JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2018MS001449

Optimization of the Eddy-Diffusivity/Mass-Flux Shallow Cumulus and Boundary-Layer Parameterization Using Surrogate Models

W. Langhans[1], **J. Mueller**[2], **and W. D. Collins**[1,3]

[1]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [2]Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [3]Department of Earth and Planetary Science, University of California, Berkeley, CA, USA

Key Points:
- Radial basis function (RBF) surrogate models provide an efficient framework to optimize physics parameterizations such as EDMF
- The RBF surrogate models detect solutions much better than those provided by a commonly used quadratic polynomial model and are less costly
- A multiplume model without eddy-diffusivity flux emerges as one of the optimal settings for the cloud-free and cloud-topped convective PBL

Correspondence to:
W. Langhans,
wlanghans@lbl.gov

Citation:
Langhans, W., Mueller, J., & Collins, W. D. (2019). Optimization of the eddy-diffusivity/mass-flux shallow cumulus and boundary-layer parameterization using surrogate models. *Journal of Advances in Modeling Earth Systems*, *11*. https://doi.org/10.1029/2018MS001449

Received 18 JUL 2018
Accepted 7 JAN 2019
Accepted article online 10 JAN 2019

Abstract Physical parameterizations in global atmospheric and ocean models typically include free parameters that are not theoretically or empirically constrained. New methods are required to determine the optimal parameter combinations for such models in an objective, exhaustive, yet computationally feasible manner. Here we propose to apply computationally inexpensive radial basis function (RBF) surrogate models to minimize a "cost," or error, function of an atmospheric model or a physical parameterization. The RBF is iteratively updated as more input-output pairs are obtained during the optimization. The approach is used to optimize the eddy-diffusivity/mass-flux boundary-layer parameterization of the convective boundary-layer in a single-column model framework. The optimization based on surrogate models is able to identify parameter combinations that reduce the error of the untuned default setting by 41%. The probability to detect a low-error solution increases rapidly especially over the first tens of single-column model evaluations. In comparison, a quadratic polynomial model only yields an error reduction of 17% since (a) high-order parameter interactions are not accounted for and (b) the construction of the polynomial is not based on an iterative sampling approach. The RBF surrogate models achieve this 17% error reduction for 40% of the polynomial model's cost. Interestingly, one of the emerging optimal settings describes a pure mass flux parameterization without eddy-diffusivity component. A second optimal solution is characterized by a plume fractional area of 20–30% and an eddy-mixing time scale of ∼700 s.

1. Introduction

Global atmospheric, ocean, and Earth system models are steadily becoming more costly due to the refinement of numerical grid resolutions and the ongoing efforts to incorporate new climate system components and physical process parameterizations. For example, global atmospheric models will attain global convection-permitting resolutions with O(1 km) grid spacings with the advent of exascale computers. In addition, parameterizations for microphysical, aerosol, turbulent, and other physical processes are growing in complexity and in the number of free parameters. The latter are unfortunately often poorly constrained by either observations or theory. As a result, parameterizations are improved through calibration, also known as "tuning," to increase the fidelity of the model relative to reference data sets. This may imply the minimization of a cost function or the introduction of a Bayesian formulation of the calibration problem (Kennedy & O'Hagan, 2001). In this paper we are interested in the former, an appreciably faster approach to parameter optimization. To minimize the computational intensity of model tuning while simultaneously reducing model error as much as practicable, computationally inexpensive methods are needed in order to identify the parameter configuration(s) that lead to the optimal outcome.

The optimization of weather and climate applications presents several challenges. First, there are typically tens of tunable parameters. Second, each individual simulation is very expensive. Third, an analytical expression of the cost function and its derivatives is not available (i.e., black box). And finally, the cost function can be expected to be multimodal with several local minima. Therefore, gradient based optimization methods (see Nocedal & Wright, 1999) are usually not applicable because they require gradient information (the approximation of gradients requires too many function evaluations) and usually stop at a locally

optimal solution. Evolutionary algorithms such as genetic algorithms (Holland, 1992) do not require gradient information, but they too require a large number of function evaluations (often many hundreds or thousands) in order to converge to a solution. In order to efficiently solve computationally expensive black-box optimization problems, Booker et al. (1999) introduced the idea of exploiting computationally cheap surrogate models to approximate the expensive simulation throughout the optimization search. The main idea is to exploit the predictions of the surrogate model when making iterative sampling decisions. Different types of surrogate models have been developed in the literature. For example, polynomial regression models (Myers & Montgomery, 1995) have widely been used to study the relationship between input parameters and the experimental or simulation response. As reviewed by Hourdin et al. (2017), such approaches have also been applied to atmospheric modeling in the past. For example, Neelin et al. (2010) successfully used a quadratic polynomial model to tune a global climate model and the idea has been extended to include cheaper lower-resolution versions of a climate model (Williamson et al., 2012). The quadratic model has led to success also in regional climate modeling (Bellprat et al., 2012) where climate fields were found to vary smoothly in parameter space. However, the applicability of quadratic polynomial models to the tuning of small-scale processes like boundary-layer clouds has not yet been explored.

Interpolation approximation models such as radial basis functions (RBFs; Powell, 1992) have been used in many optimization algorithms (Müller et al., 2013; Regis & Shoemaker, 2007, 2013) and applied to a variety of science problems, for example, watershed management (Müller & Woodbury, 2017), methane transport in the community land model (Müller et al., 2015), and the multiobjective design of airfoils (Müller, 2017). Closely related to using surrogate models for optimization purposes are Gaussian process emulation approaches that have been used to quantify the uncertainty of microphysical parameters (Johnson et al., 2015; Lee et al., 2012). In this paper, we make use of RBF models to tune a physics parameterization for weather and climate applications and compare the approach to Neelin et al.'s (2010) quadratic model.

The largest uncertainties in climate projections have been shown to originate from the parameterization of clouds (Bony & Dufresne, 2005; Bony et al., 2006; Cess et al., 1989; Soden & Held, 2006; Webb et al., 2013). The parameterization of convective clouds is of particular concern since it was identified as the leading source for differences in climate sensitivities among models (Zhao, 2014; Zhao et al., 2016). The optimization approach is therefore tested here for the optimization of a shallow-cumulus and boundary-layer parameterization. This type of parameterization is expected to remain part of global atmospheric models for several more decades until even the smallest-scale clouds become resolved on the computational grid (Schneider, Teixeira, et al., 2017). We choose to optimize a scheme based on the eddy-diffusivity/mass-flux (EDMF) approach since this scheme offers considerable potential for widespread adoption due to the option to unify PBL turbulence, moist shallow, and deep convection (e.g., Soares et al., 2004; Sušelj et al., 2013; Tan et al., 2018) and its extensibility to be scale aware (Sakradzija et al., 2016). We are particularly interested in the optimization of EDMF for the convective boundary layer (CPBL). This will serve as a relevant test bed for more complex future tasks such as optimizing global convection-permitting models with many more free parameters (e.g., due to microphysics) or training a traditional GCM through locally targeted large-eddy simulations (LES; Schneider, Lan, et al., 2017).

The total EDMF subgrid-scale flux is decomposed into two contributions from eddy-diffusivity and mass-flux terms. A theory for the relative partitioning between these two contributions does not exist. As a result, the magnitudes of the two terms depends on free parameters controlling each flux component. The optimization will thus also demonstrate which flux partitioning performs best for the CPBL. Some nonlocal convective flux is necessary to (a) establish the observed nonlocal counter-gradient flux in the top of the CPBL and to (b) admit condensation in plume-like convective clouds (see Siebesma et al., 2007). However, an optimal scale at which the continuous spectrum of boundary-layer eddies should be divided into small eddies and large plumes has never been theoretically established. Even pure mass-flux models without eddy-diffusivity flux have been proposed for the parameterization of the CPBL (Cheinet, 2003, 2004).

The goals of this paper are to

1. develop a computationally cheap framework to optimize physics parameterizations based on RBF surrogate models,
2. compare the cost and efficiency of RBF surrogate models and Neelin et al. (2010)'s quadratic polynomial model, and
3. determine the optimal EDMF parameter combination for the cloud-free and cumulus-topped CPBL.

Table 1
Summary of the Six Free Parameters in EDMF, Their Tested Range, and Traditionally Used Values

	a_{θ_v}	a_{q_t}	τ_l	$\hat{w}_{\min}; a$	τ_ϵ	b_ϵ
Description	Plume corr. coeff. for θ_v	Plume corr. coeff. for q_t	Eddy-mixing time scale	Min. velocity of plumes; fractional area	Entrainment time scale	Plume drag coefficient
Range	[0–4]	[0–4]	[0–2000] s	[0–2]; [50–2.3%]	[100–1000] s	[0–2]
Default value	0.58	0.32	400 s	1.28; 10%	400 s	0.4

Note. EDMF = eddy-diffusivity/mass-flux.

The EDMF parameterization and its free parameters are described in section 2. The optimization strategy based on RBF surrogate models is introduced in section 3. Section 4 presents the results from the optimization and from the comparison between the RBF surrogate models and the low-order polynomial model.

2. Eddy-Diffusivity/Mass-Flux Parameterization

Local eddy-diffusivity schemes without counter-gradient term are unable to model the stable stratification in the upper parts of the CPBL where nonlocal transport is considerable (Deardorff, 1966; Stull, 1984). Nonlocal mass-flux schemes, on the other hand, have traditionally been used to model moist deep convection (Arakawa, 1969). The combined EDMF approach was first introduced by Teixeira and Siebesma (2000) and Siebesma et al. (2007). It is formulated around the assumption that vertically coherent and strong organized plumes are represented by steady state plumes while the remaining turbulence is attributed to small diffusive eddies.

In a multiplume model with N_p represented classes of different rising plumes (see also Cheinet, 2003; Cheinet & Teixeira, 2003; Tan et al., 2018), the total EDMF flux of a scalar ϕ is partitioned via

$$\overline{w'\phi'}(z) = \sum_{i=1}^{N_p} M_i(z)[\phi_i^u(z) - \phi^e(z)] - (1 - \sum_{i=1}^{N_p} a_i)K(z)d_z\phi^e(z), \quad (1)$$

into a convective contribution from plumes and a diffusive contribution from eddies in the plumes' environment. Here M_i is the convective flux $M_i = a_i w_i^u$ of plume class i with fractional area a_i and vertical velocity w_i. The prime ($'$) represents the departure from the mean. Indices u and e indicate updraft and environment properties, respectively, and K is an eddy diffusion coefficient in the environment. The multiplume model is designed to represent either the full range of or parts of the continuum of weak to strong eddies. Six free parameters have been identified to be critical (see Table 1). These parameters dictate the contributions from the multiplume and the eddy-diffusivity part, respectively, and are explained in the next two sections.

2.1. Multiplume Model

The multiplume model is based on a discrete spectrum of nonprecipitating steady state entraining plumes. Its design follows closely the scheme proposed by Cheinet (2003). Each discrete plume represents a class of updrafts, which share the same initial thermodynamic and kinematic conditions. Through similarity theory the initial conditions of plumes at the surface are linked to surface fluxes and to assumed probability distributions for the vertical velocities. For $w > 0$ the latter is assumed to be a Gaussian with a standard deviation σ_w, that is, half of the grid box has rising motion near the surface. A selected range of the distribution between the minimum and maximum vertical velocities denoted by $w_{\min}$ and $w_{\max}$, respectively, is then discretized using $N_p = 40$ plume classes. In a typical EDMF setup, $w_{\min}$ is chosen to be larger than zero and the weakest eddies will be allocated to the the eddy diffusivity process. A pure mass flux model following Cheinet (2003) would dictate $w_{\min} = 0$ and cover all parts of the domain with rising motion, i.e. $\sum_i a_i = 0.5$. The lower bound is set by the scaled velocity $\hat{w}_{\min} = w_{\min}/\sigma_w$, with the hat indicating the scaling operation. The upper bound is fixed at a sufficiently large velocity $\hat{w}_{\max} = 3.5$. Therefore, $\hat{w}_{\min}$ sets the total surface plume area fraction a.

To obtain thermodynamic properties for each plume class in the surface layer, we follow Cheinet (2003) and Lenschow et al. (1980) and diagnose the standard deviations of w, θ_v, and q_t from surface layer similarity

relations. The plume properties (indicated by index u) in the surface layer are parameterized by

$$q_{ti}^{u} = \overline{q_t} + a_{q_t} w_i^u \sigma_{q_t} / \sigma_w \tag{2}$$

$$\theta_{vi}^{u} = \overline{\theta_v} + a_{\theta_v} w_i^u \sigma_{\theta_v} / \sigma_w \tag{3}$$

with the bar indicating the grid mean. For example, Cheinet (2003) uses correlation coefficients $a_{q_t} = 0.32$ and $a_{\theta_v} = 0.58$. Plumes then rise until a zero vertical velocity is reached. The velocity profiles are obtained following Simpson and Wiggert (1969), as

$$1/2 \quad d_z (w_i^u)^2 = a_b B_i^u - b_\epsilon \epsilon_i^u (w_i^u)^2, \tag{4}$$

with the right-hand side terms describing effective buoyancy acceleration and drag. Coefficients a_b and b_ϵ are the effective buoyancy and the entrainment coefficient, respectively, and ϵ_i^u is the fractional entrainment rate of the updraft. Using LES data de Roode et al. (2012) pointed out a linear correlation between the two rhs terms and suggested the relation $a_b = \gamma b_\epsilon + \eta$. In line with their study we use $\gamma = 0.4$ and $\eta = 0.3$ to derive a_b from the remaining drag coefficient b_ϵ. Fractional entrainment is described by the frequently used parameterization introduced by Neggers et al. (2002), as

$$\epsilon_i^u = (\tau_\epsilon w_i^u)^{-1}, \tag{5}$$

with τ_ϵ the entrainment time scale. At each height a saturation adjustment scheme is used to infer the condensed water content q_{li}^u.

2.2. Eddy-Diffusivity Closure

A 1.5-order closure is applied to obtain the eddy-viscosities such that $K = 0.5 l e^{1/2}$ with e the turbulent kinetic energy in the environment of plumes and l is the mixing length scale. The formulation of the prognostic equation for e follows Teixeira and Cheinet (2004) and is specifically designed to model the evolution of the CPBL. Following Sušelj et al. (2013) the mixing length l is defined as

$$l = l_{23} + (\kappa z - l_{23}) e^{-z/100 \quad \mathrm{m}} \tag{6}$$

$$l_{23}^{-1} = l_2^{-1} + l_3^{-1} \tag{7}$$

$$l_2 = \tau_l \sqrt{e} \tag{8}$$

$$l_3 = \begin{cases} \max\left(\Delta z/2, 0.7\sqrt{e/N_b^2}\right) & \text{if} \quad N_b^2 > 0 \\ \infty & \text{if} \quad N_b^2 \leq 0 \end{cases} \tag{9}$$

with τ_l an eddy-mixing time scale for the mixed layer, N_b the Brunt-Väisälä frequency, and $\kappa = 0.4$ the von Kármán constant.

In summary, the five parameters a_{θ_v}, a_{q_t}, either $\hat{w}_{min}$ or a, τ_ϵ, and b_ϵ control the flux of the multiplume model. To first order the system of governing parameterizations described above is such that if entrainment becomes larger or the fractional area becomes smaller, then the mass-flux component will become smaller. The eddy-diffusivity flux is controlled through the eddy-mixing time scale τ_l, with a larger τ_l yielding more vigorous mixing. Table 1 shows the parameter ranges tested in this study and the default values typically used in traditional EDMF specifications. In section 4 we contrast the behavior of the EDMF scheme operating under these default versus our optimized settings.

3. Optimization Strategy

3.1. Definition of the Objective Function

Two types of representative convective boundary layers are considered: a cloud-free convective boundary layer (CPBL; Nieuwstadt et al., 1993; Teixeira & Cheinet, 2004) and a specific case of a steady state maritime boundary layer topped by trade-wind cumuli observed during the Barbados Oceanographic and Meteorological Experiment (BOMEX; Siebesma et al., 2003). LESs of both cases are performed using the System for Atmospheric Modeling (SAM; Khairoutdinov & Randall, 2003) to serve as benchmarks for the evaluation of the single-column model (SCM). Domains discretized using 512 × 512 × 150 (256 × 256 × 200) points

in the two lateral and one vertical dimension, respectively, with grid point volumes of $25 \times 25 \times 20$ m^3 ($25 \times 25 \times 25$ m^3) are used for the BOMEX (CPBL) simulation.

The objective function measures the distance between our SCM simulations and these LES benchmarks and is based on vertical profiles of grid-mean values and their fluxes and on vertical profiles of the thermodynamic properties of the fastest plumes. Minimizing the objective function therefore ensures the maximization of both the large-scale and plume-scale skill of the SCM. Both grid-mean values and fluxes are included in the formulation, despite being linked. The grid-mean values are critical in terms of GCM performance since other parameterizations (e.g., radiative transfer) are only affected by them, but not by higher-order terms. The fluxes are included here since differences due to the parameter setup first materialize in different fluxes and only gradually affect the grid-mean values, especially in the steady state case. All considered quantities have been averaged in time. For the transient CPBL, two 20-min time windows centered around hour three and five are utilized. For the BOMEX case quantities have been averaged over the third simulation hour. We interpolate the averaged LES data to the equidistantly-spaced grid ($\Delta z = 20$ m) of the SCM. The root-mean-square error (or distance) between profiles of a SCM quantity ψ and the corresponding LES benchmark ψ_b is then defined for a position $\mathbf{x}$ in the parameter space, as

$$r^{\psi}(\mathbf{x}) = \left[\frac{1}{N_z} \sum_{k=1}^{N_z} (\psi(\mathbf{x}, k) - \psi_b(k))^2 \right]^{1/2}, \tag{10}$$

where N_z is the number of vertical levels in between the surface and the maximum evaluation height of 2,900 m. This distance is evaluated for the liquid water potential temperature θ_l, the total water mixing ratio q_t, the heat flux $F_{\theta_l} = c_p \rho \overline{w' \theta_l'}$, the latent heat flux $F_{qt} = L_v \rho \overline{w' q_t'}$, and the mean water and buoyancy excess (q_t' and θ_v') of the fastest plumes covering 1% of the horizontal domain. The latter have been diagnosed from the LES simulations based on grid points with a vertical velocity larger than the 99th percentile. In case of BOMEX, r^{ψ} is evaluated also for cloud liquid mixing ratio q_c. In case of the CPBL, we additionally evaluate the bulk static stability of the upper CPBL (between 700 m and the inversion) using a linear regression of θ. For this specific parameter we define r^{ψ} based on the absolute difference to the static stability of the benchmark. Thus, for both cases seven performance metrics are included. These are scaled by the respective average values $\overline{r^{\psi}}$ as determined from a grid sampling of the six-dimensional parameter space. The grid sampling divides each parameter range into equidistant intervals and evaluates the objective function at the resulting 9^6 grid points by running the SCM. Then the average of the seven scaled distances is taken as a final distance. In case of BOMEX, this final distance defines the objective function. In case of the CPBL, we average the 3- and 5-hr final distances to arrive at the objective function. The overall objective function is taken as the average of the case-specific objective functions.

3.2. Optimization Using RBF Surrogate Models

The optimization of the free parameters in EDMF is complicated by the computational expense of running the SCM repeatedly, the absence of analytic partial derivatives of its response surfaces, and the multimodality and "black box" characteristics of its emergent behavior. In response to these difficulties, we employ an adaptive surrogate model optimization approach. In fact, we compare the performance of two recent surrogate optimization methods, namely, the Metric Stochastic Response Surface (StochRBF hereafter) method (Regis & Shoemaker, 2007) and the DYnamic COordinate search using Response Surface models (DYCORS hereafter) method (Regis & Shoemaker, 2013). Both methods start with an initial experimental design to decide at what points the objective function gets evaluated first. An RBF model is computed based on this initial data. Generally speaking, an RBF model is an interpolating model that makes predictions for an unsampled parameter vector based on its distance to already evaluated parameter vectors. We use a cubic RBF model, that is, we use the cubed distance for interpolation (see Appendix A for the general definition of the RBF model). A large set of candidate points is generated by (1) perturbing the best parameter vector that has been found so far and (2) by randomly generating points from the parameter space. The candidate points are ranked based on (a) their distance to the set of already evaluated points and (b) their function values as predicted by the RBF model. A weighted combination of these two criteria determines the best candidate point at which the next expensive function evaluation is done. The RBF model is then updated with the new data and the method iterates through candidate point generation, computation of the criteria, evaluation of the expensive simulation at the newly selected point, and the update of the RBF model until

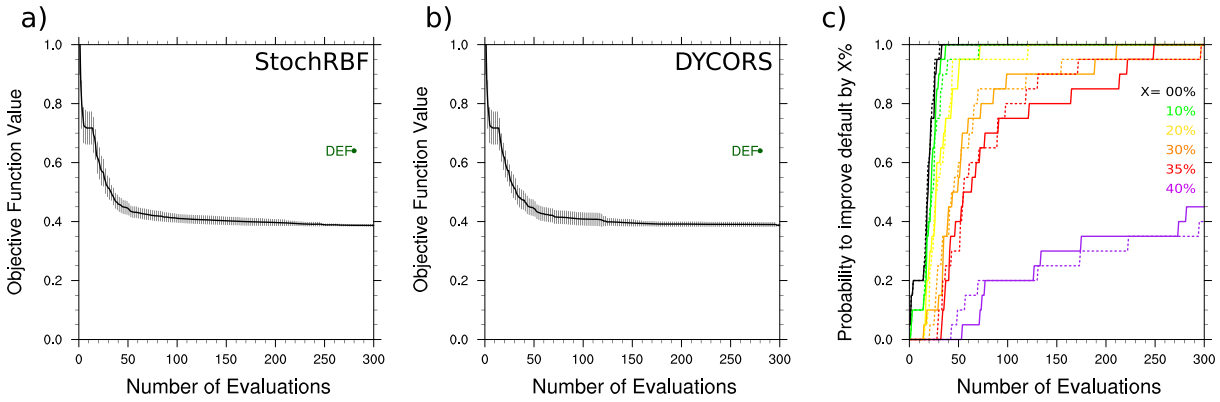

Figure 1. The average (lines) and standard deviation (vertical bars) of the single-best objective function values of 20 trials are shown as a function of the number of SCM evaluations performed for (a) StochRBF and (b) DYCORS. The objective function value obtained with the default parameter combination (DEF) is shown as reference (green dot in (a) and (b)). (c) The fraction of StochRBF (solid) and DYCORS (dashed) trials that detected a parameter combination that reduces the error of the default setting by 0%, 10%, 20%, 30%, 35%, and 40% as function of the number of evaluations.

a specified cutoff criterion is reached. Here iterations stop after 300 evaluations. We refer the reader to the above cited works for the implementation details.

The main difference between StochRBF and DYCORS lies in the approach to perturbing the values of the best parameter vector found so far. While StochRBF perturbs all parameters, DYCORS only perturbs a subset of the parameters. In particular, DYCORS perturbs each parameter with a probability that decreases as the number of function evaluations increases and approaches the maximum number of evaluations (here, 300). Thus, the DYCORS search can be interpreted as one that becomes increasingly more local as fewer and fewer parameters are perturbed. Second, when random perturbations are added to the best point found so far, it is possible that the resulting candidate point falls outside the bounds of the parameter space. StochRBF simply projects the point onto the corresponding boundary that has been exceeded, whereas DYCORS reflects the point to inside the parameter space. Thus, StochRBF is more likely to generate and sample points that lie on the boundary.

Note that both algorithms are inherently stochastic because the candidate points are generated by adding random perturbations to the parameters (or a subset of them) of the best point found so far. Thus, in order to obtain statistics about the average performance of the algorithms, for example, we run each algorithm 20 times starting from a different initial experimental design. This allows us to investigate the robustness of the results to this randomness. The goal of using the two algorithms is to examine if one outperforms the other for our specific application, and thus to derive recommendations regarding whether or not an increasing focus on the local search as done in DYCORS leads to better solutions and should therefore be preferred for the type of optimization applications that we consider here.

Table 2
Overview of Parameter Combinations and Their Performances

Name	a_{θ_v}	a_{q_t}	τ_l (s)	$\hat{w}_{\min}; a$ (%)	τ_ϵ (s)	b_ϵ	*OF*
DEF	0.58	0.32	400.00	1.28; 10	400.00	0.40	0.640
OPTS	0.52	0.74	729.00	0.88; 19	430.00	0.49	0.378
OPTD	0.49	0.73	719.00	0.58; 28	438.00	0.49	0.376
OPTSτ_l0	0.23	0.38	0.00	0.72; 24	457.00	0.56	0.378
MM_G	0.50	0.50	750.00	0.50; 31	550.00	1.00	0.532
MM_S	0.66	0.42	825.00	0.77; 22	493.00	0.99	0.529
OPTG	0.50	0.50	750.00	0.75; 23	438.00	0.50	0.380

Note. Shown are parameter values and objective function values *OF* for configurations discussed in the text.

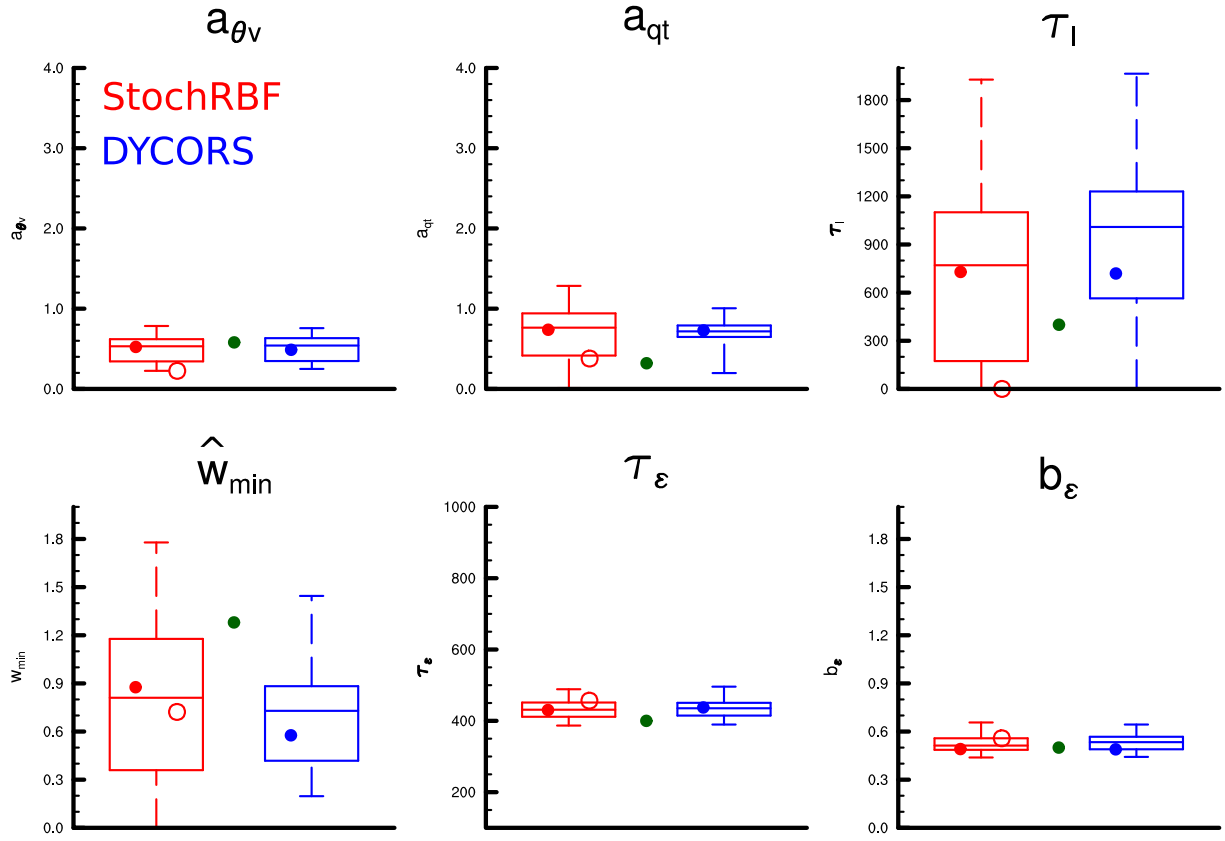

Figure 2. Box plots showing parameter distributions (min, 25%, median, 75%, max) for the single-best configurations of 20 trials from (red) StochRBF and (blue) DYCORS. The green dots show the default parameter values. The filled red and blue dots indicate the OPTS and OPTD configurations, respectively, and red circles indicate the OPTSτ_l0 configuration, the best configuration found with $\tau_l = 0$ s.

3.3. Neelin's Quadratic Polynomial Model

The basic idea proposed by Neelin et al. (2010) is to design a second-order polynomial, as

$$\tilde{\phi}(\tilde{\mathbf{x}}) = \phi_{\text{ref}} + \tilde{\mathbf{x}}^T\mathbf{a} + \tilde{\mathbf{x}}^T\mathbf{B}\tilde{\mathbf{x}}, \tag{11}$$

to describe the behavior of ϕ across the parameter space. The linear and quadratic term describe the deviation from a known reference value ϕ_{ref} as a function of the parameter perturbation vector $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}_{\text{ref}}$. The reference point $\mathbf{x}_{\text{ref}}$ is defined to be the center of the parameter domain. Here vector $\mathbf{a}$ contains the linear coefficients for each parameter, and matrix $\mathbf{B}$ contains the quadratic and interaction terms in the diagonal and off-diagonal elements, respectively. In Neelin et al. (2010), ϕ represents a climatological field such as surface precipitation. Here we utilize this polynomial to model the profiles of our quantities of interest. For example, ϕ represents the total water mixing ratio at a specific model level or the bulk stability of the CPBL.

To obtain the coefficients $\mathbf{a}$ and $\mathbf{B}$ a few evaluations of the SCM are required at so-called saturation points in the six-dimensional parameter space. As further described in Appendix B, the number of evaluations needed is $2d^2$, which equals 72 for $d = 6$. The objective function for a specific parameter combination can then be evaluated by following the same procedure as outlined above for the SCM.

4. Results

4.1. Performance of the RBF Surrogate Models

Figures 1a and 1b show the average and standard deviation of the single-best objective function value (also referred to as *error* from now on) of the 20 trials as a function of the number of SCM evaluations performed. For each trial the single-best error is the lowest objective function value detected over all evaluations conducted so far. For both algorithms the average error decreases rapidly over the first ~90 evaluations. The average then asymptotes to a value of about 0.38. After 300 evaluations the two algorithms reach a single-best parameter combination denoted by OPTS (for StochRBF) and OPTD (for DYCORS) with a remaining error

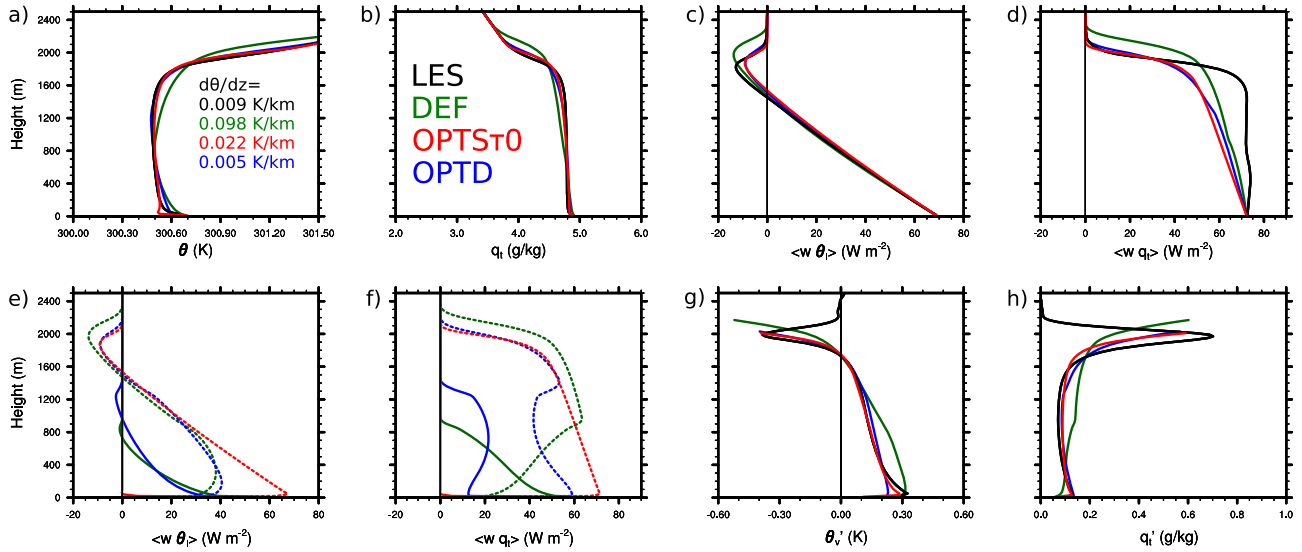

Figure 3. Results obtained for the CPBL simulation using the default parameter combination DEF, the two optimized configurations OPTSτ_l0 and OPTD, and the LES. Shown are (a) θ, (b) q_t, (c) F_{θ_l}, (d) F_{q_t}, (e) eddy-diffusivity (solid) and mass-flux (dashed) parts of F_{θ_l}, (f) eddy-diffusivity (solid) and mass-flux (dashed) parts of F_{q_t}, (g) perturbation virtual potential temperature θ_v', and (h) perturbation water mixing ration q_t' of the fastest plumes covering a fractional area of 1%. Panel (a) lists the bulk stability of the upper CPBL for the various configurations.

of 0.378 and 0.376, which corresponds to an error reduction relative to the default setting (DEF) by 41% (see Table 2).

The probability of a single trial to reduce the default error by a certain percentage is shown in Figure 1c. Generally, the chance to improve by a certain percentage increases mostly during the first tens of evaluations and less during subsequent evaluations. The probability to perform as well or better than the default configuration reaches 100% after only 32 and 30 evaluations with the StochRBF and DYCORS algorithm, respectively. As another example, for StochRBF the probability to reduce the error by 30% increases from 0% after 13 evaluations to 90% after 98 evaluations.

Parameter distributions based on the 20 single-best configurations (including OPTD and OPTS) are shown in Figure 2 for both algorithms. The two algorithms find very similar medians and interquartile ranges. The four parameters a_{θ_v}, a_{q_t}, τ_ϵ, and b_ϵ have narrow interquartile ranges. Since all of these configurations result in a low error, this points at a significant sensitivity to these parameters. In contrast, τ_l and $\hat{w}_{\min}$ have wide interquartile ranges indicating that the error might be less sensitive to the eddy-mixing time scale and the plume area fraction. The alternative explanation could be that these distributions are multimodal and therefore have wider ranges. We will explore this idea further later in the paper. Many configurations exist with plume fractional areas larger than in the default setup (i.e., a smaller $\hat{w}_{\min}$). Note that the median of the τ_l distribution is at least two times as large as the default value of 400 s. The OPTS and OPTD configurations are shown by the red and blue markers, respectively. These single-best configurations detected by the algorithms are very similar (see also Table 2) and parameter values fall into the interquartile range of the respective parameter distribution. The main differences to the default setting are larger plume fractions, larger eddy-mixing time scales, and larger surface moisture perturbations of plumes.

We further find that for StochRBF three out of the 20 trials find a single-best configuration with $\tau_l = 0$, that is, a pure plume model without eddy-diffusivity mixing. DYCORS on the other hand never evaluates the SCM at a point with $\tau_l = 0$ because it reflects points falling outside the parameter space back inside rather than placing them on the boundary. StochRBF appears more suitable for detecting such outlier configurations. In fact, 38 and 1 out of the best 60 and best 6, respectively, detected configurations in StochRBF have $\tau_l < 100$ s while only 3 of the 60 best and none of the 6 best configurations in DYCORS have $\tau_l < 100$ s. Note that of those 38 with $\tau_l < 100$ s, 36 have a $\tau_l = 0$ s. This is an important finding as more than half of the best 1% of all 6,000 sampled configurations in StochRBF have $\tau_l = 0$ s and none in DYCORS. This difference is a consequence of the different treatments of the boundaries and the fact that StochRBF does a thorough global search, while DYCORS spends more time on the local search. The best configuration with $\tau_l = 0$,

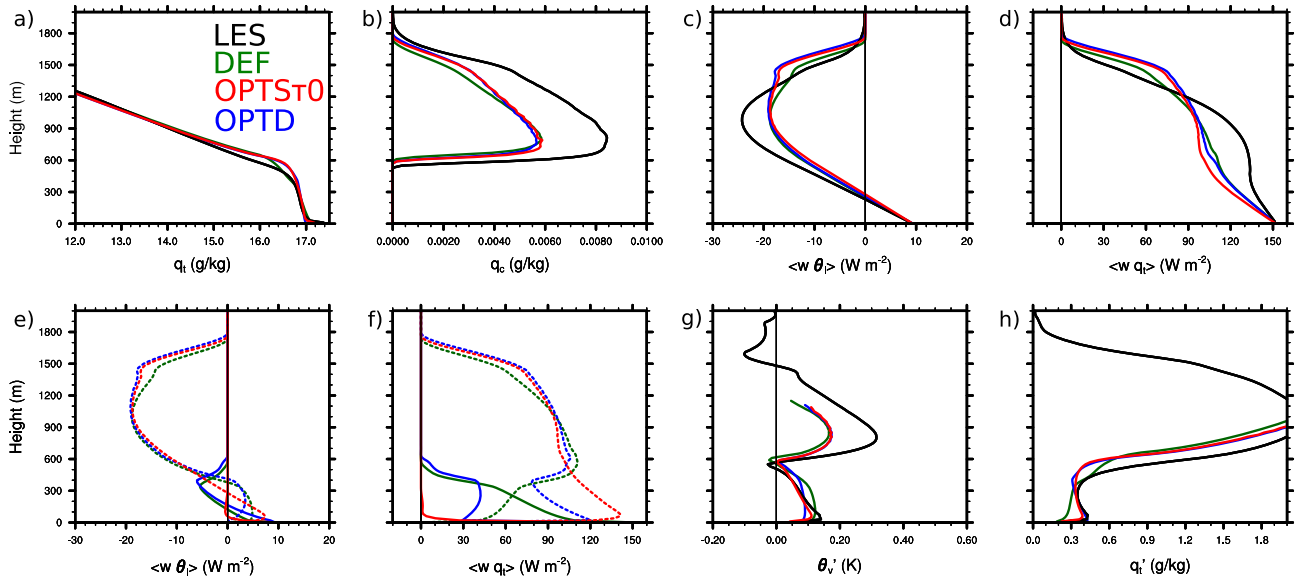

Figure 4. Results obtained for the BOMEX simulation using the default parameter combination DEF, the two optimized configurations OPTSτ_l0 and OPTD, and the LES. Shown are (a) q_t, (b) q_c, (c) F_{θ_l}, (d) F_{q_t}, (e) eddy-diffusivity (solid) and mass-flux (dashed) parts of F_{θ_l}, (f) eddy-diffusivity (solid) and mass-flux (dashed) parts of F_{q_t}, (g) perturbation virtual potential temperature θ_v', and (h) perturbation water mixing ration q_t' of the fastest plumes covering a fractional area of 1%.

OPTSτ_l0, is also listed in Table 2 and indicated by the red circles in Figure 2. To check if we could find even better configurations for the pure plume model we reduce the number of free parameters to five and fix $\tau_l = 0$ s. Indeed, if we repeat the optimization this way then both algorithms yield single-best configurations with an error of 0.374, slightly smaller than those of OPTS and OPTD. We also tested a version of EDMF in which we set the mixing length l to zero ($\tau_l = 0$ still allows for minuscule mixing in the surface layer) and confirmed our findings regarding the high quality of a plume-only parameterization. OPTS and OPTD are based on very similar setups, but the equally-well performing OPTSτ_l0 configuration is conceptually completely different. As seen in Table 2, OPTS and OPTD apply similar parameter configurations, result in the same considerable improvement over the default setup, and yield the same profiles for the CPBL and BOMEX cases (therefore, only OPTD is shown and discussed below).

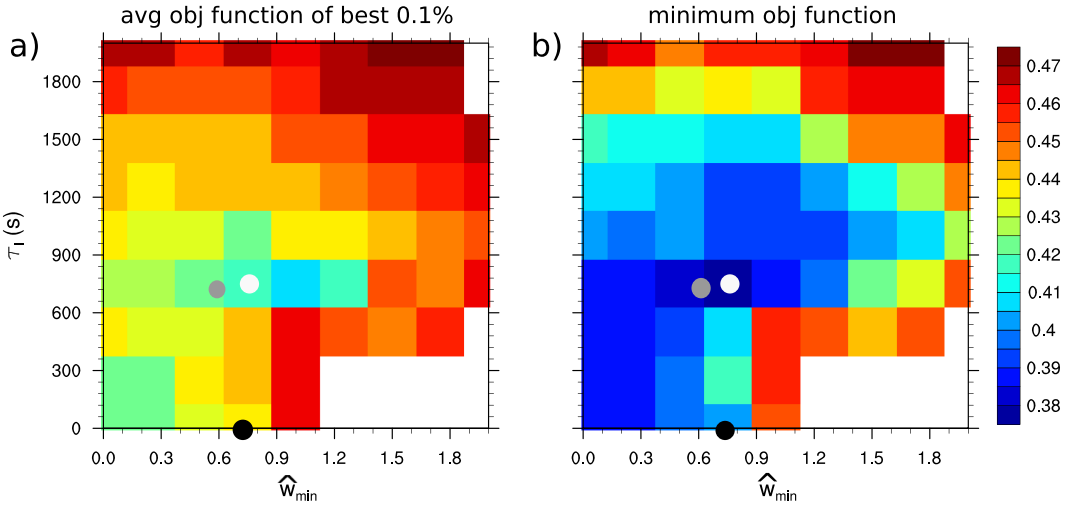

Figure 5. (a) The average objective function value in the $\hat{w}_{min} - \tau_l$ parameter space computed for the best 0.1% of the grid-sampled configurations. (b) The minimum values detected in this parameter plane. In both panels white areas indicate that no configurations belong to the best 0.1%. The bullets show (white) OPTG detected by grid sampling, (gray) OPTD detected by DYCORS, and (black) OPTSτ_l0 detected by StochRBF.

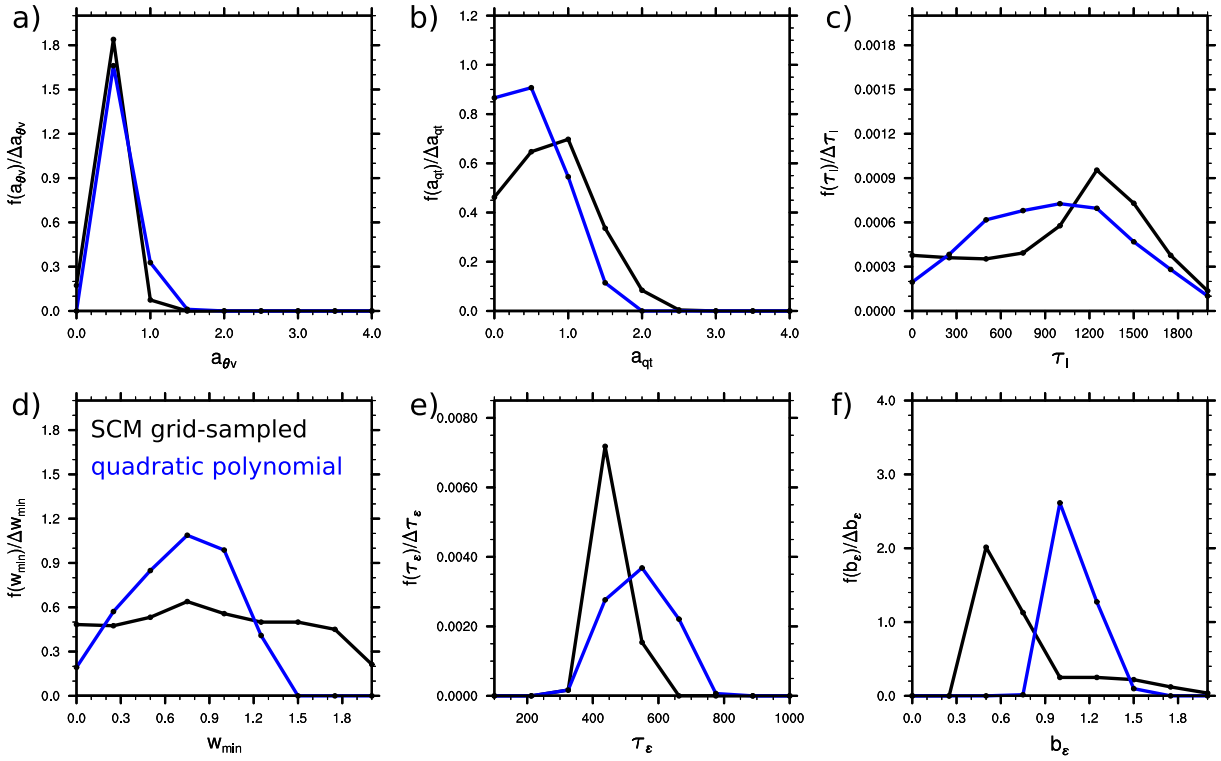

Figure 6. Frequency distributions for (black) the best 0.1% combinations of the 9^6 single column model results and (blue) the best 0.1% of the 9^6 grid-sampled quadratic polynomial model results.

Figure 3 shows the profiles used when constructing the objective function for the CPBL case. Only the time averages centered around hour three are shown. With the default setup the eddy-diffusivity flux is active in the lowest 1,000 m only, while above that only the mass-flux component is active (Figures 3e and 3f). The described transition causes an irregularity at around 1,000 m that affects all profiles shown in Figure 3 and is particularly well visible in Figures 3a, 3g, and 3h. While this particular partitioning of the flux does allow for a stably stratified upper CPBL, the bulk stability is too large in this setup (Figure 3a). Figure 4 shows the profiles for BOMEX. In this case, the default setup leads to too small vertical fluxes (Figures 4c and 4d), to a too small cloud liquid water content (Figure 4b), and slightly off profiles of the plume properties (Figures 4g and 4h).

The OPTSτ_l0 configuration is just as skillful as OPTS and OPTD but, interestingly, based on a completely different flux partitioning (Figures 3e and 3f and 4e and 4f). Basically all flux is carried by the mass-flux component. The increase in the convective flux appears to be mostly made possible by a larger entrainment time scale. The OPTD configuration increases the eddy-diffusivity time scale such that the layer with active eddy-diffusivity flux becomes slightly deeper than in the default setup (Figures 3e and 3f). OPTD also increases the convective moisture flux near the surface by using a larger surface moisture perturbation for plumes.

While both OPTSτ_l0 and OPTD improve each of the two cases individually, the improvements are much more significant for the CPBL case. Both optimized settings, OPTD and OPTSτ_l0, improve the simulation of the CPBL by capturing the stability of the upper CPBL more accurately (Figure 3a) and by providing better thermodynamic profiles of the fastest plumes (Figures 3g and 3h). The profiles of the two solutions are mostly indistinguishable but are based on a completely different flux partitioning.

We now want to address the question whether the larger variances of τ_l and $\hat{w}_{min}$ (seen in Figure 2) are a result of a multimodal behavior of the objective function. To this end, we analyze the best configurations in the $\hat{w}_{min} - \tau_l$ parameter plane. Figure 5a shows the average objective function value for the best 0.1% of all grid-sampled configurations. Figure 5b shows the minimum value found for each combination of the two parameters. In both plots excellent solutions can be found across a wide range of $\hat{w}_{min}$ and τ_l values. The best

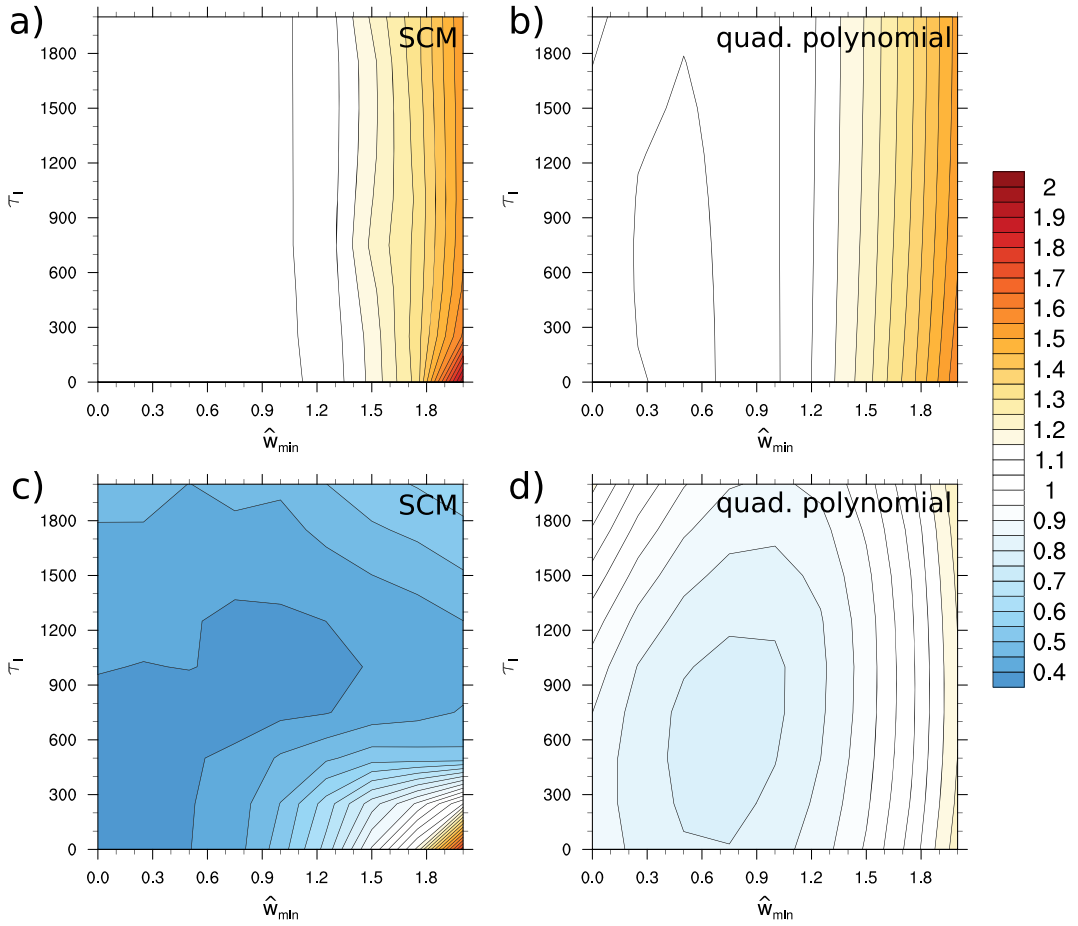

Figure 7. The objective function value shown for the plane spanned by the scaled plume vertical velocity $\hat{w}_{\ \min}$ (which sets the plume area fraction) and the eddy-mixing time scale τ_l for (a and c) the single column model and (b,d) the quadratic polynomial model. The coordinates of the remaining four parameters are those of (a and b) the reference point $\mathbf{x}_{\text{ref}}$ and those of (c and d) the OPTG solution.

among the top 0.1% configurations are found in the bottom-left quadrant and the center of this parameter plane. There is no strong evidence for the existence of isolated regions of very small errors. Rather, a wide spectrum of combinations is found to lead to low errors. Let us define the best configuration detected by the grid sampling as OPTG. The OPTG and OPTD configurations are indicated by white and gray bullets, respectively. With plume fractions around 20–30% and a τ_l of about 750 s they are located closer toward the center of the plain. Very low objective function values can be found for many parameter pairs between those and the bottom-left corner (plume-only configurations with 50% plume area fraction). StochRBF samples the low errors in the bottom-left corner of the plane but detects an even better sample, OPTSτ_l0 (black bullet), that is not captured by the grid sampling. This indicates that more detailed distributions would be obtained by increasing the density of the grid sampling and that a 50% plume fraction is not necessary to achieve excellent skill of the plume-only configuration, at least for this specific multiplume model.

4.2. Comparison Against the Quadratic Polynomial Model

The quadratic polynomial model has been used to predict the objective function values on a six-dimensional grid using nine points in each parameter direction (i.e., 9^6 points) equally spaced between the minimum and maximum of each parameter range. To test the influence of this sampling strategy, we also stochastically selected 10^6 points at which we evaluated the objective function with the quadratic polynomial model. The best parameter combinations obtained with the quadratic polynomial model on the grid and stochastically (MM_G and MM_S) result in errors of 0.532 and 0.529 (Table 2). This corresponds to an error reduction of only 17%. The grid sampling of the SCM using the described 9^6 points yields a best configuration, OPTG,

that reduces the error of the default setup by 41% (see Table 2) indicating that the poor performance of the quadratic polynomial cannot be blamed on the grid-sampling strategy.

Remember that the RBF surrogate model optimizations used more evaluations than the 72 used to fit the quadratic polynomial. Still, after 72 evaluations 20 and 19 out of StochRBF's and DYCORS' 20 trials found a single-best solution with an error smaller than the error of the best estimate from the quadratic polynomial. The average single-best error after 72 evaluations is 0.422 and 0.416 for StochRBF and DYCORS, respectively. This still amounts to an average error reduction of 35%. In other words, for the same cost the surrogate models still on average reduce the error of the quadratic polynomial by 21%. Further, StochRBF and DYCORS reach an average single-best error equal to or smaller than the error of the quadratic polynomial after only 27 evaluations. The same accuracy is therefore accomplished by the RBF surrogate models with only ∼40% of the quadratic polynomial's cost.

To better understand why the quadratic polynomial model performs much worse, we construct frequency distributions for all six parameters. We are interested here in the combinations that perform best. Figure 6 thus shows distributions based only on the best 0.1% of all configurations from both the 9^6 grid-sampled SCM simulations and the 9^6 grid-sampled quadratic polynomial model evaluations. The results from the stochastically-sampled quadratic polynomial model are very similar to the grid-sampled ones and therefore not shown.

The quadratic polynomial model suffers from two major deficiencies. First, the frequency of small and large $\hat{w}_{\min}$ (i.e., area fraction) is underestimated (Figure 6d). Instead of having a pronounced peak in the center of the range, the SCM shows a rather homogeneous distribution indicating that good solutions exist across all convective area fractions. Second, the distribution for b_ϵ peaks at a too large value (Figure 6f). Further, the subtle increase in frequencies toward the minimum value of τ_l is not captured by the polynomial model which instead models a rapid drop in frequencies.

Apparently, the second-order truncation of the quadratic polynomial model is insufficient to model the behavior of the CPBL and BOMEX cases across the parameter space. This is further illustrated in Figure 7 which shows the error as a function of two parameters. As an example, we again show here the behavior in the $\hat{w}_{\min} - \tau_l$ plane, but insufficiencies could be demonstrated for any other parameter pair. By design the quadratic polynomial model performs best for planes through the reference point since higher-order terms involving other parameters are zero. This indeed leads to a reasonable agreement with the SCM at the reference point (Figures 7a and 7b). However, the error is no longer reproduced accurately by the quadratic polynomial model for planes that run through points away from the reference point. Figures 7c and 7d illustrate this for the same parameter pair but analyzed at the global minimum, OPTG, detected by grid sampling of the SCM. The quadratic polynomial model does not pick up the low errors found in the center and bottom-left corner of this plane. This indicates that terms describing correlations between three parameters or more cannot be neglected if the true boundary-layer state is to be replicated in parameter space. The RBF surrogate models are not subject to the same limitation since they are by construction more flexible and unlike a polynomial model, they do not by design impose a certain "shape" (e.g., quadratic) on the objective function.

5. Conclusions

To our knowledge this is the first time that RBF surrogate models have been applied to the optimization of a physical parameterization of an atmospheric model. Two algorithms based on RBFs, the Metric Stochastic Response Surface (StochRBF) method (Regis & Shoemaker, 2007) and the DYnamic COordinate search using Response Surface models (DYCORS) method (Regis & Shoemaker, 2013), have been utilized to optimize an EDMF boundary layer and shallow cumulus scheme for the convective boundary layer. The results are very encouraging since both algorithms lead to a significant error reduction of 41% relative to the default model setup. In contrast, a frequently used quadratic polynomial model (see Neelin et al., 2010) leads to a reduction of only 17%. While the quadratic polynomial model was found sufficient for the optimization of global climate and regional climate models, it was here shown that for the optimization of a boundary-layer parameterization the higher-order terms and/or a better sampling strategy are critical to the functional form of the error. In contrast to polynomial models, RBFs do not impose a certain functional form (such as quadratic) on the objective function, and they are therefore more flexible when modeling nonlinearities. The major disadvantages of using polynomial models of a certain degree is that (1) they are highly unreliable for

making predictions globally if the true underlying function does not look like a polynomial of the chosen degree, and (2) the number of samples that are needed to fit a polynomial of higher order quickly becomes infeasible when function evaluations are expensive. Thus, even if fitting a cubic polynomial to our data, we do not expect it to outperform the RBF surrogate model because much more than 72 points are needed to fit a cubic polynomial, and as shown above, the RBF models on average converge during the first ∼100 evaluations. Another major difference between the methods is the adaptivity of the RBF surrogate models. The quality of the latter is iteratively improved as new information is obtained. The method proposed by Neelin et al. (2010), on the other hand, is a one-shot approach in which the quadratic model is built once and then never updated or improved.

Both RBF algorithms are inherently stochastic due to the randomness applied during the selection of evaluation points. It is found that the probability to detect configurations with a certain error reduction rapidly increases over the first tens of evaluations. Thus, on average the skill of the quadratic polynomial model is reached for only 40% of its cost. It is hoped that the approach can therefore be applied also to problems with a larger number of tunable parameters such as, for example, encountered in a cloud-resolving model with a full physics package. The stochasticity of these algorithms could even be exploited to represent the model uncertainty of physical parameterizations (e.g., Weisheimer et al., 2014) or to account for the natural variability of small-scale processes (e.g., Dawson & Palmer, 2015).

By design DYCORS reflects out-of-bounds evaluation points back into the parameter domain while StochRBF places them right on the boundary that was overshot. Further, StochRBF performs a more thorough global search in the parameter space while DYCORS focuses on a more local search. As a consequence, StochRBF is able to detect excellent solutions with a zero eddy-mixing time scale, while DYCORS is unable to discover any solutions at the boundary of the parameter space. More than half of the best 1% of all sampled configurations in StochRBF have $\tau_l = 0$ s (i.e., the value of the lower boundary) and none in DYCORS. Both algorithms pick up another optimal solution with $\tau_l \sim 700$ s and a fractional area of about 20-30%.

This suggests that the default EDMF can be improved significantly by either turning to a pure plume model or by specifying larger fractional areas and simultaneously increasing the eddy-mixing time scale. Of course, both versions require an EDMF formulation that does not apply the assumption of small plume fractional areas. Interestingly, high-density grid sampling also revealed that excellent configurations can be found for a wide range of combinations of the eddy-mixing time scale and the convective fractional area.

While conceptually different in terms of flux partitioning, the two optimal configurations yield a nearly indistinguishable boundary layer behavior. It is unclear at this point why these two models deliver the same behavior, and we hope future research will address this question. Note that multiplume schemes without eddy-diffusivity term have had great success with the modeling of convective boundary layers (e.g., Cheinet, 2003, 2004) and even successful implementations of EDMF into operational global models are found to have negligible eddy-diffusivity fluxes (Sušelj et al., 2014). Also note that our conclusions are based only on two cases with convective boundary layers. An extension to other cases would be needed to provide general guidelines for the tuning of EDMF for, for example, operational applications. Further, we only studied the convective boundary layer without shear. It seems plausible, however, that shear-driven turbulence can be added successfully to a pure plume model and be parameterized with an eddy-diffusivity term.

Appendix A: RBFs

The algorithms we employ in this work use RBFs (Powell, 1992) to approximate the expensive objective function. Generally, one could use other types of surrogate models such as kriging (Matheron, 1963), polynomial regression models (Myers & Montgomery, 1995), or ensembles of surrogate models(Müller & Piché, 2011; Müller & Shoemaker, 2014). RBFs have the property that they are interpolating, which is convenient in our case because the objective function is deterministic and an interpolating model will repredict the true function value at an already evaluated point. Moreover, RBFs scale relatively well with the problem dimension, and they have been shown to perform well in previous work (Müller, 2017; Müller & Woodbury, 2017; Regis & Shoemaker, 2007, 2013). The RBF interpolant is defined as

$$s(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \phi(\|\mathbf{x} - \mathbf{x}_i\|_2) + \rho(\mathbf{x}) \tag{A1}$$

where $s(\mathbf{x})$ denotes the prediction of the RBF model at the point $\mathbf{x}$, λ_i, $i = 1 \dots n$, are that RBF model parameters, $\phi(\cdot)$ denotes the RBF (here the cubic $\phi(r) = r^3$), $\rho(\mathbf{x})$ is a polynomial tail whose order depends on the chosen RBF type (for the cubic RBF, we need at least a linear tail $\rho(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T\mathbf{x}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^T$), and $\mathbf{x}_i, i = 1 \dots n$, are the points at which we have already evaluated the expensive function $f(\mathbf{x})$. $\| \cdot \|_2$ denotes the Euclidean norm. The parameters of the RBF model are determined by solving a linear system of equations. We have to evaluate initially at least $d + 1$ affinely independent points in order to solve the linear system, with $d = 6$ the number of unknown parameters.

Appendix B: Construction of Neelin's Quadratic Polynomial Model

In a first step, the two end points of each parameter range and the reference point will provide the diagonal elements of $\mathbf{B}$ and elements of vector $\mathbf{a}$. That amounts to a total of $2d$ additional points. In a second step, at least one corner point of a pairwise plane is needed to solve for the corresponding off-diagonal element of $\mathbf{B}$. Since $b_{ij} = b_{ji}$, this amounts to at least another $(d^2 - d)/2$ points. Same as in Neelin et al. (2010) and Bellprat et al. (2012), the off-diagonal elements are computed here using a linear regression based on all four corner points. Therefore, the total amount of points needed is $2d + 4(d^2 - d)/2 = 2d^2$, which is equal to 72 if $d = 6$. We apply this procedure to obtain $\mathbf{a}$ and $\mathbf{B}$ for each atmospheric variable and for each height. The coefficients of the bulk stability of the upper CPBL have obviously no height dependence. In case of plume properties q_t' and θ_v' we need to consider an alternative strategy. While each parameter configuration will result in plumes rising over at least a few model levels of the single column model, the top height of the plumes is not know a priori. Therefore, at a given model level one or more saturation points might not have any plumes in which case the thermodynamic perturbations q_t' and θ_v' would not be defined. For this reason we decide to compute the coefficients directly for the errors (see equation (10)) of these two-parameter profiles.

Acknowledgments
This material is based upon work supported by the U.S. Department of Energy, Office of Biological and Environmental Research under contract number ESD14091, and the Office of Advanced Scientific Computing Research, Applied Mathematics program under contract number DEAC02005CH11231. Optimization algorithms used in this paper are available at https://optimization.lbl.gov/downloads, and data presented in this paper is available at https://git.io/edmfopt. The first author thanks J. Teixeira and K. Sušelj for help with EDMF, and two reviewers are thanked for their contributions.

References

Arakawa, A. (1969). Parameterization of cumulus convection, *Proc. IV WMO/IUGG Symp. on Numerical Prediction* (Vol. 8, pp. 1–6). Tokyo, Japan: Japan Meteorological Agency.

Bellprat, O., Kotlarski, S., Lüthi, D., & Schär, C. (2012). Objective calibration of regional climate models. *Journal of Geophysical Research*, *117*, D23115. https://doi.org/10.1029/2012JD018262

Bony, S., Colman, R., Kattsov, V. M., Allan, R. P., Bretherton, C. S., Dufresne, J.-L., et al. (2006). How well do we understand and evaluate climate change feedback processes? *Journal of Climate*, *19*(15), 3445–3482.

Bony, S., & Dufresne, J. L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*, L20806. https://doi.org/10.1029/2005GL023851

Booker, A., Dennis, J. Jr., Frank, P., Serafini, D., Torczon, V., & Trosset, M. (1999). A rigorous framework for optimization of expensive functions by surrogates. *Structural Multidisciplinary Optimization*, *17*, 1–13.

Cess, R. D., Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Ghan, S. J., et al. (1989). Interpretation of cloud-climate feedback as produced by 14 atmospheric general circulation models. *Science*, *245*, 513–516.

Cheinet, S. (2003). A multiple mass-flux parameterization for the surface-generated convection. Part I: Dry plumes. *Journal of the Atmospheric Sciences*, *60*, 2313–2327.

Cheinet, S. (2004). A multiple mass flux parameterization for the surface-generated convection. Part II: Cloudy cores. *Journal of the Atmospheric Sciences*, *61*, 1093–1113.

Cheinet, S., & Teixeira, J. (2003). A simple formulation for the eddy-diffusivity parameterization of cloud-topped boundary layers. *Geophysical Research Letters*, *30*(18), 1930. https://doi.org/10.1029/2003GL017377

Dawson, A., & Palmer, T. N. (2015). Simulating weather regimes: Impact of model resolution and stochastic parameterization. *Climate Dynamics*, *44*(7), 2177–2193. https://doi.org/10.1007/s00382-014-2238-x

de Roode, S. R., Siebesma, A. P., Jonker, H. J. J., & de Voogd, Y. (2012). Parameterization of the vertical velocity equation for shallow cumulus clouds. *Monthly Weather Review*, *140*(8), 2424–2436.

Deardorff, J. W. (1966). Counter-gradient heat flux in lower atmosphere and in laboratory. *Journal of the Atmospheric Sciences*, *23*, 503–506.

Holland, J. (1992). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Cambridge, Massachusetts: MIT Press/Bradford Books.

Hourdin, F., Mauritsen, T., Gettelman, G., Golaz, J.-C., Balaji, V., Duan, Q., et al. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602.

Johnson, J. S., Cui, Z., Lee, L. A., Gosling, J. P., Blyth, A. M., & Carslaw, K. S. (2015). Evaluating uncertainty in convective cloud microphysics using statistical emulation. *Journal of Advances in Modeling Earth Systems*, *7*, 162–187. https://doi.org/10.1002/2014MS000383

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*, 425–464.

Khairoutdinov, M. F., & Randall, D. A. (2003). Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *Journal of the Atmospheric Sciences*, *60*, 607–625.

Lee, L. A., Carslaw, K. S., Pringle, K. J., & Mann, G. W. (2012). Mapping the uncertainty in global CCN using emulation. *Atmospheric Chemistry and Physics*, *12*(20), 9739–9751. https://doi.org/10.5194/acp-12-9739-2012

Lenschow, D. H., Wyngaard, J. C., & Pennell, W. T. (1980). Mean-field and second-moment budgets in a baroclinic, convective boundary layer. *Journal of the Atmospheric Sciences*, *37*, 1313–1326.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, *58*, 1246–1266.

Müller, J. (2017). SOCEMO: Surrogate Optimization of Computationally Expensive Multiobjective problems. *INFORMS Journal on Computing*, *29*(4), 581–596.
Müller, J., Paudel, R., Shoemaker, C., Woodbury, J., Wang, Y., & Mahowald, N. (2015). CH4 parameter estimation in CLM4.5bgc using surrogate global optimization. *Geoscientific Model Development Discussions*, *8*, 141–207.
Müller, J., & Piché, R. (2011). Mixture surrogate models based on Dempster-Shafer theory for global optimization problems. *Journal of Global Optimization*, *51*, 79–104.
Müller, J., & Shoemaker, C. (2014). Influence of ensemble surrogate models and sampling strategy on the solution quality of algorithms for computationally expensive black-box global optimization problems. *Journal of Global Optimization*, *60*, 123–144.
Müller, J., Shoemaker, C., & Piché, R. (2013). SO-MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. *Computers and Operations Research*, *40*, 1383–1400.
Müller, J., & Woodbury, J. (2017). GOSAC: global optimization with surrogate approximation of constraints. *Journal of Global Optimization*, *69*, 117–136. https://doi.org/10.1007/s10898-017-0496-y
Myers, R., & Montgomery, D. (1995). *Response surface methodology, process and product optimization using designed experiments*. Hoboken, NJ: Wiley-Interscience Publication.
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), 21,349–21,354.
Neggers, R. A. J., Siebesma, A. P., & Jonker, H. J. J. (2002). A multiparcel model for shallow cumulus convection. *Journal of the Atmospheric Sciences*, *59*, 1655–1668.
Nieuwstadt, F. T. M., Mason, P. J., Moeng, C.-H., & Schumann, U. (1993). A comparison of four computer codes, Proceedings of the 8th Symposium on Turbulent Shear Flows, Springer, pp. 343–367.
Nocedal, J., & Wright, S. (1999). *Numerical Optimization*. New York: Springer.
Powell, M. (1992). Advances in numerical analysis, vol. 2: Wavelets, subdivision algorithms and radial basis functions. Oxford University Press, Oxford, *chap. The theory of radial basis function approximation in 1990* (pp. 105–210). London: Oxford University Press.
Regis, R., & Shoemaker, C. (2007). A stochastic radial basis function method for the global optimization of expensive functions. *INFORMS Journal on Computing*, *19*, 497–509.
Regis, R., & Shoemaker, C. (2013). Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization. *Engineering Optimization*, *45*, 529–555.
Sakradzija, M., Seifert, A., & Dipankar, A. (2016). A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems*, *8*, 786–812. https://doi.org/10.1002/2016MS000634
Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*, 12,396–12,417. https://doi.org/10.1002/2017GL076101
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*, 3–5.
Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., et al. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of the Atmospheric Sciences*, *60*, 1201–1219.
Siebesma, A. P., Soares, P. M. M., & Teixeira, J. (2007). A combined eddy-diffusivity mass-flux approach for the convective boundary layer. *Journal of the Atmospheric Sciences*, *64*, 1230–1248.
Simpson, J., & Wiggert, V. (1969). Models of precipitating cumulus towers. *Monthly Weather Review*, *97*, 471–489.
Soares, P. M. M., Miranda, P. M. A., Siebesma, A. P., & Teixeira, J. (2004). An eddy-diffusivity/mass-flux parametrization for dry and shallow cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, *130*, 3365–3383.
Soden, B. J., & Held, I. M. (2006). An assessment of climate feedbacks in coupled ocean-atmosphere models. *Journal of Climate*, *19*(14), 3354–3360.
Stull, R. B. (1984). Transilient turbulence theory. 1. The concept of eddy-mixing across finite distances. *Journal of the Atmospheric Sciences*, *41*, 3351–3367.
Sušelj, K., Hogan, T. F., & Teixeira, J. (2014). Implementation of a stochastic eddy-diffusivity/mass-flux parameterization into the Navy Global Environmental Model. *Weather Forecasting*, *29*, 1374–1390.
Sušelj, K., Teixeira, J., & Chung, D. (2013). A unified model for moist convective boundary layers based on stochastic eddy-diffusivity/mass-flux parameterization. *Journal of the Atmospheric Sciences*, *70*, 1929–1953.
Tan, Z., Kaul, C. M., Pressel, K. G., Cohen, Y., Schneider, T., & Teixeira, J. (2018). An extended eddy-diffusivity mass-flux scheme for unified representation of subgrid-scale turbulence and convection. *Journal of Advances in Modeling Earth Systems*, *10*, 770–800. https://doi.org/10.1002/2017MS001162
Teixeira, J., & Cheinet, S. (2004). A simple mixing length formulation for the eddy-diffusivity parameterization of dry convection. *Boundary-Layer Meteorology*, *110*(3), 435–453.
Teixeira, J., & Siebesma, A. P. (2000). A mass-flux/k-diffusion approach for the parametrization of the convective boundary layer: global model results. In *Proc. 14th Symp. on Boundary Layers and Turbulence*, Amer. Meteor. Soc., Aspen, CO, pp. 231–234.
Webb, M. J., Lambert, F. H., & Gregory, J. M. (2013). Origins of differences in climate sensitivity, forcing and feedback in climate models. *Climate Dynamics*, *40*, 677–707.
Weisheimer, A., Corti, S., Palmer, T., & Vitart, F. (2014). Addressing model error through atmospheric stochastic physical parametrizations: Impact on the coupled ECMWF seasonal forecasting system. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *372*(2018), 20130290. https://doi.org/10.1098/rsta.2013.0290
Williamson, D., Goldstein, M., & Blaker, A. (2012). Fast linked analyses for scenario-based hierarchies. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, *61*, 665–691.
Zhao, M. (2014). An investigation of the connections among convection, clouds, and climate sensitivity in a global model. *Journal of Climate*, *27*, 1845–1862.
Zhao, M., Golaz, J.-C., Held, I. M., Ramaswamy, V., Lin, S.-J., Ming, Y., et al. (2016). Uncertainty in model climate sensitivity traced to representations of cumulus precipitation physics. *Journal of Climate*, *29*, 543–560.