dynamics*, *33*(1), 141–157. https://doi.org/10.1007/s00382-008-0443-1

Zadra, A., Williams, K., Frassoni, A., Rixen, M., Ángel, F. A., Berner, J., et al. (2018). Systematic errors in weather and climate models: Nature, origins, and ways forward. *Bulletin of the American Meteorological Society*, *99*(4), ES67–ES70. https://doi.org/10.1175/BAMS-D-17-0287.1

References From the Supporting Information

Ellingson, R. G., Yanuk, D. J., Lee, H.-T., & Gruber, A. (1989). A technique for estimating outgoing longwave radiation from HIRS radiance observations. *Journal of Atmospheric and Oceanic Technology*, *6*(4), 706–711. https://doi.org/10.1175/1520-0426(1989)006⟨0706:ATFEOL⟩2.0.CO;2