Systematic Calibration of A Convection-Resolving Model: Application over Tropical Atlantic

Shuchang Liu[1], Christian Zeman[2], Silje Lund Sørland[3], and Christoph Schär[4]

[1]IAC ETHz
[2]IAC ETHZ
[3]Institute for Atmospheric and Climate Science
[4]ETH Zurich

November 26, 2022

Abstract

Non-hydrostatic km-scale weather and climate models show significant improvements in simulating clouds, especially convective ones. However, even km-scale models need to parameterize some physical processes and are thus subject to the corresponding uncertainty of parameters. Systematic calibration has the advantage of improving model performance with transparency and reproducibility, thus benefiting model intercomparison projects, process studies, and climate-change scenario simulations. In this paper, the regional atmospheric climate model COSMO v6 is systematically calibrated over the Tropical South Atlantic. First, the parameters' sensitivities are evaluated with respect to a set of validation fields. Five of the most sensitive parameters are chosen for calibration. The objective calibration then closely follows a methodology extensively used for regional climate simulations. This includes simulations considering the interaction of all pairs of parameters, and the exploitation of a quadratic-form metamodel to emulate the simulations. In the current set-up with 5 parameters, 51 simulations are required to build the metamodel. The model is calibrated for the year 2016 and validated in two different years using slightly different model setups (domain and resolution). Both years demonstrate significant improvements, in particular for outgoing shortwave radiation, with reductions of the bias by a factor of 3 to 4. The results thus show that parameter calibration is a useful and efficient tool for model improvement. Calibrating over a larger domain might help improve the overall performance, but could potentially also lead to compromises among different regions and variables, and require more computational resources.

Systematic Calibration of A Convection-Resolving Model: Application over Tropical Atlantic

Shuchang Liu[1], Christian Zeman[1], Silje Lund Sørland[2], Christoph Schär[1]

[1]Institute for Atmospheric and Climate Science, ETH Zürich
[2]NORCE Norwegian Research Centre, Norway

Key Points:

- A systematic calibration method is applied to improve the performance of a km-resolution regional climate model over the tropical Atlantic.
- Cloud-related model performance at the km-scale is significantly improved through systematic calibration.
- The calibrated parameter setting is robust among different years.

Corresponding author: Shuchang Liu, `Shuchang.liu@env.ethz.ch`

Abstract

Non-hydrostatic km-scale weather and climate models show significant improvements in simulating clouds, especially convective ones. However, even km-scale models need to parameterize some physical processes and are thus subject to the corresponding uncertainty of parameters. Systematic calibration has the advantage of improving model performance with transparency and reproducibility, thus benefiting model intercomparison projects, process studies, and climate-change scenario simulations.

In this paper, the regional atmospheric climate model COSMO v6 is systematically calibrated over the Tropical South Atlantic. First, the parameters' sensitivities are evaluated with respect to a set of validation fields. Five of the most sensitive parameters are chosen for calibration. The objective calibration then closely follows a methodology extensively used for regional climate simulations. This includes simulations considering the interaction of all pairs of parameters, and the exploitation of a quadratic-form metamodel to emulate the simulations. In the current set-up with 5 parameters, 51 simulations are required to build the metamodel. The model is calibrated for the year 2016 and validated in two different years using slightly different model setups (domain and resolution). Both years demonstrate significant improvements, in particular for outgoing shortwave radiation, with reductions of the bias by a factor of 3 to 4.

The results thus show that parameter calibration is a useful and efficient tool for model improvement. Calibrating over a larger domain might help improve the overall performance, but could potentially also lead to compromises among different regions and variables, and require more computational resources.

1 Introduction

While the critical role of anthropogenic greenhouse gases for the climate system is widely accepted (IPCC, 2021), the uncertainties in climate projections are still staggeringly large. Current uncertainties limit the ability to plan climate-change adaptation measures, weakening the debate about climate-change mitigation. Reducing these uncertainties is thus of key importance.

Studies have found that the uncertainty in global mean warming in response to anthropogenic greenhouse gases in climate models is closely related to the representation of cumulus and stratocumulus clouds over tropical oceans, since they are controlled by dynamic processes at small scales (typically 0.1-10 km), which is significantly lower than the grid spacing of global climate models (50-100 km) (Bony & Dufresne, 2005; Sherwood et al., 2014; Bony et al., 2015; Schneider et al., 2017). Due to computational constraints, most global climate models still parameterize the moist-convective vertical exchange of energy, moisture and momentum, even in the tropics, where it is the key agent of atmospheric motion. However, during the last decade, tremendous efforts have become evident towards explicitly resolving convective clouds rather than using semi-empirical parameterization schemes (Satoh et al., 2019; Stevens et al., 2019; Schär et al., 2020). Several studies using limited area modeling have shown that the convection-resolving approach yields a significantly improved simulation of the diurnal cycle of precipitation (Prein et al., 2013), as well as a better representation of hourly precipitation statistics, wet and dry extremes (Kendon et al., 2019; Ban et al., 2014, 2015; Prein et al., 2017), cloud cover (Hentgen et al., 2019; Miyamoto et al., 2013) and wind (Belušić et al., 2018).

While the progress of convection-resolving models (CRMs) in the extratropics has been highly promising, recent studies suggest that the potential of CRMs in the tropics is even more exciting (Stevens et al., 2019; Hentgen et al., 2020). In the tropics, convection is a key process throughout all seasons and is closely linked to the Hadley circulation that features air rising near the Equator, flowing poleward in the upper tropical atmosphere, descending in the subtropics, and then returning equatorwards. This is one of the most important circulations in our climate system that functions as an atmospheric heat engine, and many studies have demonstrated that the spatial organization of subtropical and tropical clouds associated with the Hadley circulation can be represented more credible at high resolutions (Bretherton & Khairoutdinov, 2015; Heim & Hentgen, 2021). This concerns especially shallow cumulus and stratocumulus clouds (Hohenegger et al., 2020).

In spite of these improvements when going towards higher resolution, there are still some challenges. Although CRMs run at a relatively high resolution (typically lower than 4 km) (Prein et al., 2015), some processes still need to be parameterized, such as cloud microphysics and turbulence (Schär et al., 2020), which are approximations of subgrid-scale processes and rely on semi-empirical parameters that are poorly constrained by observations. Thus, when applying CRMs over the tropics, the simulations are subject to high parametric uncertainty related to poorly confined model parameters. In practice, the values of uncertain parameters are determined using subjective expert tuning. Normally, the tuning does not follow a unique well-defined methodology (Hourdin et al., 2017). Subjective model tuning implies some difficult challenges. For instance, differences in model results reflect both differences in model structure (such as dynamical cores and type of parameterizations) and model tuning, thereby hazing the value of model intercomparison projects. This is particularly important for cloud-radiative feedback, as the magnitudes of the anthropogenic forcing and cloud-radiative feedbacks are small, often smaller than the systematic model biases in terms of radiation budget (Stocker, 2014).

Compared with subjective tuning, systematic calibration methods, using a predefined mathematical framework to perform model tuning, possess the advantage of making the process more explicit and reproducible (Hourdin et al., 2017). The framework encompasses the validation strategy, the set of to-be-calibrated parameters, and the modeling strategy (period and domain). Within such a stipulated framework, the calibration is objective, but the definition of the framework is subjective. Thus, to ensure a valid intercomparison of different model versions (e.g., different resolutions or parameterizations) and an assessment of the parametric uncertainty, a systematic model calibration method is preferable (García-Díez et al., 2015; Bellprat et al., 2012, 2016).

Current calibration techniques mainly include two categories in terms of the optimization (Hourdin et al., 2017). One is fast optimization of some cost function, evaluating model performance given specific metrics like averaged radiation or precipitation (Neelin et al., 2010; Bellprat et al., 2012; Bracco et al., 2013; Duan et al., 2017; Langenbrunner & Neelin, 2017; Tett et al., 2017; Gorman & Oliver, 2018). The other, instead of trying to find the optimum parameter setting, involves using Bayesian approaches to provide the uncertainty for the parameters (Bony & Dufresne, 2005; Rougier, 2007; Sanderson, 2011; Sexton et al., 2012; Salter et al., 2019; Couvreux et al., 2021). Except for some studies that use particle-based approaches (Lee et al., 2020) or adaptive sampling algorithms (Phipps et al., 2021). Most of the research uses emulators, mapping model inputs with outputs to reduce computational resources. In terms of the emulators, the calibration methods can also be divided into those that use statistical models (Voudouri et al., 2021) and machine learning methods (Li et al., 2019).

In this study, We choose a fast optimization method given limited computational resources, and applied a simple statistical emulator for clearer input-output relationships. We systematically calibrated the non-hydrostatic fully compressible limited-area model of the Consortium for Small-Scale Modeling (COSMO) in climate mode (Steppeler et al., 2003; Doms & Förstner, 2004) and obtained optimistic parameter settings over the tropical Atlantic. The objective of this study is to examine the potential of systematic calibration in improving the model performance of cloud simulation over the tropics. Future applications will address the role of cloud-radiative feedbacks in climate change.

2 Materials and Methods

2.1 Numerical Simulations

The European Center for Medium-Range Weather Forecast (ECMWF) Re-Analysis (ERA5) data (Hersbach et al., 2020) is used as lateral boundaries to drive the COSMO v6 model. The parameterization schemes applied are similar as Heim and Hentgen (2021): deep and shallow convection parameterizations are switched off, radiative fluxes are computed following the δ-two-stream approach after Ritter and Geleyn (1992), the single-moment bulk scheme after Reinhardt and Seifert (2006) is used as cloud microphysics parameterisation, a 1D TKE-based model (Raschendorfer, 2001) is employed for the computation of subgrid-scale vertical turbulent flux and we use prescribed sea-surface temperature over the ocean.

All simulations are run with 60 vertical levels and a horizontal grid spacing of 4 km. For the sensitivity and calibration simulations, domain D01 is applied as displayed in Figure 1 with a size of 1000x575 grid columns. The simulation period covers 4 months (Feb., May, Aug., Nov.) in 2016, each with a 5-day-spin-up period. Based on previous calibration studies (Voudouri et al., 2018; Russo et al., 2020), 13 parameters that are thought to exert a significant impact on model results were tested, shown in Table 1. In the end, five of these parameters are selected for calibration, and the reasoning is elaborated in section 3.1. For validation of the optimized parameter setting, we proceed in two steps. First we present a validation over D01 with the same set-up as for the calibration. Second, we a larger validation domain D02 is used at a refined horizontal grid spacing of 3 km. It has a size of 2750x2065 grid columns. Both validation periods consider another year than the one used in calibration, to avoid overfitting of parameters.

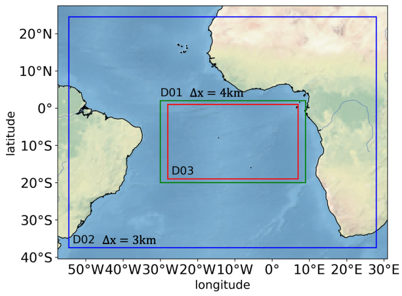

Figure 1. Simulation, calibration and validation domains. Domain D01 (green line) is used for the COSMO sensitivity and model calibration simulations. The calibration takes place in the subdomain D03 (red line). In addition, the large domain D02 (blue line) is also used for further validation.

2.2 Calibration

The calibration with N parameters optimises the parameter choice in an N-dimensional cube spanned by the min/max ranges of the selected parameters (see Table 1. To construct a metamodel, the following set of simulations are employed: the default simulation (all parameters at default value), pairs of sensitivity simulations (one parameter changed to min/max values), and quadruplets of interaction simulations (two parameters changed to min/max values). The total number of simulations is then $1 + 2N + 2N(N-1) = 2N^2 + 1$, and for $N = 5$ this yields 51 simulations. Based on this set of simulations, a metamodel is constructed, and the optimal value of the parameters is selected. The restriction to using only quadratic interactions (with two non-default values) in the set of simulations is consistent with the choice of the metamodel (see below). The set of sim-

Table 1. Perturbed parameters. The parameters selected for calibration are denoted in bold. The range covers the parameter values explored. The bold entries denote the default values in simulations. The same values have also been used by Hentgen et al. (2020).

Parameter/property	Acronym	Value Range
Turbulence		
Minimal diffusion coefficients for vertical heat and momentum transport (m^2s^{-1})	tkmin	[0, **0.4**, 2]
Maximal turbulent length scale (m)	**tur_len**	[60, **100**, 500]
Factor for turbulent momentum dissipation	d_mom	[12, **16.6**, 20]
Land surface		
Scaling factor for laminar boundary layer depth	rlam_heat	[0.1, **0.5249**, 2]
Scaling factor for laminar boundary layer depth over sea	**rat_sea**	[1, **20**, 100]
Surface area index of the waves over sea	c_sea	[1, **1.5**, 10]
Exponent to get the effective surface area	e_surf	[0.1, **1**, 10]
Microphysics		
Cloud ice threshold for autoconversion	**qi0**	[0, **5e-6**, 0.01]
Variable for computing the rate of cloud liquid water in unsaturated cases	**clc_diag**	[0.2, **0.5**, 1]
Cloud droplet number concentration	**cloud_num**	[1e7, **5e8**, 1e9]
Radiation		
Variable for computing the rate of cloud cover in unsaturated cases	uc1	[0, **0.0626**, 1.6]
Critical value for normalized oversaturation	q_crit	[1, **1.6**, 10]
Portion of gridscale qc seen by the radiation	radqc_fact	[0.5, **0.5**, 1]

ulations considered in the current study is shown in Table 2. The technical details of the calibration closely follow Bellprat et al. (2012). Significant differences concern the choice of the validation data, differences in the performance score, and the use of scaled parameter ranges (see below).

2.2.1 Performance score

Since the target is to improve cloud-related performance, top of atmosphere (TOA) radiative fluxes (outgoing longwave radiation (OLR) and outgoing shortwave radiation (OSR)) are chosen to calibrate the model results. Besides, the surface latent heat flux (LHFL) is also included as a target validation field, since it plays an important role in humidifying the atmosphere. Furthermore, LHFL also enables us to take a surface field into consideration, apart from the TOA fields. The TOA observation data is from Satellite Application Facility on Climate Monitoring (CM SAF) (Schulz et al., 2009). Since LHFL observation data is limited, ERA5 reanalysis data (Hersbach et al., 2020) is used to constrain this field. This special choice of validation data is owed to the limited availability of in-situ observations in the area of interest. A critical element of this choice is the use of ERA5 data for LHFL. The use of such data in the calibration hinges upon an appropriate estimate of the data's uncertainties.

The variables are evaluated using monthly means, averaged spatially for 28 rectangular regions ($5° \times 5°$ each, 4 rows and 7 columns over the calibration domain D03 as displayed in Figure 1). The error of these time series is measured using a performance score (PS):

$$PS = exp[-\frac{1}{2\,V\,RTY}\sum_{v}\sum_{r}\sum_{t}\sum_{y}\frac{(m_{v,r,t,y}-o_{v,r,t,y})^2}{\sigma^2_{o_{v,t}}+\sigma^2_{\epsilon_{v,t,y}}}]. \quad (1)$$

The Y, T, R, V in (1) denote the number of years used in the calibration framework (Y=1 with the year 2016), number of months used (T=4 monthly averages including Feb., May., Aug., Nov.), averaged over each region (R=28 regions), and for the three validation variables (OLR, OSR, LHFL, V = 3). PS is therefore an estimate of likelihood obtained by normalizing the simulated error ($m-o$) with interannual observation variation (σ_o) and observational uncertainty (σ_ϵ). The interannual variability (σ_o) is expressed as the interannual standard deviations of the monthly mean observations (2013-2017) averaged over the whole domain. The observational uncertainty (σ_ϵ) of OLR and OSR are from Urbain et al. (2017). The σ_ϵ of LHFL is from the standard deviation of the ERA5 assimilation ensemble members, which provides background-error estimates for the deterministic reanalysis system (Hersbach et al., 2019, 2020). Table 3 displays the σ_o and σ_ϵ used for the calibration.

Table 2. Summary of simulations: the sensitivity ensemble includes 2 simulations per parameter (with min and max parameter values); the interaction ensemble includes sensitivity simulations with all quadratic interactions; and the validation simulations include two simulations with default and calibrated parameter sets over two domains.

Ensemble	Domain	Period	Resolution	Parameters	Simulations
Default simulation	D01	Feb. May., Aug., Nov. 2016	4.4 km	def	1
Sensitivity tests	D01	Feb. May., Aug., Nov. 2016	4.4 km	13	26
Parameter interactions	D01	Feb., May., Aug., Nov. 2016	4.4 km	5	40
Validation01	D01	the whole year of 2013	4.4 km	-	2
Validation02	D02	Feb., May., Aug., Nov. 2006	3.3 km	-	2

Table 3. σ_o and σ_ϵ used for calibration.

σ	Fields (Wm^{-2})	Feb.	May	Aug.	Nov.
	OLR	10.0	16.0	8.7	17.2
σ_o	OSR	35.3	26.8	29.5	31.6
	LHFL	28.8	40.6	37.3	10.2
	OLR	4.9			
σ_ϵ	OSR	1.3			
	LHFL	11.5			

2.2.2 Metamodel

Since direct simulations with the convection-resolving model (CRM) are computationally expensive, a quadratic metamodel (MM) was chosen to emulate the output of the CRM (Neelin et al., 2010; Bellprat et al., 2012). The MM is based on the assumption that the climate model results from parameter perturbation are smooth and can be approximated by a 2^{nd} order polynomial regression. Interactions of parameter perturbations are approximated by a non-linear term for each parameter pair.

Relative parameter values μ_* and model fields Φ_* are used as independent and dependent variables separately to fit the MM. For each field, month and domain pixel, the corresponding formulations can be written as:

$$\boldsymbol{\mu}_* = \boldsymbol{\mu}_p - \boldsymbol{\mu}_{def}, \tag{2}$$

$$\Phi_* = \Phi_p - \Phi_{def}, \tag{3}$$

$$\Phi_p = f_{MM}(\boldsymbol{\mu}_*) + \Phi_{def}, \tag{4}$$

where subscripts def and p refer to default and perturbed parameter values, and f_{MM} indicates the polynomial function of MM. It includes one linear and one quadratic term for each relative parameter value and also one interactive term for every parameter pair (1^{st} order for each parameter in the pair). Depending on the number of parameters (N), f_{MM} can be expressed in the vector notation as

$$\Phi_* = \boldsymbol{\mu}_*^T \boldsymbol{a} + \boldsymbol{\mu}_*^T \boldsymbol{B} \boldsymbol{\mu}_*, \tag{5}$$

where the vector $\boldsymbol{a}$ contains the N linear coefficients for each parameter, and the matrix $\boldsymbol{B}$ includes coefficients for N quadratic terms on its diagonal and for $N(N-1)/2$ interactive terms in the off-diagonal elements (with the general assumption $\boldsymbol{B}_{i,j} = \boldsymbol{B}_{j,i}$). Together this yields $N(N+3)/2$ coefficients defining the MM. For example, if two parameters (μ_1, μ_2) are calibrated, f_{MM} would be

$$\Phi_* = \mu_1 a_1 + \mu_2 a_2 + \mu_1^2 b_1 + \mu_2^2 b_2 + 2\mu_1\mu_2 b_{1,2}, \tag{6}$$

where a_1, a_2, b_1, b_2 and $b_{1,2}$ are coefficients to be solved.

Perturbed parameter ensembles used to fit the MM are simulated through sampling parameters with their maximum and minimum possible values based on previous studies (Voudouri et al., 2018; Bellprat et al., 2016). Consequently, there are $2N^2$ simulations used to fit the MM, which is more than the number $N(N+3)/2$ of unknown coefficients. The resulting linear system of equations is thus overdetermined, and optimal interaction parameters are estimated using least squares error measures.

In general, the default value μ_{def} will not be in the center of the parameter range $[\mu_{min}, \mu_{max}]$, and this may lead to unsatisfactory results when fitting the MM. Parabolic fitting works best with a default value at the center, therefore we applied a logarithmic transformation of parameter values to fit the MM as Voudouri et al. (2018),

$$x \rightarrow \hat{x} \equiv log(\alpha \frac{x - x_{min}}{x_{max} - x_{min}} + \beta), \tag{7}$$

where the α and β are determined by parameter default values and ranges enabling $\hat{x}_{def} = (\hat{x}_{min} + \hat{x}_{max})/2$.

After the construction of the MM, 3,000,000 parameter sets are sampled with the Latin hypercube design (McKay et al., 2000). The set of parameter values with maximum PS was chosen as the optimal parameter set.

3 Results

3.1 Optimized parameters

Figure 2 presents the PS's of the sensitivity tests of the 13 parameters. The default PS (the black dots) indicates that LHFL performance is quite good, which is reasonable since we use the prescribed sea-surface temperature. Besides, as the domain D01 is mainly affected by low clouds, which hardly modify emitted longwave radiation from surface, the longwave radiation performance is also good. One of the target is to improve the representation of low clouds, which is related to variations in the OSR-field. Therefore, when choosing the final parameters for calibration, the ones that strongly impact OSR are the priority. Based on this figure the following parameters are selected for the calibration: tur_len, clc_diag, cloud_num, qi0 and rat_sea.

The choice follows the following considerations: First, tur_len, clc_diag and cloud_num have the largest potential in increasing OSR performance, with the largest OSR PS around 0.6. We also include two parameters to constrain OLR and LHFL. OLR is most sensitive to qi0, which controls the autoconversion of cloud ice and has almost no impact on OSR and LHFL. This makes qi0 a suitable parameter for calibration. For LHFL, rlam_heat and rat_sea exert the most significant impact. Since they have a similar impact over the ocean (rlam_heat controls the overall latent heat flux and rat_sea is a scaling factor exerted on the rlam_heat to distinguish sea and land) and the domain located over the ocean, rat_sea is chosen for calibration. Besides, according to Possner et al. (2014), it's better to use a small value for tkmin, thus in the calibration, it's set as 0.25.

Figure 3 displays the biases of longwave and shortwave radiation based on the sensitivity tests averaged over the four months (Feb., May, Aug. Nov.) in 2016. The OLR, OSR, LHFL are all defined as upward positive in this paper. Only the five calibrated parameters are displayed. The drastic impact of qi0 on longwave radiation can be seen when setting it to the maximum value. Because larger qi0 indicates less conversion of cloud ice to precipitable snow and more cloud ice would accumulate, thus preventing longwave radiation from escaping. The remaining parameters effectively control the shortwave radiation.

3.2 Calibration results

Once the coefficients of the metamodel have been determined from the calibration simulations, the optimal parameter setting is chosen based on a sampling of the five-dimensional cube. Figure 4 shows the resulting distribution of the PS. The PS increases from the default 0.62 (black line) to the optimum 0.86 (red line). This improvement is very substantial, but will require independent validation (see section 3.3). Figure 5 displays the corresponding distributions of PS as a function of the parameters. The default and optimized parameter values are shown by the black and red vertical lines. Results show that the parameter qi0 mainly affects high clouds and controls longwave radiation. Increasing qi0 results in lower values for OLR due to larger cloud ice content. The parameter for computing the rate of cloud liquid water in unsaturated cases (clc_diag) approaches 1, which indicates no subgrid-scale clouds. That is reasonable for high-resolution modeling due to smaller grid cells. The optimal value for tur_len is a bit lower than its default. This leads to less vertical mixing within the planetary boundary layer. This indicates decreased moisture supply and cloud amount. Besides, turbulence also affects the boundary layer stability and the inversion height (Heim & Hentgen, 2021), which indirectly influences the amount of low clouds. A shallower boundary layer favors the formation of low clouds, especially of persistent stratocumulus decks, yet a too shallow boundary layer top might be lower than the surface-determined lifting condensation level (LCL) and thus not allow clouds to form (Wood, 2012). Lower values of rat_sea favour higher surface latent heat fluxes. Clouds react to decreased rat_sea mainly in two ways. One is higher PBL moisture which allows for more cloud water. The other is decreased boundary layer stability, which may not favor the formation of low clouds. Furthermore, a lower value of cloud_num results in a larger cloud droplet size. That leads to increased precipitation, and might thus decrease cloud amount. In the mean time, reduced cloud_num also suppresses buoyant turbulence kinetic energy (TKE) production, thus may decrease

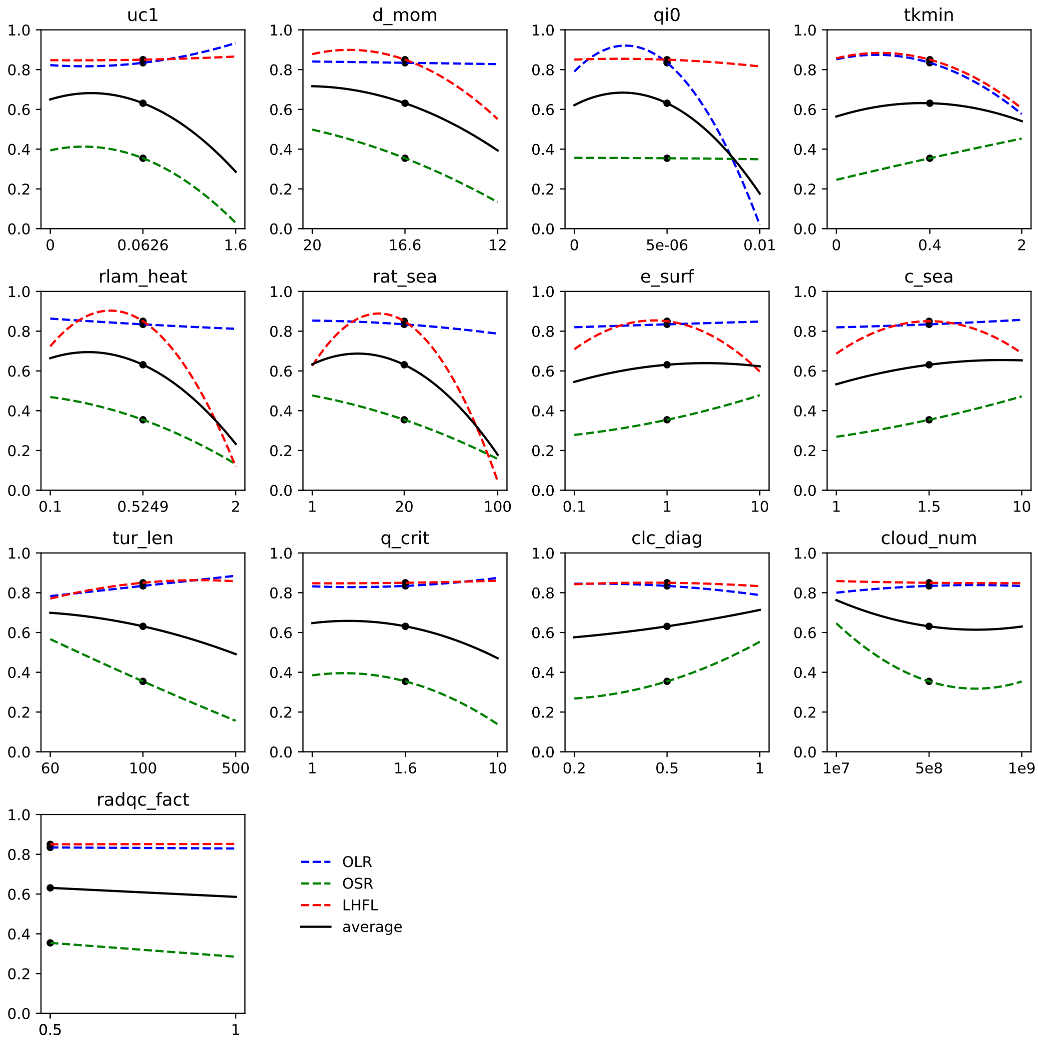

Figure 2. PS calculated separately for OLR (blue dashed line), OSR (green dashed line) and LHFL (red dotted line) and the PS for all of the 3 fields (black solid line) for the three tested parameter values. The results are the averaged over the four months and the analysis domain. The black dots indicate the respective PS with the default parameter setting. The horizontal axes shows the parameter values after the logarithmic transformation, and the lines represent quadratic fits.

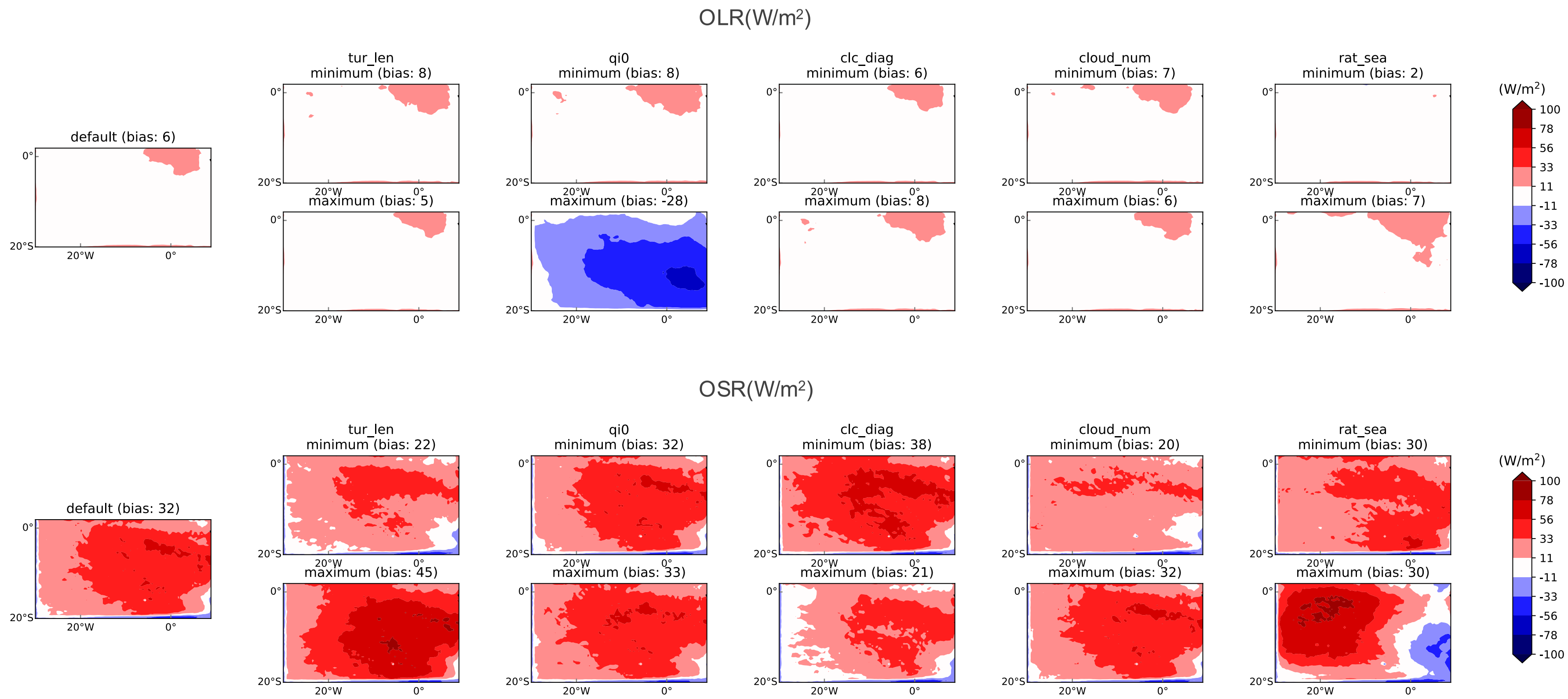

Figure 3. Biases of outgoing longwave radiation (OLR, upper panels) and outgoing shortwave radiation (OSR, lower panels) of the 4 km simulation for the sensitivity tests averaged over the four months (Feb., May, Aug. Nov.) in 2016. In each sub-figure the biases (model - observation) for the minimum and maximum parameter value are shown in the upper and lower panels, respectively (see also Table 1).

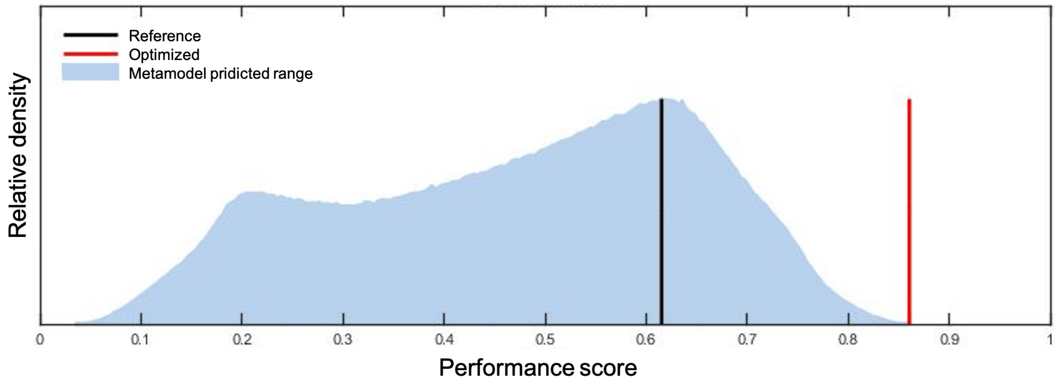

Figure 4. Metamodel predicted PS distributions for the 3,000,000 sampled parameter combinations (blue histogram) with the Latin hypercube method along with the original score of the reference (black line) and the optimized (red line) simulation.

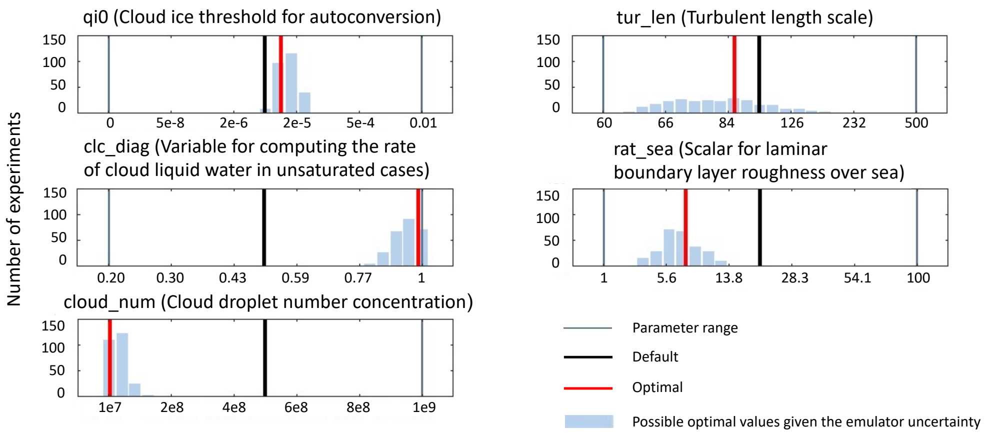

Figure 5. Number of experiments (blue histograms) of the parameter settings, which perform equally well, given the uncertainty of the metamodel in predicting the model performance (with an uncertainty of 0.015). The blue lines indicate the parameter range, the black line indicates the default parameter value and the red line indicates the optimum parameter values.

cloud-top entrainment and increase cloud amount (Coakley Jr & Walsh, 2002; Ackerman et al., 2004).

3.3 Robustness of the optimized parameter setting

To verify the calibration and the key result in Figure 4, the default simulation for the year 2016 has been repeated with the calibrated parameter settings. This confirmed the results and showed an improvement in PS from 0.62 before calibration, to 0.86 after calibration. The agreement with the metamodel is surprisingly good, as the optimal performance score is missed by less than a percent.

To test whether the calibrated parameter setting also works for another year, Figure 6 displays the comparison between simulations using the optimized parameter setting as described before and the default simulation during four full seasons in 2013 with domain D01: December, January and February (DJF), March, April and May (MAM),

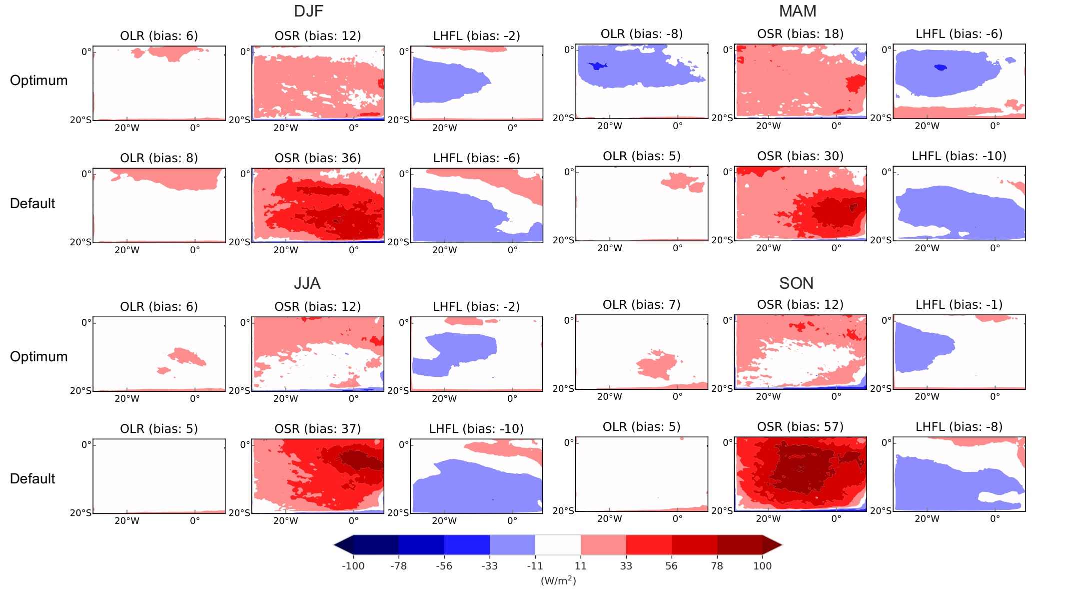

Figure 6. Validation of the optimum parameter setting in 2013 for December, January and February (DJF), March, April and May (MAM), June, July and August (JJA), September, October and November (SON). (bias = model - observation)

June, July and August (JJA), September, October and November (SON). The model performance is significantly improved in all seasons for shortwave radiation and surface latent heat flux. OLR is mainly affected by high clouds, whereas the spatial domain is dominated by low clouds for most of the seasons. Therefore, the change in OLR is minor. In MAM, when the ITCZ is southernmost and partially within the simulation domain, there is a significant underestimation of OLR, and an increase of the bias with the calibration. This kind of effect is to be expected with, as with the use of a PS there may be compensation of errors. In this particular case, the large OSR bias in the default is being reduced, but at the cost of increases in the OLR bias. The underestimation in MAM is mainly due to the overestimated ice cloud in the ITCZ. Therefore, the longwave radiation bias in MAM might indicate a deficiency of the model in simulating the high clouds with the same set of optimum parameters obtained over the current domain (since more weight is given to the low clouds due to the selection of the domain). However, overall PS is reduced, corresponding to a net reduction of the weighted overall bias.

The daily bias over the domain D01 in 2013 is presented in Figure 7. For the longwave radiation, the bias is almost the same between the optimum and default setting for most of the time. However, in April and May, where the ITCZ moves to the Southernmost, the bias with the optimum parameter setting is significantly higher than with the default setting. For shortwave radiation, there is a systematic decrease of bias using the optimum parameter setting, especially in austral winter and spring, when low cloud prevails. It should be noted that the consideration of daily biases includes biases due to predictability limitations and chaotic processes in the model domain.

To further explore how robust the optimum parameter setting is, we use another year (2006) and an extended simulation domain (D02 as displayed in Figure 1) for validation. Due to the limitation of computational resources, we only simulated 4 months (Feb., May, Aug., Nov.) to represent each season. Figure 8 shows the comparison between the optimized parameter setting and the default ones averaged over four months (Feb., May, Aug., Nov.). Table 4 lists the biases for the simulations with the optimum

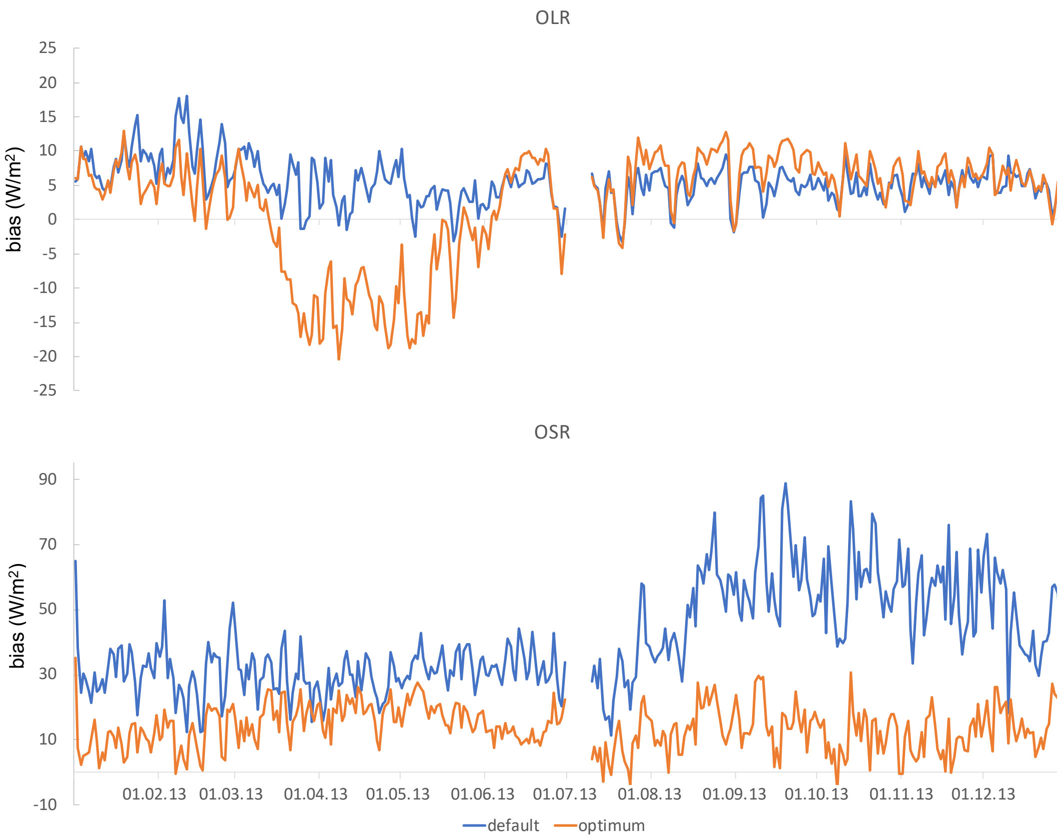

Figure 7. Comparison of daily bias averaged over domain D01 in 2013 between the optimum and default setting. (The data gap between July 1st-9th is due to missing satellite data.)

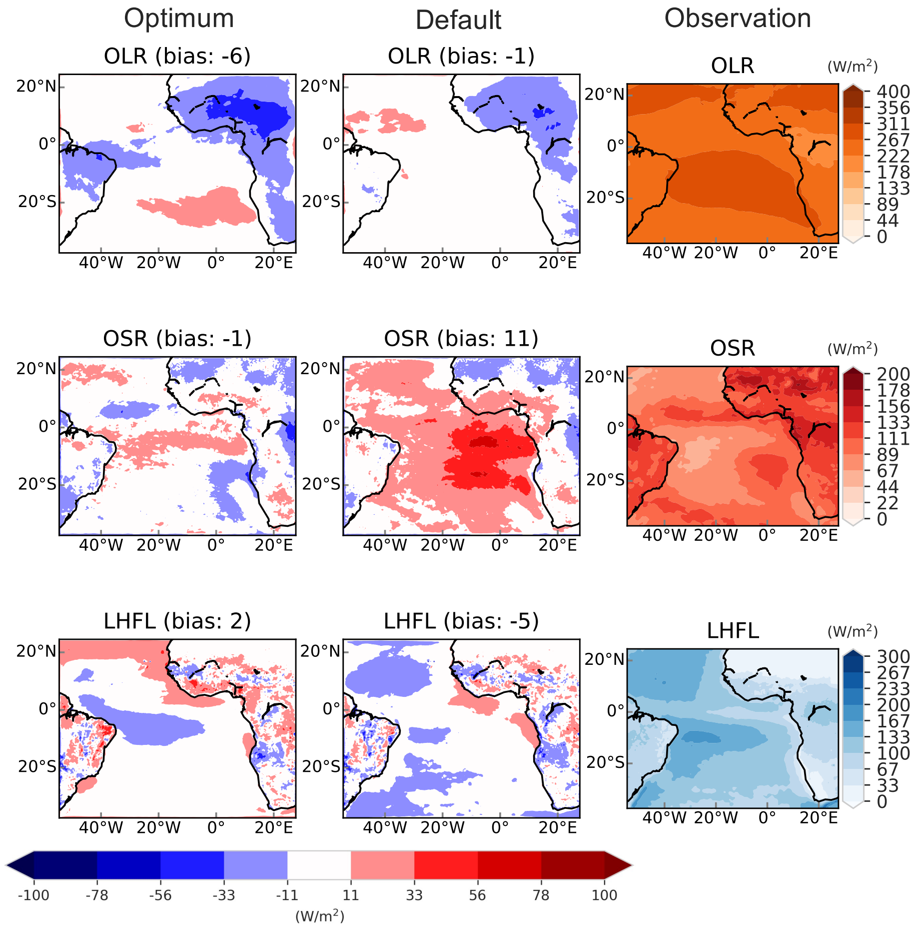

Figure 8. Validation of the optimum parameter setting averaged over four months (Feb., May, Aug., Nov.) in 2006. (bias = model - CM SAF observation)

and default setting for 2006 over the whole domain D02 and calibration domain D03. Within D03 (Figure 1), the performance improved substantially, where the OSR bias decreased from 25, 25, 36, 53 Wm^{-2} under the default setting to 4, 12, 2, 3 Wm^{-2} under the optimum setting in Feb., May, Aug., Nov. respectively. OLR performance has also improved, except for May. The deteriorated underestimation of OLR in May with the optimum setting might be due to the impact of the ITCZ, which is a similar case as the validation results in 2013 (Figure 6). These results indicate that the optimum parameter setting is robust for different years and slightly different resolutions (4 km versus 3 km). When taking the remaining part of the domain D02 (Figure 1) into consideration, the performances still improve significantly for OSR and LHFL. The four months average bias in 2006 decreased from 11 to -1 Wm^{-2} for OSR and from -5 to 2 Wm^{-2} for LHFL. For OLR, it is evident that the optimum simulation underestimates OLR over the ITCZ (Figure 8), and Table 4 shows that overall D02 domain average OLR is underestimated in all four months. Because D02 encompasses the ITCZ during all four months. This is consistent with the aforementioned result that the set of parameters that suits low clouds over sea might not apply as well for ITCZ.

Table 4. Comparison of bias between optimum and default simulation in 2006

Month	Spacial range	OLR (Wm^{-2})		OSR (Wm^{-2})		LHFL (Wm^{-2})	
		Default	Optimum	Default	Optimum	Default	Optimum
Four months average	D03	4	-1	35	5	-2	-6
	D02	-1	-6	11	-1	-5	2
Feb.	D03	6	1	25	4	-2	-5
	D02	-2	-6	11	0	-5	2
May	D03	2	-14	25	12	-2	-8
	D02	0	-7	7	-1	-5	3
Aug.	D03	6	6	36	2	-6	-9
	D02	0	-4	13	-2	-6	1
Nov.	D03	3	1	53	3	2	-2
	D02	-2	-7	15	-3	-4	2

4 Summary and conclusions

In this paper, the regional climate model COSMO v6 was systematically calibrated over the Tropical South Atlantic. First, the most sensitive parameters were identified with respect to the target fields that are important for the representation of clouds (shortwave/longwave radiation and surface latent heat flux). Based on sensitivity studies, a total of 5 parameterization parameters were selected for calibration. The calibration is based on single-parameter sensitivity experiments and simulations considering quadratic interactions. A metamodel (MM) is then used to emulate the model simulations. We applied Latin hypercube sampling and chose the set of parameters with the best performance score (PS) as the optimal parameter set.

We calibrated the COSMO v6 model in 2016 and validated the results in 2013 and 2006 in two different computational domains. With the calibrated optimal parameter settings, the performance improved significantly compared with the default parameter setting, especially for OSR. Even when we applied the optimal setting over a signficantly extended domain with a slightly higher resolution (3 km versus 4 km), the optimal setting also showed significant improvements. However, since the calibrated domain is dominated by the ocean and the impact of ITCZ in the domain is small, applying the obtained optimal parameter setting over land and the northern part of the domain encounters problems, especially for OLR, which is highly relevant with ITCZ high clouds. Thus, calibrating over a larger domain might improve the overall performance, but would potentially also lead to compromises among different regions and variables, and would require more computational resources to achieve improved results for the whole domain.

Besides the aforementioned performance improvements, another advantage of the systematic calibration applied in this study is that it could benefit model intercomparisons, process studies and climate-change scenario simulations. The traditional way of tuning a model does not follow a unique well-defined methodology and thus hazes the value of model intercomparisons. Instead, systematic calibration, based on a well-defined methodology, is promising in constraining parameterization-related uncertainties with transparency and reproducibility. Moreover, the calibration methodology, which is provided as an open source code with this paper, is independent of the target model and validation fields, and could be easily applied to other models and research domains.

Using regional climate model (RCM) simulations with prescribed lateral boundary conditions from reanalysis fields in model calibration, as presented in the current study, provides substantial advantages over using calibration with global climate models (GCMs). In a GCM there will in general be significant circulation biases. For instance, biases in polar regions will affect the circulation in tropical regions, and a calibration will at least partly attempt to compensate for associated circulation biases. With RCMs driven by reanalyses, the calibration targets the parameterization suite with realistic large-scale circulations. As a result, the RCM approach requires much shorter calibration and validation periods, as demonstrated by our study. Indeed, we used merely 4 months of a particular year for the calibration, and have demonstrated that this significantly improves simulations in other years and extended domains. It is thus attractive to consider a combined GCM/RCM calibration framework, that considers both approaches. Indeed, there is an increasing number of GCMs that are available in both limited-area and global configurations, such as the ICON model (Pham et al., 2021) or the Unified Model (Bush et al., 2020). With such models, it is feasible to combine RCM-style calibrations in subdomains. For instance, one could calibrate boundary-layer and warm microphysics parameters over tropical oceans, snow and ice microphysics parameters over polar regions, and land-surface parameters over major continental regions. We believe that this kind of approach would be superior in comparison with conventional GCM model tuning, and provide a more physically based set of model parameters.

There are a number of fundamental limitations with model calibration. First of all, it can only improve parameterization-related model performance of the subjectively predefined validation fields. It is thus important to select a broad range of validation data sets. Second, there are compensations of errors between different variables and areas. Since the model itself is not perfect (i.e. will have biases irrespective of the parameter choices), compensation of errors cannot be completely avoided. Third, emulators are necessary within the calibration framework, since it is impossible to traverse the parameter space with the climate model. In this study, we used deterministic polynomial regression to build the emulator, which already provided enough accuracy as indicated in section 3.3, but emulators inevitably bring in uncertainties. Nevertheless, we believe that the results achieved in this study are very promising and suggest that regional climate models should more systematically be calibrated than in the past.

Acknowledgments

The calibration source code is available under https://github.com/shucliu/Systemetic_calibration. The corresponding model output data is available open source. This research is funded by the Swiss National Science Foundation (SNSF) funds. We also acknowledge PRACE for awarding compute resources for the COSMO simulations on Piz Daint at the Swiss National Supercomputing Centre (CSCS). The authors declare no conflicts of interest relevant to this study.

References

Ackerman, A. S., Kirkpatrick, M. P., Stevens, D. E., & Toon, O. B. (2004). The impact of humidity above stratiform clouds on indirect aerosol climate forcing. *Nature*, *432*(7020), 1014–1017.

Ban, N., Schmidli, J., & Schär, C. (2014). Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmospheres*, *119*(13), 7889–7907.

Ban, N., Schmidli, J., & Schär, C. (2015). Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophysical Research Letters*, *42*(4), 1165–1172.

Bellprat, O., Kotlarski, S., Lüthi, D., De Elía, R., Frigon, A., Laprise, R., & Schär, C. (2016). Objective calibration of regional climate models: application over

Europe and North America. *Journal of Climate*, *29*(2), 819–838.
Bellprat, O., Kotlarski, S., Lüthi, D., & Schär, C. (2012). Objective calibration of regional climate models. *Journal of Geophysical Research: Atmospheres*, *117*(D23).
Belušić, A., Prtenjak, M. T., Güttler, I., Ban, N., Leutwyler, D., & Schär, C. (2018). Near-surface wind variability over the broader Adriatic region: insights from an ensemble of regional climate models. *Climate dynamics*, *50*(11), 4455–4480.
Bony, S., & Dufresne, J. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*(20).
Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., ... Sobel, A. H. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, *8*(4), 261–268.
Bracco, A., Neelin, J. D., Luo, H., McWilliams, J. C., & Meyerson, J. E. (2013). High dimensional decision dilemmas in climate models. *Geosci. Model Dev.*, *6*(5), 1673–1687. doi: 10.5194/gmd-6-1673-2013
Bretherton, C. S., & Khairoutdinov, M. F. (2015). Convective self-aggregation feedbacks in near-global cloud-resolving simulations of an aquaplanet. *Journal of Advances in Modeling Earth Systems*, *7*(4), 1765–1787.
Bush, M., Allen, T., Bain, C., Boutle, I., Edwards, J., Finnenkoetter, A., ... Zerroukat, M. (2020, 4). The first Met Office Unified Model–JULES Regional Atmosphere and Land configuration, RAL1. *Geosci. Model Dev.*, *13*(4), 1999–2029. doi: 10.5194/gmd-13-1999-2020
Coakley Jr, J. A., & Walsh, C. D. (2002). Limits to the aerosol indirect radiative effect derived from observations of ship tracks. *Journal of the atmospheric sciences*, *59*(3), 668–680.
Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., ... Xu, W. Z. (2021). Process-Based Climate Model Development Harnessing Machine Learning: I. A Calibration Tool for Parameterization Improvement. *Journal of Advances in Modeling Earth Systems*, *13*(3). doi: 10.1029/2020MS002217
Doms, G., & Förstner, J. (2004). Development of a kilometer-scale NWP-system: LMK. *COSMO newsletter*, *4*, 159–167.
Duan, Q., Di, Z., Quan, J., Wang, C., Gong, W., Gan, Y., ... Fan, S. (2017). Automatic Model Calibration: A New Way to Improve Numerical Weather Forecasting. *Bulletin of the American Meteorological Society*, *98*(5), 959–970. doi: 10.1175/BAMS-D-15-00104.1
García-Díez, M., Fernández, J., & Vautard, R. (2015). An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation. *Climate dynamics*, *45*(11), 3141–3156.
Gorman, R. M., & Oliver, H. J. (2018). Automated model optimisation using the Cylc workflow engine (Cyclops v1.0). *Geosci. Model Dev.*, *11*(6), 2153–2173. doi: 10.5194/gmd-11-2153-2018
Heim, C., & Hentgen, L. (2021). *Inter-model Variability in Convection-resolving Simulations of 2 Subtropical Marine Low Clouds* (Tech. Rep.).
Hentgen, L., Ban, N., Kröner, N., Leutwyler, D., & Schär, C. (2019). Clouds in convection-resolving climate simulations over Europe. *Journal of Geophysical Research: Atmospheres*, *124*(7), 3849–3870.
Hentgen, L., Ban, N., Vergara-Temprado, J., & Schär, C. (2020). Improving the simulation of tropical clouds in explicit high-resolution climate models. *Submitted to Journal of Advances in Modeling Earth Systems*(December), 1–23.
Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., ... Rozum, I. (2019). *ERA5 monthly averaged data on single levels from 1979 to present, Copernicus Climate Change Service (C3S) Climate Data Store*

(CDS).
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . Schepers, D. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.
Hohenegger, C., Kornblueh, L., Klocke, D., Becker, T., Cioni, G., Engels, J. F., . . . Stevens, B. (2020). Climate statistics in global simulations of the atmosphere, from 80 to 2.5 km grid spacing. *Journal of the Meteorological Society of Japan. Ser. II*.
Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., . . . Qian, Y. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602.
IPCC. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.
Kendon, E. J., Stratton, R. A., Tucker, S., Marsham, J. H., Berthou, S., Rowell, D. P., & Senior, C. A. (2019). Enhanced future changes in wet and dry extremes over Africa at convection-permitting scale. *Nature communications*, *10*(1), 1–14.
Langenbrunner, B., & Neelin, J. D. (2017). Multiobjective constraints for climate model parameter choices: Pragmatic Pareto fronts in CESM1. *Journal of Advances in Modeling Earth Systems*, *9*(5), 2008–2026. doi: 10.1002/2017MS000942
Lee, B., Haran, M., Fuller, R. W., Pollard, D., & Keller, K. (2020). A Fast Particle-Based Approach for Calibrating a 3-D Model of the Antarctic Ice Sheet. *Annals of Applied Statistics*, *14*(2), 605–634. doi: 10.1214/19-AOAS1305
Li, S. H., Rupp, D. E., Hawkins, L., Mote, P. W., McNeall, D., Sparrow, S. N., . . . Wettstein, J. J. (2019). Reducing climate model biases by exploring parameter space with large ensembles of climate model simulations and statistical emulation. *Geosci. Model Dev.*, *12*(7), 3017–3043. doi: 10.5194/gmd-12-3017-2019
McKay, M. D., Beckman, R. J., & Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *42*(1), 55–61.
Miyamoto, Y., Kajikawa, Y., Yoshida, R., Yamaura, T., Yashiro, H., & Tomita, H. (2013). Deep moist atmospheric convection in a subkilometer global simulation. *Geophysical Research Letters*, *40*(18), 4922–4926.
Neelin, J. D., Bracco, A., Luo, H., McWilliams, J. C., & Meyerson, J. E. (2010). Considerations for parameter optimization and sensitivity in climate models. *Proceedings of the National Academy of Sciences*, *107*(50), 21349–21354. doi: 10.1073/pnas.1015473107
Pham, T. V., Steger, C., Rockel, B., Keuler, K., Kirchner, I., Mertens, M., . . . Früh, B. (2021, 2). ICON in Climate Limited-area Mode (ICON release version 2.6.1): a new regional climate model. *Geosci. Model Dev.*, *14*(2), 985–1005. doi: 10.5194/gmd-14-985-2021
Phipps, S. J., Roberts, J. L., & King, M. A. (2021). An iterative process for efficient optimisation of parameters in geoscientific models: a demonstration using the Parallel Ice Sheet Model (PISM) version 0.7.3. *Geosci. Model Dev.*, *14*(8), 5107–5124. doi: 10.5194/gmd-14-5107-2021
Possner, A., Zubler, E., Fuhrer, O., Lohmann, U., & Schär, C. (2014). A case study in modeling low-lying inversions and stratocumulus cloud cover in the Bay of Biscay. *Weather and forecasting*, *29*(2), 289–304.
Prein, A. F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N. K., Keuler, K., & Georgievski, G. (2013). Added value of convection permitting seasonal simulations. *Climate Dynamics*, *41*(9-10), 2655–2677.
Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., . . . Feser, F. (2015). A review on regional convection-permitting climate modeling:

Demonstrations, prospects, and challenges. *Reviews of geophysics*, *53*(2), 323–361.
Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., & Holland, G. J. (2017). The future intensification of hourly precipitation extremes. *Nature Climate Change*, *7*(1), 48–52.
Raschendorfer, M. (2001). The new turbulence parameterization of LM. *COSMO newsletter*, *1*, 89–97.
Reinhardt, T., & Seifert, A. (2006). A three-category ice scheme for LMK. *Cosmo Newsletter*, *6*, 115–120.
Ritter, B., & Geleyn, J.-F. (1992). A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Monthly weather review*, *120*(2), 303–325.
Rougier, J. C. (2007). Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change*, *81*, 247–264. doi: 10.1007/s10584-006-9156-9
Russo, E., Sørland, S. L., Kirchner, I., Schaap, M., Raible, C. C., & Cubasch, U. (2020). Exploring the Parameters Space of the Regional Climate Model COSMO-CLM 5.0 for the CORDEX Central Asia Domain. *Geoscientific Model Development Discussions*, *2020*, 1–33.
Salter, J. M., Williamson, D. B., Scinocca, J., & Kharin, V. (2019). Uncertainty Quantification for Computer Models With Spatial Output Using Calibration-Optimal Bases. *Journal of the American Statistical Association*, *114*(528), 1800–1814. doi: 10.1080/01621459.2018.1514306
Sanderson, B. M. (2011). A Multimodel Study of Parametric Uncertainty in Predictions of Climate Response to Rising Greenhouse Gas Concentrations. *Journal of Climate*, *24*(5), 1362–1377. doi: 10.1175/2010JCLI3498.1
Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports*, *5*(3), 172–184.
Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Di Girolamo, S., … Leutwyler, D. (2020). Kilometer-scale climate models: Prospects and challenges. *Bulletin of the American Meteorological Society*, *101*(5), E567-E587.
Schneider, T., Teixeira, J., Bretherton, C. S., Brient, F., Pressel, K. G., Schär, C., & Siebesma, A. P. (2017). Climate goals and computing the future of clouds. *Nature Climate Change*, *7*(1), 3–5.
Schulz, J., Albert, P., Behr, H.-D., Caprion, D., Deneke, H., Dewitte, S., … Hechler, P. (2009). Operational climate monitoring from space: the EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF). *Atmospheric Chemistry and Physics*, *9*(5), 1687–1709.
Sexton, D. M. H., Murphy, J. M., Collins, M., & Webb, M. J. (2012). Multivariate probabilistic projections using imperfect climate models part I: Outline of methodology. *Climate Dyn.*, *38*, 2513–2542. doi: 10.1007/s00382-011-1208-9
Sherwood, S. C., Bony, S., & Dufresne, J.-L. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, *505*(7481), 37–42.
Steppeler, J., Doms, G., Schättler, U., Bitzer, H. W., Gassmann, A., Damrath, U., & Gregoric, G. (2003). Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteorology and atmospheric Physics*, *82*(1), 75–96.
Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., … Zhou, L. (2019, 12). *DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains* (Vol. 6) (No. 1). Springer Berlin Heidelberg. doi: 10.1186/s40645-019-0304-z
Stocker, T. (2014). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press.
Tett, S. F. B., Yamazaki, K., Mineter, M. J., Cartis, C., & Eizenberg, N. (2017).

Calibrating climate models using inverse methods: case studies with HadAM3, HadAM3P and HadCM3. *Geosci. Model Dev.*, *10*(9), 3567–3589. doi: 10.5194/gmd-10-3567-2017

Urbain, M., Clerbaux, N., Ipe, A., Tornow, F., Hollmann, R., Baudrez, E., ... Moreels, J. (2017). The CM SAF TOA radiation data record using MVIRI and SEVIRI. *Remote Sensing*, *9*(5), 466.

Voudouri, A., Avgoustoglou, E., Carmona, I., Levi, Y., Bucchignani, E., Kaufmann, P., & Bettems, J. M. (2021). Objective Calibration of Numerical Weather Prediction Model: Application on Fine Resolution COSMO Model over Switzerland. *ATMOSPHERE*, *12*(10). doi: 10.3390/atmos12101358

Voudouri, A., Khain, P., Carmona, I., Avgoustoglou, E., Kaufmann, P., Grazzini, F., & Bettems, J. M. (2018). Optimization of high resolution COSMO model performance over Switzerland and Northern Italy. *Atmospheric research*, *213*, 70–85.

Wood, R. (2012). Stratocumulus clouds. *Monthly Weather Review*, *140*(8), 2373–2423.