Modeling the GABLS4 strongly-stable boundary layer with a GCM parameterization: parametric sensitivity or intrinsic limits?

Olivier Audouin[1], Romain Roehrig[2], Fleur Couvreux[3], and Danny Williamson[4]

[1]CNRM, 9 University of Toulouse, Meteo-France, CNRS
[2]CNRM, Université de Toulouse, Météo-France, CNRS
[3]Université Toulouse, CNRM, Meteo-France, CNRS
[4]University of Exeter

November 21, 2022

Abstract

The representation of stable boundary layers (SBLs) still challenges turbulence parameterizations implemented in current weather or climate models. The present work assesses whether these model deficiencies reflect calibration choices or intrinsic limits in currently-used turbulence parameterization formulations and implementations. This question is addressed for the ARPEGE-Climat 6.3 CNRM atmospheric model in a single-column model/large-eddy simulation (SCM/LES) comparison framework, using the history matching with iterative refocusing statistical approach. The GABLS4 case, which samples a nocturnal strong SBL observed at Dome C, Antarctic Plateau, is used. The standard calibration of the ARPEGE-Climat 6.3 turbulence parameterization leads to a too deep SBL, a too high low-level jet and misses the nocturnal wind rotation. This behavior is found for low and high vertical resolution model configurations. The statistical tool then proves that these model deficiencies reflect a poor parameterization calibration rather than intrinsic limits of the parameterization formulation itself. In particular, the role of two lower bounds that were heuristically introduced during the parameterization implementation to increase mixing in the free troposphere and to avoid runaway cooling in snow- or ice-covered region is emphasized. The statistical tool identifies the space of the parameterization free parameters compatible with the LES reference, accounting for the various sources of uncertainty. This space is non-empty, thus proving that the ARPEGE-Climat 6.3 turbulence parameterization contains the required physics to capture the GABLS4 SBL. The SCM framework is also used to validate the statistical framework and a few guidelines for its use in parameterization development and calibration are discussed.

Modeling the GABLS4 strongly-stable boundary layer with a GCM parameterization: parametric sensitivity or intrinsic limits?

O. Audouin[1], R. Roehrig[1], F. Couvreux[1], and D. Williamson[2,3]

[1]CNRM, University of Toulouse, Meteo-France, CNRS, Toulouse, France
[2]Exeter University, Exeter, United Kingdom
[3]The Alan Turing Institute, 96 Euston Road, London, United Kingdom

Key Points:

- The ARPEGE-Climat 6.3 performance in modeling a stable boundary layer is assessed in a single-column model/large-eddy simulation comparison
- The use of history matching with iterative refocussing indicates turbulence parameterization deficiencies due to poor calibration
- Validation of the statistical framework and guidelines for its use in parameterization development and calibration are discussed

Corresponding author: Olivier Audouin, `olivier.audouin@meteo.fr`

Abstract

The representation of stable boundary layers (SBLs) still challenges turbulence parameterizations implemented in current weather or climate models. The present work assesses whether these model deficiencies reflect calibration choices or intrinsic limits in currently-used turbulence parameterization formulations and implementations. This question is addressed for the ARPEGE-Climat 6.3 CNRM atmospheric model in a single-column model/large-eddy simulation (SCM/LES) comparison framework, using the history matching with iterative refocusing statistical approach. The GABLS4 case, which samples a nocturnal strong SBL observed at Dome C, Antarctic Plateau, is used. The standard calibration of the ARPEGE-Climat 6.3 turbulence parameterization leads to a too deep SBL, a too high low-level jet and misses the nocturnal wind rotation. This behavior is found for low and high vertical resolution model configurations. The statistical tool then proves that these model deficiencies reflect a poor parameterization calibration rather than intrinsic limits of the parameterization formulation itself. In particular, the role of two lower bounds that were heuristically introduced during the parameterization implementation to increase mixing in the free troposphere and to avoid runaway cooling in snow- or ice-covered region is emphasized. The statistical tool identifies the space of the parameterization free parameters compatible with the LES reference, accounting for the various sources of uncertainty. This space is non-empty, thus proving that the ARPEGE-Climat 6.3 turbulence parameterization contains the required physics to capture the GABLS4 SBL. The SCM framework is also used to validate the statistical framework and a few guidelines for its use in parameterization development and calibration are discussed.

Plain Language Summary

During the night or in snow- or ice-covered region, a stable atmospheric boundary layer (SBL) often develops. Their representation still challenges turbulence parameterizations implemented in numerical weather or climate models. The present work assesses whether the ARPEGE-Climat atmospheric model deficiencies reflect calibration choices or intrinsic limits in its turbulence parameterization using statistical approach from the Uncertainty Quantification Community. A single-column version of the model is evaluated on the GABLS4 case, a nocturnal strong SBL observed at Dome C, Antarctic plateau, and compared to high-resolution simulations. The standard calibration of the ARPEGE-Climat 6.3 turbulence parameterization leads to a too deep SBL and an incorrect wind pattern and so for different vertical resolutions. The statistical tool proves that these model deficiencies are rectified with proper calibration of the turbulence parameterization. In particular, it is shown that two lower bounds, introduced to increase turbulent mixing, are key to capture the GABLS4 SBL. Finally, the potential and relevance of the Uncertainty Quantification approach applied to the Single-Column Model/Large-Eddy Simulation comparison framework for the calibration of climate models are highlighted and few guidelines for its use are proposed.

1 Introduction

In the atmospheric boundary layer, stably stratified conditions generally develop above a surface colder than the overlying air. These Stable Boundary Layers (SBLs) often result from the advection of warm air over a cold surface or from the cooling of the surface. They are frequently observed over ice or snow surfaces (e.g., polar regions, high-latitude continental regions in wintertime) and over land during nighttime. Their development and intensity (e.g., the vertical stratification) are also strongly modulated by the atmospheric synoptic conditions. Weak cloud cover during nighttime also favors SBL occurrence, as such conditions enhance the radiative cooling of the surface (Mahrt, 1998). In contrast, the occurrence of strong near-surface wind reduces the SBL stratification, or even inhibits their development, through the maintenance of significant mechanical mixing (e.g., Van de Wiel et al., 2012).

SBL can be classified according to the intensity of their stratification (at first order the vertical gradient of potential temperature), ranging from weak SBLs in which the turbulence remains significant, to strong SBLs, in which the turbulence become intermittent or even disappears (e.g. Mahrt, 1998; Acevedo et al., 2016; van Hooft et al., 2017). In the latter conditions, a mechanical decoupling between the atmosphere and the surface can occur (Derbyshire, 1999): the temperature inversion close to the surface becomes driven by radiation and soil diffusion and the surface turbulent heat flux cannot sustain the surface energy demand enhanced by a strong net surface radiative cooling (e.g. Van de Wiel et al., 2012; van Hooft et al., 2017). Such strong SBLs mostly occur under clear-sky and weak wind conditions, with a strong increase of the near-surface temperature inversion below a critical wind speed (e.g. van Hooft et al., 2017; Vignon, van de Wiel, et al., 2017) .

The representation of SBLs in operational climate and weather models is a challenge (e.g. Holtslag et al., 2013): the turbulence is particularly weak and sometimes intermittent (e.g. Mauritsen & Svensson, 2007), and interacts with other small-scale processes (e.g., gravity waves Steeneveld et al., 2008; Tsiringakis et al., 2017). Under the umbrella of the Global Energy and Water Cycle Exchanges (GEWEX) project, the GEWEX Atmospheric Boundary Layer Study (GABLS) has initiated four model intercomparison projects (Cuxart et al., 2006; Svensson et al., 2011; Bosveld et al., 2014; Bazile et al., 2015) to evaluate and improve the SBL representation in weather and climate models. So far, the GABLS intercomparison exercises revealed:

1. Large-eddy simulations (LES) are able to consistently capture the main properties of stable boundary layers, at least when their resolution is below a few meters for weak to moderate SBL (e.g. Beare et al., 2006) or 1 m for strong SBL (Couvreux, Bazile, et al., 2020). As a result, such LES provide relevant process-level information to evaluate turbulence parameterizations in an LES/Single-column model (SCM) comparison framework (e.g. D. A. Randall et al., 1996; D. Randall et al., 2003)
2. State-of-the-art turbulence parameterizations, such as those with a 1.5-order turbulence closure, are able to reasonably capture the physics of SBLs for a wide range of forcing (e.g. Cuxart et al., 2006; Baas et al., 2018; Vignon, Hourdin, et al., 2017)
3. Nevertheless, current weather and climate models still simulate SBLs that are too deep, with surface drag that is too strong, low-level jets that are too weak and too high and wind veering with height that is too weak (e.g. Cuxart et al., 2006; Holtslag et al., 2013).

The apparent contradiction between the two last conclusions results from the calibration of weather or climate models which, so far, has required an increased turbulent mixing to reduce the activity of synoptic systems, and thereby improve operational scores (e.g. Sandu et al., 2013), or to prevent runaway surface cooling through long-term mechanical decoupling with the atmosphere (Derbyshire, 1999). Such a calibration probably reflects the lack of mixing due to processes that are currently not accounted for in weather and climate models (e.g., surface heterogeneities, internal gravity waves, impact of subgrid orography).

Recently Vignon, Hourdin, et al. (2017), Vignon et al. (2018) and Hourdin et al. (2020) showed that it is possible to achieve a reasonable representation of SBLs in a climate model (LMDZ), while maintaining reasonable large-scale performance. Starting with an SCM framework built on the very stable boundary layer of GABLS4 (Bazile et al., 2015), (Vignon, Hourdin, et al., 2017) underline the importance of (i) the coupling with the surface (snow albedo and thermal inertia) and (ii) the turbulent mixing thresholds usually used in current operational turbulence parameterizations (e.g., for the mixing length or in stability functions). More specifically, the appropriate calibration of surface (snow) properties and the removal of those thresholds in the turbulence parameterization allows the LMDZ SCM to capture well the strong temperature gradient close to the surface (in the first 20 meters), observed at Dome C, Antarctica, during an austral summer night. (Vignon et al., 2018) follow on with 3D LMDZ simulations facing the observations collected at Dome C during one year. On the one hand, the SCM improved results are consistently reported in this 3D configuration, stressing the relevance of the SCM framework for developing and calibrating parameterizations. On the other hand, the new version of the LMDZ model adequately reproduces the annual cycle of Dome C, the two main SBL regimes discussed above, and preserves satisfying large-scale skills. Finally, the vertical resolution in the lower part of the boundary layer is also shown to be critical for capturing SBLs that cover only a few tens of meters (see also Steeneveld et al., 2006).

Following (Vignon, Hourdin, et al., 2017; Vignon et al., 2018), the general objective of the present work is to document the performance of the CNRM climate atmospheric model, namely ARPEGE-Climat 6.3 (Roehrig et al., 2020), to represent SBLs. The focus is here on its turbulence parameterization, which is based on the work of Cuxart et al. (2000). The parameterization of a given process seeks to represent its effects on the large-scale (or resolved) state of the model. It is based on a set of physical theories or empirical relationships to numerically describe the subgrid-scale processes and their effects. Parameterizations introduce a number of constants, called free parameters in the following, which are often difficult to constrain with observations or other references. A parameterization can thus be seen as a function of the model state variables and of these free parameters. Their calibration, or "tuning", is a critical step in model development for weather or climate applications (e.g. Hourdin et al., 2017). In the present paper, we therefore propose to address the following specific question: Is it possible to calibrate the ARPEGE-Climat turbulence parameterization to achieve a satisfying representation of SBLs, especially those with a strong thermal stratification ? In other words, does the ARPEGE-Climat turbulence parameterization contain the required physics to represent appropriately strong SBLs ?

(Cuxart et al., 2006) shows that the current turbulence parameterization of ARPEGE-Climat 6.3 is able to capture the main properties of the moderate SBL for the first GABLS exercise. We seek to extend this result to the strongly-stratified SBL of the GABLS4 nocturnal phase. We rely on SCM simulations, which have been shown relevant for 3D model configuration (Hourdin et al., 2013; Neggers, 2015; Vignon et al., 2018; Gettelman et al., 2019). We also make use of GABLS4 LES as references, as they have been shown to capture well the properties of the GABLS4 nocturnal phase (Couvreux, Bazile, et al., 2020). Following (Couvreux, Hourdin, et al., 2020), we use statistical tools developed in the Uncertainty Quantification community, in particular the history matching proposed by (D. Williamson et al., 2013) and applied to the SCM/LES comparison. This tool provides the sensitivity analysis of our turbulence parameterization to its free parameters and identifies which part of the full free parameter space provides SCM simulations consistent with the chosen reference, accounting for the various sources of uncertainty (D. Williamson et al., 2013; D. B. Williamson et al., 2017; Couvreux, Hourdin, et al., 2020). Instead of optimizing the ARPEGE-Climat turbulence parameterization over the GABLS4 SBL, and thus possibly facing overfitting issues, the approach provides useful information to continue the calibration process over other 1D cases and in the full 3D model configuration, while keeping an acceptable behavior for the GABLS4 SBL.

Section 2 introduces the ARPEGE-Climat 6.3 atmospheric model and its turbulence parameterization. The relevant free parameters of the parameterization to be used for calibration are emphasized. Section 3 presents the case study used for the SCM/LES intercomparison and the LES results that serve as a reference. Section 4 describes the statistical framework. Section 5 details the results obtained for three different configurations of the ARPEGE-Climat 6.3 SCM. Section 6 discusses several aspects of the methodology and Section 7 finally concludes the present study.

2 ARPEGE-Climat 6.3

ARPEGE-Climat is a global atmospheric model developed at CNRM for climate studies. Its latest version (6.3, Roehrig et al., 2020) is the atmospheric component of the CNRM ocean-atmosphere climate model CNRM-CM6-1 (Voldoire et al., 2019), and Earth System model CNRM-ESM2-1 (Séférian et al., 2019). The following work uses the single column model (SCM) version of ARPEGE-Climat (e.g. Abdel-Lathif et al., 2018), in the context of the GABLS4 framework (see 3.1). The model physical package is fully described in Roehrig et al. (2020) and therefore we only insist hereafter on the model features relevant for the present study. ARPEGE-Climat 6.3 standard vertical grid consists of 91 vertical levels, following the progressive hybrid σ-pressure discretization of Simmons and Burridge (1981). The altitude of the first 5 model levels is approximately 8, 29, 56, 91 and 132 m. The model timestep is 15 minutes. A version of ARPEGE-Climat 6.3 with higher vertical resolution (2 m up to 400 m) is also used in section 5.1. To prevent instabilities, the timestep of this version is reduced to 60 seconds. Note that the use of this 60-s timestep in the 91-level version of ARPEGE-Climat does not impact much the results of the present work. As described in section 3, the SCM configuration is run on a idealized case (stable boundary layer, no moisture, no radiation), in which only the turbulence and surface flux parameterizations are activated. These parameterizations are described hereafter, in a dry context.

2.1 Turbulence parameterization

The turbulence scheme used in ARPEGE-Climat 6.3 follows the work of J. Redelsperger and Sommeria (1982), J.-L. Redelsperger and Sommeria (1986), and Cuxart et al. (2000). It relies on the eddy diffusivity approach, coupled to a prognostic equation for the grid-scale-averaged turbulence kinetic energy (TKE) $\overline{e}$. Given the standard horizontal resolution of ARPEGE-Climat ($\mathcal{O}$(100km)), only the vertical component of turbulent mixing is parameterized. For any variable ψ impacted by turbulent mixing (e.g., wind component u and v, potential temperature θ), the associated second-order turbulent flux $\overline{w'\psi'}$ reads (primes denote fluctuations with respect to the grid-scale average, noted $\bar{\psi}$):

$$\overline{w'\psi'} \quad = \quad -K_\psi \frac{\partial \bar{\psi}}{\partial z} \quad ; \quad K_\psi \quad = \quad \alpha_\psi \mathbf{CM} L_m \sqrt{\overline{e}}\phi_\psi \tag{1}$$

where α_ψ and $\mathbf{CM}$ are free parameters of the parameterization, L_m is the mixing length, and ϕ_ψ is a stability function. ϕ_ψ is taken to 1 for momentum and turbulence kinetic energy ($\psi \in \{u, v, e\}$). For the potential temperature θ, the following formulation is used:

$$\phi_\theta = \frac{1}{1 + C\frac{g}{\overline{\theta}}\frac{L_m^2}{\overline{e}}\frac{\partial\overline{\theta}}{\partial z}} \text{ where } C \text{ is a free parameter.} \tag{2}$$

In Equation 1, $\mathbf{CM}$ modulates all turbulent fluxes in the same way. α_u and α_v are taken to 1, and α_θ is the inverse Prandtl number in neutral condition (i.e. when $\phi_\theta = 1$). In the following, α_e and α_θ will be referred to as **AE** and **AT**, respectively.

Eddy diffusivity coefficients K_ψ depend on the intensity of $\overline{e}$. The time evolution of $\overline{e}$ is given by:

$$\frac{\partial \overline{e}}{\partial t} = -\frac{1}{\overline{\rho}}\frac{\partial}{\partial z}(\overline{\rho}\overline{e'w'}) - \left(\overline{u'w'}\frac{\partial \overline{u}}{\partial z} + \overline{v'w'}\frac{\partial \overline{v}}{\partial z}\right) + \frac{g}{\overline{\theta}}\overline{w'\theta'} - \frac{\overline{e}\sqrt{\overline{e}}}{L_\epsilon} \tag{3}$$

where ρ is the air density, g is the gravity acceleration, and L_ϵ the dissipation length. L_ϵ is assumed to be proportional to the mixing length: $L_\epsilon = \mathbf{CE}L_m$, with $\mathbf{CE}$ a free parameter.

The mixing length follows the non-local formulation of Bougeault and Lacarrere (1989) and reads

$$L_m^{\mathrm{BL89}} = \left[\frac{1}{2}\left((L_{\mathrm{up}})^{-2/3} + (L_{\mathrm{down}})^{-2/3}\right)\right]^{-3/2} \tag{4}$$

where L_{up} and L_{down} are respectively the maximum upward and downward displacements a parcel can travel within the ambient thermal stratification, given its turbulence kinetic energy, and accounting only for the work of its buoyancy. A minimum mixing length, **LMIN**, is introduced to maintain a minimum vertical mixing in stable boundary layers. Close to the surface, the mixing length is also supposed to be larger than κz where $\kappa = 0.4$ is the Von Kármán constant. Thus the mixing length L_m reads:

$$L_m = \max\left[L_m^{\mathrm{BL89}}, \min(\mathbf{LMIN}, \kappa z)\right] \tag{5}$$

In case of shallow stable boundary layer, another lower bound, which applies mainly close to the surface, is introduced directly on the turbulent fluxes, to avoid runaway cooling of the surface (especially in snow- or ice-covered regions):

$$\overline{w'\psi'} = \max\left[-K_\psi \frac{\Delta\overline{\psi}}{\Delta z}, \alpha_\psi \left(\mathbf{KOZMIN}(1 - \frac{z}{\mathbf{ZMAX}})\right) \Delta\overline{\psi}\right] \tag{6}$$

where **KOZMIN** and **ZMAX** are two free parameters, and $\Delta\overline{\psi}$ is the vertical difference of $\overline{\psi}$ between two consecutive model layers (distant of Δz). Above **ZMAX**, no lower bound is used. This formulation, through $\Delta\overline{\psi}$ and Δz depends on the vertical discretization of the model (see also section 5).

The turbulence parameterization thus includes several free parameters that have to be calibrated. Eight parameters have been identified here. The calibration of the parameterization consists in choosing a value for each of them, accounting for both parameterization performance and physical constraints. In the standard configuration of ARPEGE-Climat 6.3 (see Roehrig et al., 2020), the parameter values follow the work of Cheng et al. (2002) except the parameters **LMIN**, **KOZMIN** and **ZMAX** that were introduced in the course of the parameterization implementation in ARPEGE-Climat and set in a more empirical way. Table 1 provides the values of these parameters as currently used in ARPEGE-Climat 6.3 as well as those initially proposed in Cuxart et al. (2000). Note that the parameter C in Equation 2 is set to 0.143. As the model is not much sensitive to it, the following work does not consider this parameter.

Table 1. Free parameters of the turbulence parameterization. The values in the standard version of ARPEGE-Climat are those from Cheng et al. (2002). Those from the work of Cuxart et al. (2000) are also included. The bottom two lines provide the range of values that we explore for each parameter.

	CM	**AE**	**AT**	**CE**	**LMIN**	**KOZMIN**	**ZMAX**
ARPEGE-Climat 6.3	0.126	2.70	1.13	0.85	10.0	5e-3	200
Cuxart et al. (2000a)	0.0667	6.0	2.5	0.70	10.0		
Lower bound	0.05	0.50	0.20	0.33	0.0	0.0	30
Upper bound	0.30	6.00	3.00	5.00	10.0	5e-3	400

2.2 Surface flux parameterization

The SCM configuration of ARPEGE-Climat will be used in two different configurations with respects to the surface boundary conditions, one with prescribed surface sensible heat flux and one with prescribed surface temperature. In both, the roughness lengths for momentum (z_0) and heat (z_{0_h}) are prescribed.

2.2.1 Configuration with prescribed surface sensible heat flux

The friction velocity u_* is computed following Paulson (1970):

$$u_* = \frac{\kappa \overline{U_1}}{\ln\left(\frac{z_1}{z_0}\right) - \varphi\left(\frac{z_1}{L_{\mathrm{MO}}}\right) + \varphi\left(\frac{z_0}{L_{\mathrm{MO}}}\right)} \tag{7}$$

where $\overline{U}_1$ is the wind intensity ($\overline{U}_1 = \sqrt{\overline{u}_1^2 + \overline{v}_1^2}$) at the first model level of altitude z_1, and L_{MO} is the Monin-Obukhov length. The similarity function φ is given by Paulson (1970). As L_{MO} depends on u_*, the computation is done iteratively, initialized from a neutrally-stable state (i.e. $L_{\mathrm{MO}} = \infty$, knowing $\varphi(0) = 0$). The surface momentum flux is finally given by:

$$F_u = \overline{\rho} u_*^2 \tag{8}$$

2.2.2 Configuration with prescribed surface temperature

In this configuration, the standard version of the ARPEGE-Climat surface scheme is used. The surface momentum and heat fluxes are computed based on the formulations of Mascart et al. (1995) involving the bulk Richardson number Ri_b^0:

$$\mathrm{Ri}_b^0 = \frac{g z_1 (\overline{\theta}_1 - \theta_s)}{\frac{1}{2}(\overline{\theta}_1 + \theta_s)\overline{U}_1^2} \tag{9}$$

where $\overline{\theta}_1$ and $\overline{\theta}_s$ are the potential temperature at the first model level and at the surface, respectively. A critical Richardson number $\mathrm{Ri}_c = 0.1$ is used as a lower bound of the bulk Richardson number: $\mathrm{Ri}_b = \min(\mathrm{Ri}_b^0, \mathrm{Ri}_c)$. The exchange coefficients for memtum and heat in the case of stable states ($\mathrm{Ri}_b > 0$) following Mascart et al. (1995) and Noilhan and Mahfouf (1996) read:

$$C_d = \frac{\kappa^2}{\left(\ln\left(\frac{z}{z_0}\right)\right)^2} \frac{1}{1 + \frac{\mathrm{B}_1 \mathrm{Ri}_b}{\sqrt{1 + \mathrm{B}_2 \mathrm{Ri}_b}}} \tag{10}$$

$$C_h = \frac{\kappa^2}{\left(\ln\left(\frac{z}{z_0}\right)\right)^2} \left(\frac{\ln(z/z_0)}{\ln(z/z_{0h})}\right) \left(\frac{1}{1 + \mathrm{B}_3 \mathrm{Ri}_b \sqrt{1 + \mathrm{B}_2 \mathrm{Ri}_b}}\right) \tag{11}$$

and are used to compute the surface fluxes:

$$F_u = \overline{\rho} C_d \overline{U_1}^2 \quad \text{and} \quad F_s = \overline{\rho} C_p C_h \overline{U_1} (\theta_s - \overline{\theta}_1) \tag{12}$$

where C_p is the heat capacity of air at constant pressure.

Note that the present surface flux parameterizations include internal free parameters ($\mathrm{B}_1 = 10$, $\mathrm{B}_2 = 5$, $\mathrm{B}_3 = 15$), which are not considered in the following analysis.

They are possibly critical for SBLs (e.g. Vignon, Hourdin, et al., 2017), and will be analysed in a future work.

3 Experimental setup and reference simulations

3.1 The GABLS4 framework

The present study is based on the GABLS4 model intercomparison case (Bazile et al., 2014, 2015; Couvreux, Bazile, et al., 2020). It focuses on an austral summer diurnal cycle of the boundary layer at Dome C, Antarctic Plateau (123.3E, 75.1S, 3223 m above sea level, local time (LT) = UTC+8 hours) as observed from 11 December 0800 LT to 12 December 0800 LT. During that day, the boundary layer evolved from a 400-m deep convective regime during daytime to a nighttime very stable regime covering a depth shallower than 30 m (Vignon, Hourdin, et al., 2017).

The GABLS4 model intercomparison encompasses three different stages. The first one is dedicated to the intercomparison of SCMs with an interactive snow surface scheme. The second stage prescribes observed surface temperature, thus suppressing several feedbacks between the atmosphere and the surface. The third stage consists in an idealization of GABLS4 stage 2, in which no moisture, no radiation, no large-scale subsidence and no large-scale advection of temperature are considered.

Couvreux, Bazile, et al. (2020) emphasize that the representation of the full GABLS4 diurnal cycle is a challenge for LES as it requires a large domain for the 400-m deep daytime convective boundary layer and a very high-resolution for the 30-m deep nocturnal very stable boundary layer. Therefore, Couvreux, Bazile, et al. (2020) proposed a complementary setup focused on the GABLS4 nocturnal stable phase, starting at the end of the convective period (1800 LT, i.e. 10 hours after the start of the original version of GABLS4 stages) and covering 11 hours (until 0500 LT). This new setup, referred to as GABLS4-Stage3-10hr, corresponds to the setup used in the present work.

The initial conditions are obtained from the ensemble mean of three LES that took part to the GABLS4 LES intercomparison (Couvreux, Bazile, et al., 2020). GABLS4-Stage3-10hr uses the same large-scale forcing as in GABLS4 Stage 3, which thus only includes a large-scale horizontal pressure gradient through a prescribed geostrophic wind. This geostrophic wind is constant in time ($u_g = 1.25\,\mathrm{ms}^{-1}$ and $v_g = 4.5\,\mathrm{ms}^{-1}$) and along height. GABLS4 Stage 3 (and thus GABLS4-Stage3-10hr) assumes a dry atmosphere, with no radiation. The surface pressure is held constant to 651 hPa (Dome C is at 3223 m above sea level), and the surface temperature is prescribed and evolves with time, following the observations made at Dome C. For the computation of the surface wind stress, the surface roughness length is set to $z_0 = 10^{-3}\,\mathrm{m}$ for momentum and $z_{0_h} = 10^{-4}\,\mathrm{m}$ for heat, following Vignon, van de Wiel, et al. (2017) and Couvreux, Bazile, et al. (2020).

3.2 GABLS4 Large-Eddy Simulations

Couvreux, Bazile, et al. (2020) compare 7 LES models over GABLS4-Stage3-10hr. Two of them are not considered here because of a slightly different setup compared to the other five (slightly coarser resolution or different roughness lengths). The five remaining LES use an isotropic resolution of 1 m over a 500 m x 500 m x 150 m domain. The LES compute their surface fluxes (momentum and heat) from the prescribed surface temperature and roughness lengths using their own parameterization. In such a setup, especially thanks to the high resolution, the spread among the LES ensemble is rather small, except very close to the surface. In particular, they are substantial differences for the surface sensible heat flux (e.g., -6 to -13 W m^{-2} at 2300 LT) or for the friction velocity (0.07

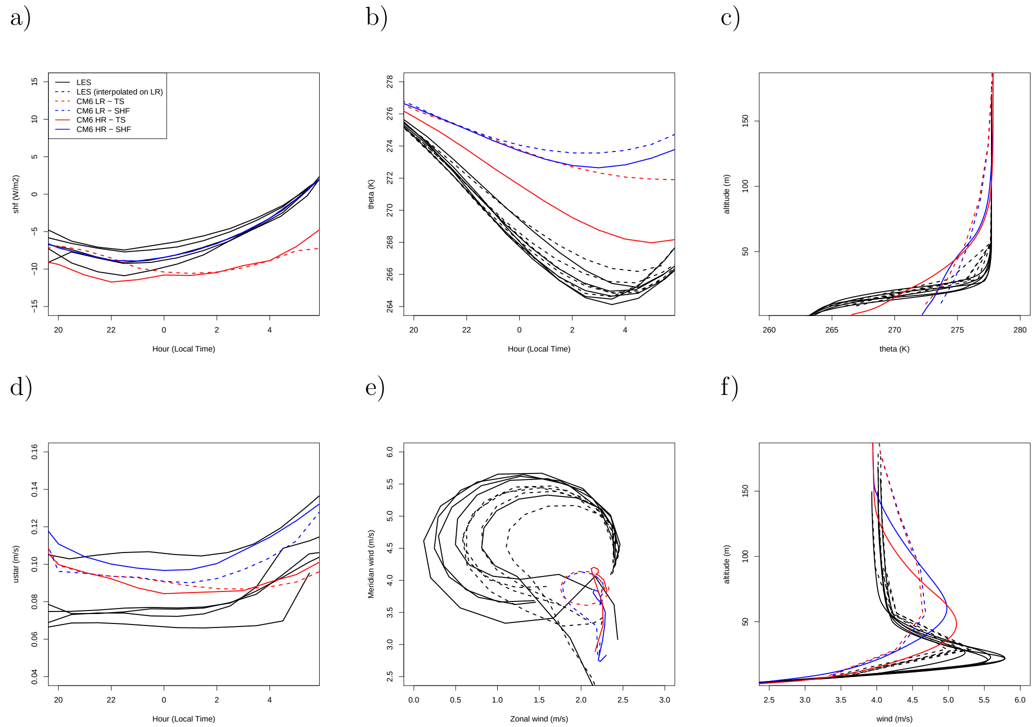


Figure 1. (a) Sensible heat flux (W m^{-2}), (b) 8.5-m potential temperature (K), (c) potential temperature vertical profile at 0300 LT (K), (d) surface friction velocity (m s^{-1}), (e) wind rotation at 29 m and (f) wind speed vertical profile at 0100 LT for the LES (solid black lines), LES interpolated on CM6-LR levels (dashed black lines), and CM6-LR (dashed lines) and CM6-HR (solid lines) SCM simulations. CM6 simulations are either forced by the surface temperature (red lines) or by the surface sensible heat flux (blue lines).

to 0.11 ms^{-1}) - see Fig. 1a and 1d. This spread is however much smaller than the spread among the full Stage 3 LES ensemble (Couvreux, Bazile, et al., 2020).

Although the setup is idealized compared to the observed situation, observations have been used to evaluate the simulation behavior. The heights of the stable boundary layer and of the low-level jet are overestimated compared to observations but this might be due to the neglect of subsidence in the LES simulations. Otherwise, the intensity of the jet and that of the stratification are consistent with observations. The wind turning at 41 m is also much more realistic with the high-resolution LES than the results obtained with the coarse-resolution LES (Couvreux, Bazile, et al., 2020).

3.3 SCM configuration

The ARPEGE-Climat Single Column model is run for two different setups. The first one follows exactly the GABLS4-Stage3-10hr setup detailed in section 3.1. The second one, derived from GABLS4-Stage3-10hr, and referred to as GABLS4-Stage3-10hr-shf, prescribes the surface sensible heat flux, instead of the surface temperature, and thus further removes the coupling between the surface and the atmosphere. In this latter setup, the prescribed surface sensible heat flux corresponds to the ensemble mean of the 5 LES.

In both setups, all model parameterizations are deactivated, except those for the atmosphere turbulence and the surface fluxes. Note that, in the GABLS4-Stage3-10hr setup, the surface flux parameterization follows the work of Mascart et al. (1995) (Section 2.2), while in the GABLS4-Stage3-10hr-shf, it is replaced by the simplified version described in Paulson (1970) (Section 2.2).

We also explore the turbulence parameterization behavior for two different vertical resolutions, namely the standard vertical resolution of ARPEGE-Climat 6.3 (91 vertical levels, with about 15 levels below 1500 m), and a constant high-vertical resolution of 2 m up to an altitude of about 400 m and then decreasing as the standard vertical grid up to the model top, called CM6-HR in the following.

4 Statistical framework : History matching

The calibration of the free parameters of a parameterization is a difficult task due to many degrees of freedom (i.e. the number of parameters) and the computational cost of operational configuration simulations. For example, the turbulence parameterization used here has eight different free parameters and seven are kept for the calibration experiences. If one wanted to systematically explore this 7-dimension space, let say with 10 values in each parameter range, this would require 10^7 simulations. This is clearly prohibitive for most model configurations, even in a SCM framework. Therefore surrogate models (i.e. mathematical functions that approximate SCM outputs for a negligible computational cost), trained on a few simulations, are used to emulate the full model behavior in the parameter space. In the following work, we use history matching with iterative refocussing as proposed in D. Williamson et al. (2013). The main objective of the statistical approach is not to find the optimal set of parameters, but rather to remove regions of parameter space in which the model behavior is inappropriate for a given set of performance metrics, accounting for several sources of uncertainty. The approach thus reduces the risk for overtuning (Hourdin et al., 2017; D. Williamson et al., 2015). The algorithm is iterative, in the sense that several consecutive waves are considered. Each wave consists of a small number of simulations performed with the full model, which allow us to improve the surrogate model accuracy where needed (i.e within the acceptable range from the previous wave) and to refocus the search for acceptable parameter values in a reduced space during the following wave. The process stops when convergence is achieved for the space of acceptable parameters. Here, this approach is applied in a SCM-LES comparison framework as detailed in Couvreux, Hourdin, et al. (2020) and briefly presented below.

1. Targeted metrics are first selected. They aim at summarizing the model performance. Reference metrics are computed, here LES results are used.
2. Free parameters of the physic parameterization are selected and their possible range (generally determined from the modeler expertise) are identified. In our case, only parameters from the turbulence scheme are selected (see Section 2.1).
3. From the selection of the free parameters (Table 1), the first wave experimental design is built. To ensure an optimal sampling of the initial parameter space, a latin hypercube method is used (D. Williamson, 2015).
4. Metrics are computed from the first wave simulations and used as a training sample to build the surrogate models based on machine learning methods. Among the many possible approaches, we use Gaussian Processes, which have the advantage to predict both the metric and its uncertainty (Salter & Williamson, 2016).
5. The parameter space is then systematically explored using surrogate models and emulated metrics are compared to the reference. The space of acceptable parameter values is determined iteratively by ruling out the parts of the full parameter space which lead to metric values too far from the reference value, accounting for the uncertainty on the reference, the SCM and the surrogate model. For a given metric f, the following measure $I_f(\boldsymbol{\lambda})$, referred to as implausibility, is thus introduced:

$$I_f(\boldsymbol{\lambda}) = \frac{|r_f - \mathrm{E}[e(\boldsymbol{\lambda})]|}{\sqrt{\sigma^2_{r,f} + \sigma^2_{d,f} + \mathrm{Var}[e(\boldsymbol{\lambda})]}} \tag{13}$$

where $\boldsymbol{\lambda}$ is a point of the parameter space, $e(\boldsymbol{\lambda})$ the metric value predicted by the surrogate model. More precisely, $\mathrm{E}[e(\boldsymbol{\lambda})]$ is the expectation of the metric, $\mathrm{Var}[e(\boldsymbol{\lambda})]$ is the variance of the metric, which is a measure of the surrogate model uncertainty at the point $\boldsymbol{\lambda}$. r_f is the reference metric value (i.e. the mean of the metric f computed from the LES ensemble). $\sigma^2_{r,f}$ is the reference uncertainty and is computed as the variance of the metric f computed from the LES ensemble. $\sigma^2_{d,f}$ is the SCM discrepancy or structural error for this metric. This last value is not known *a priori* and its estimation could be challenging. The implausibility thus measures the distance between the reference and the predicted metric value, normalized by the sum of uncertainties (supposed to be independent). The implausibility can thus be small either because the predicted value is close to the reference, or because the uncertainties are large. The threshold used to rule out the inappropriate regions of the initial parameter space, and thus to define the Not Ruled Out Yet (NROY) space is a free parameter of the approach. We take 3, a rather conservative value, which reduces the risk of ruling out an acceptable point. This value is chosen following the 3-σ rule for any unimodal distribution (Pukelsheim, 1994) which states that at least 95% of the distribution lies in the range of 3 σ around the mean. This threshold can be reduced once the surrogate model is sufficiently accurate. In the case of multiple metrics, we can either form a multivariate implausibility (taken as the maximum of the implausibilities computed for each metric) or define the NROY space as the space where most metrics meet the constraints. The latter option seeks to avoid multiple testing problems (e.g. Vernon et al., 2010; Couvreux, Hourdin, et al., 2020). As the number of metrics used in this work is relatively small (less than 4), we subsequently use the multivariate implausibility option.
Following the iterative refocussing philosophy, once the NROY space is determined at the end of the current wave, it is further resampled to define the next wave, which will provide a few more simulations performed with the full model and thus improve the surrogate model accuracy (i.e. reduce $\mathrm{Var}[e(\boldsymbol{\lambda})]$) within the NROY space. It can be noted here that $\sigma_{r,f}$ and $\sigma_{d,f}$ are independent of the surrogate model and remain constant along the history matching process. Once the surrogate model uncertainty is sufficiently reduced so that the other uncertainties dominate the implausibility, the NROY space is not further reduced by new waves. The iterative process has thus converged and it is not necessary to perform additional waves. The convergence questionis discussed in section 6.2.

5 Ability of ARPEGE-Climat to simulate the GABLS4 stable boundary layer

The tool introduced in the previous section is used to assess whether the ARPEGE-Climat turbulence scheme is able to capture the main properties of the GABLS4 stable boundary layer. In particular, we determine which part of the space of model free parameter values is compatible with the LES references. The vertical resolution is potentially a critical aspect for the scheme and therefore we start with a high-resolution configuration of ARPEGE-Climat (referred to as CM6-HR, cf Section 2). We also remove a degree of freedom in the surface-atmosphere coupling by prescribing the surface sensible heat flux instead of the surface temperature, to focus only on the turbulence scheme (referred to as CM6-HR-SHF). Later in this section, subsections 5.2 and 5.3 discuss results with the standard vertical resolution of ARPEGE-Climat and with two different surface boundary conditions.

5.1 High-resolution SCM configuration forced by surface sensible heat flux (SCM-HR-SHF)

5.1.1 ARPEGE-Climat standard calibration

As a starting point of the present work, we evaluate the standard calibration of ARPEGE-Climat turbulence scheme (parameter values indicated in Table 1), when forced by the LES ensemble mean surface sensible heat flux (CM6-HR-SHF). Figure 1b presents the time evolution of the 8.5-m potential temperature from 1800 to 0700 LT. 8.5 m is approximately the altitude of the first level in the standard vertical resolution model (between the third and fourth level in the present high resolution configuration). Linear interpolation is used in the high-resolution model configuration and in LES to compute the 8.5-m potential temperature. The LES models (solid grey lines) simulate a significant cooling until 0200 LT, from about 277 K to about 264 K. The minimum potential temperature is reached around 0300 LT. At that time, the potential temperature vertical gradient between 15 and 25 m varies between 0.4 and 0.8 $\mathrm{K m^{-1}}$ among the LES (Fig. 1c). At 0100 LT, a low-level jet is well formed in all the LES (Fig. 1f). The altitude of its peak is similar in all LES, around 22 m, and its intensity ranges between 5 and 6 $\mathrm{m s^{-1}}$. The inertial rotation of the wind at 25 m is further emphasized in Fig. 1e. It is consistent with the theory (e.g. Blackadar, 1957) and representative of the observations collected at Dome C (Gallée et al., 2015).

CM6-HR-SHF severely underestimates the 8.5-m potential temperature cooling during the first half of the night (Fig. 1b, solid blue line). The minimum potential temperature reaches about 273 K at 0200 LT, about 8 K warmer than the LES corresponding value. The potential temperature vertical profile at 0300 LT (Fig. 1c) emphasizes that CM6-HR-SHF simulates a boundary layer, which is too thick and which stability is underestimated: the potential temperature vertical gradient is at least six times weaker (0.06 $\mathrm{K m^{-1}}$) than in the LES. Consistently, the CM6-HR-SHF low-level jet is too high, located near 55 m (Fig. 1f). Its intensity is significantly weaker than in all LES but one. The wind rotation is also strongly underestimated at 25 m (Fig. 1e).

In order to assess the model sensitivity to its internal turbulent parameters, we choose to synthesize the model behavior with four scalar metrics. The sensitivity to the choice and number of metrics is discussed in Section 6.1. The nocturnal cooling and boundary layer stability are quantified using the potential temperatures at 2 m and 8 m (referred to as θ_{2m} and θ_{8m} respectively); these two vertical levels allow to constrain the θ vertical gradient. These two metrics are computed at 0300 LT, when θ is minimum in the LES. The low-level jet structure is measured using the maximum of the supergeostrophic wind speed and the wind speed at 55 m (referred to as $\mathrm{jet_{MAX}}$ and w_{55m} respectively). The latter altitude corresponds to the level where the wind returns to its geostrophic value in the LES (it is also the altitude of the third level in the standard resolution model version CM6-LR). These two last metrics are taken at 0100 LT when the low-level jet is well established.

5.1.2 Defining the acceptable range of the turbulence free parameters

70 simulations (Wave 1) are run with the ARPEGE-Climat SCM (SCM-HR-SHF) for varying values of the seven parameters identified in Section 2.1 and following the experimental design proposed in Section 4. They are shown by the orange lines in Fig. 2. Although the majority of these simulations exhibits a too weak cooling, some of them capture the LES behavior, with both a correct θ vertical profile at 0300 LT and a correct overnight evolution of θ at 8 m. Concerning the wind, in most simulations, the low-level jet at 0100 LT is too high, so that the return to the geostrophic wind occurs above 100 m. The wind rotation at 25 m is poorly represented, with a too weak meridian component and a too strong zonal component. Reflecting these first conclusions, the metrics computed for each simulation are most of the time fairly far from those computed

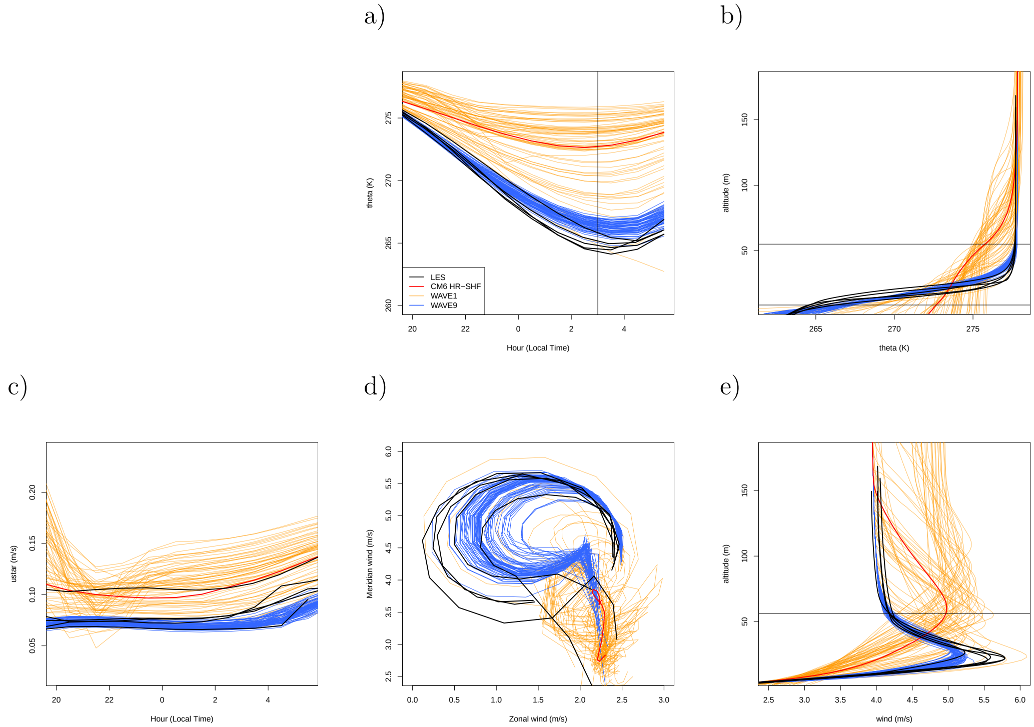


Figure 2. Same as Figure 1 but for the SCM-HR-SHF calibration experiment. The standard calibration simulation CM6-HR-SHF is indicated with the red line, the Wave 1 70 SCM simulations with the orange lines and the Wave 9 70 SCM simulations with the blue lines. Note that since it is prescribed, the sensible heat flux is not shown.

with the LES. Nevertheless, there exist a few simulations, thus a few sets of parameters, for which the chosen metrics have values close to those computed with LES. This suggests that, at least for each individual metric, there exists an appropriate calibration of ARPEGE-Climat. Based on this result, and for the sake of simplicity, in the following, we choose to explore the model performance considering no structural (or tolerance to) error (i.e. $e_m = 0$ in Equation 13). Such a choice may lead to overtuning (D. B. Williamson et al., 2017) and will need to be reconsidered in the context of the full model calibration. It will also be discussed in 6.2.

The metric values, computed for each simulation of this Wave 1, are used as a training sample to build a surrogate model for each of the four metrics. These surrogate models then allow us to explore the parameter space more exhaustively by estimating the value of each metric at as many new points as desired. In practice, the complete parameter space is resampled using a new latin hypercube of $\mathcal{O}(10^6)$ points and the surrogate models provide estimate (and uncertainty) of the four metrics for all these new points. As explained in Section 4, for each point sampled in the parameter space, the implausibility with respect to each metric is calculated, and the maximum implausibility over the four metrics (i.e. the most discriminating metric) is used to characterize the NROY space. Sampled points with an implausibility greater than our threshold set to 3 (see Section 4), are ruled out. After this first wave, the remaining space is 3.0% of the initial space, so a large part of the full parameter space is rejected. The NROY space obtained as a result of this first iteration (not shown) shows that **AE**, **KOZMIN** and **ZMAX** have a negligible influence on the results after this first iteration. The model behavior as a function of the other four parameters indicates a preference for values that significantly

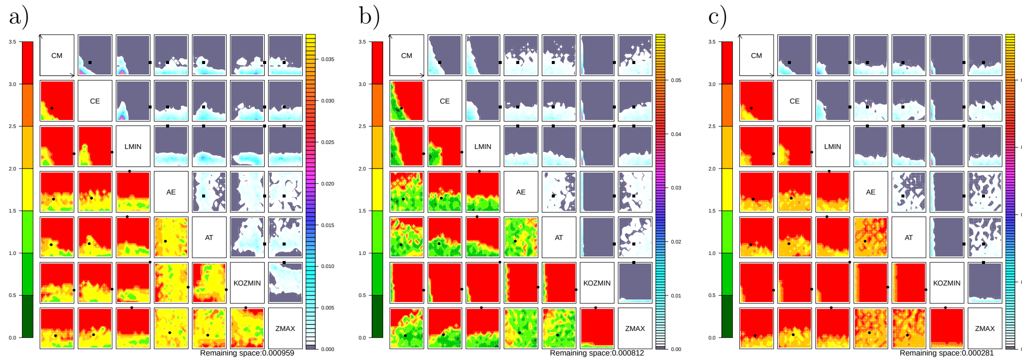


Figure 3. a) NROY of the SCM-HR-SHF, b) NROY of the SCM-LR-SHF and c) SCM-LR-TS calibration experiments at the end of Wave 9.

reduce the turbulent mixing. Table 2 details for each metric the performance of the simulations for this Wave 1 and for each of the following waves. For the first wave, all the metrics are very discriminating (more than two thirds of the simulations are incompatible with the LES reference values) except the one representing the jet intensity.

Seventy points in the NROY space estimated after this first wave are then sampled and the corresponding simulations are carried out. These new simulations form the Wave 2. The process is repeated until it is no longer possible to reduce the parameter space. Figure 2 illustrates the vertical profiles and time series of temperature and wind for the simulations of Waves 1 (orange lines) and 9 (blue lines) compared with the reference LES simulations. There is no clear further reduction of the NROY space from Wave 6 suggesting convergence of the results and that we can stop the iterations (see Table 2). From Wave 4, $\mathrm{jet}_{\mathrm{MAX}}$, $\mathrm{w}_{55\mathrm{m}}$ and the one on $\theta_{8\mathrm{m}}$ are no longer discriminating, i.e. for these metrics, almost no more part of the parameter space is ruled out. For θ_{2m}, the process evolves only slowly. At Wave 5, 10 simulations are still ruled out. From Wave 6 onwards, the number of rejected simulations varies between 2 and 4. The difference between two consecutive waves is then only due to the sampling of the remaining space. Figure 2b shows that the θ vertical profile at 0300LT (time used to compute θ-related metrics) is very close to that of the LES. However, it seems that there is still room for improvement close to the surface but performing an additional iteration does not reduce the near-surface spread of the SCM. The very low number of rejected simulations means that the same space is resampled at each iteration. Pushing the calibration exercise further would require more simulations but the results obtained seem satisfying. It should be noted here that at a given iteration, the simulations performed previously are not used to build emulators. Looking at the time series of θ at 8 m (Fig. 2a), there is also a very clear improvement in the results, which applies all along the simulation although we only use the metric at one given time (0300 LT). The results for the wind are good also. The Wave 9 simulations capture well the wind vertical profile at 0100 LT and the 25-m wind veering, falling within the LES spread (Fig. 2e and 2d). Note that the four instantaneous constraints put on the SCM simulations are sufficient to achieve model calibrations that perform well over the entire GABLS4 stable boundary layer, thus illustrating the consistency of the ARPEGE-Climat turbulence scheme physics.

Figure 3a shows the final NROY space (obtained after nine iterations). The upper right-hand side of the figure represents the two-dimensional density of the acceptable parameter space for each pair of parameters: for a given point in each two-dimensional space, the shading indicates the NROY space density in all other parameter dimensions. Grey color means that this density is close to zero, given the color scale used for the plot. The closer the color to yellow, the greater the density, thus indicating that this combination of parameters cannot be discarded yet. The lower left-hand side represents the minimum of implausibility. As shown by Fig. 3a, the iterative refocusing method selects parameter values that reduce turbulent mixing. The exchange coefficients are significantly reduced (mainly through **CM** and **LMIN**). The calibration also leads to small values of **CE** , which controls the dissipation length scale: the TKE dissipation is increased and thus the TKE and the turbulent mixing are reduced. These results are consistent with previous works that generally indicate a too strong turbulent mixing in numerical weather and climate models (e.g. Sandu et al., 2013; Beljaars & Viterbo, 1998). The case of **LMIN** is also consistent with the work of Vignon, Hourdin, et al. (2017) which showed the importance of removing most of the bounds in the turbulence parameterization used in global models to better represent stable boundary layers. The iterative refocussing highlights the high sensitivity of the model results to this parameter (Fig. 3a), and illustrates the difficulty to calibrate the model for stable boundary layers with **LMIN** values greater than about 4-5 m. The influence of **AT**, which controls the turbulent flux of θ, is significant but not as decisive as that of the three previous parameters. The value currently used in ARPEGE-Climat is appropriate. The results are not sensitive to **AE**, so that there is a priori no need to change its current value. This parameter influences the vertical diffusion of TKE. It therefore appears that this term is negligible compared to the other terms in the equation for the evolution of TKE in the case of very stable boundary layer of GABLS4 (not shown). Finally, **KOZMIN** and **ZMAX** which directly limit the turbulent fluxes, are not relevant to adjust this high-resolution SCM configuration. This is expected given the bound formulation (Eq. 6) which tends to zero as the vertical grid spacing goes to zero. These parameters are more critical for the model standard resolution (see Sections 5.2 and 5.3).

As a preliminary conclusion, the iterative refocussing demonstrates that the ARPEGE-Climat turbulence scheme contains the required physics to represent the GABLS4 strongly-stable boundary layer, at least for resolution of $\mathcal{O}$(1m). Besides, if the ARPEGE-Climat standard calibration is not appropriate (i.e. the free parameter standard values are not retained in the NROY space), the standard values of **CM** and **CE** are reasonable (i.e. very close or within the NROY space). It is rather the lower bound **LMIN** on the mixing length that is the most discriminating, a conclusion similar to the one obtained by Vignon, Hourdin, et al. (2017).

5.2 Standard-resolution SCM configuration forced by surface sensible heat flux (SCM-LR-SHF)

The behavior of the standard-resolution SCM configuration, forced by the LES ensemble-mean surface sensible heat flux (CM6-LR-SHF), is summarized in Fig. 1 (dashed blue line). In order to compare the low-resolution SCM results with those of the LES, the latter are regridded onto the vertical grid of CM6-LR. The CM6-LR-SHF overnight cooling is weak, and slightly weaker than in CM6-HR-SHF, thereby indicating sensitivity of the model results to vertical resolution. The minimum potential temperature is reached earlier than in the LES (as for CM6-HR-SHF) and is 8 K warmer. The low-level jet is weakly marked. The altitude of the maximum wind speed is too high (55 m) and the wind speed remains too strong above 55 m, where it should be geostrophic. The wind rotation is slightly better represented than in CM6-HR-SHF but still too weak. The CM6-LR-SHF behavior is thus broadly similar to its high-resolution counterpart, and consistent with an overestimated turbulent mixing. We now investigate whether this behavior is intrinsic to the parameterization (for this standard resolution), or results from a poor calibration for stable boundary layers. Similar scalar metrics to those used in Section 5.1 for CM6-HR-SHF are chosen to constrain the θ vertical gradient and the low-level jet structure. The associated altitudes are slightly adapted to be consistent with the SCM standard vertical resolution. The θ gradient is characterized by θ at the first and third model levels at 0300 LT, namely 8 m ($\theta_{8\mathrm{m}}$) and 55 m ($\theta_{55\mathrm{m}}$). Because of the rather large uncertainty of θ at the second SCM level (29 m) in the LES (about 5 K),

Table 2. Evolution of the different metrics over the successive waves (lines) for the SCM-HR-SHF tuning experiment. For each metric, the first column indicates the spread of the metric among the 70 simulations of each wave(min – max). The second column gives for each wave, the number of simulations rejected because of an implausibility greater than 3.

	$\theta_{2\mathrm{m}}$		$\theta_{8\mathrm{m}}$		$\mathrm{jet}_{\mathrm{MAX}}$		$\mathrm{w}_{55\mathrm{m}}$	
WAVE	Metric	I>3	Metric	I>3	Metric	I>3	Metric	I>3
1	260.3 – 277.9	58	266.3 – 276.5	64	4.8 – 5.7	2	4.1 – 5.3	46
2	252 – 269.4	10	263.2 – 270.7	3	4.7 – 5.8	2	4.1 – 5.6	31
3	255.2 – 265.2	10	262.3 – 268.7	5	5.0 – 5.9	0	4.1 – 4.5	6
4	259.1 – 266.1	7	264.4 – 267.7	4	5.0 – 5.5	0	4.1 – 4.3	0
5	258.9 – 266.5	10	264.4 – 267.3	0	5.0 – 5.4	0	4.1 – 4.5	1
6	261 – 265.1	4	264.7 – 267.4	0	5.0 – 5.4	0	4.1 – 4.3	2
7	261.5 – 264.9	3	264.9 – 267.4	2	5.0 – 5.5	0	4.1 – 4.3	2
8	261.7 – 264.8	4	265.4 – 266.9	0	5.0 – 5.4	0	4.1 – 4.3	0
9	261.7 – 265	2	265.7 – 267	0	5.1 – 5.3	0	4.1 – 4.3	0
LES	263.4		265.0		5.6		4.2	

choosing this metric would have been less efficient in reducing the free parameter space. As in the high-resolution experiment, the jet structure is summarized by the wind speed at 29 m (second SCM level and altitude of the wind maximum – $\mathrm{w}_{29\mathrm{m}}$) and at 55 m (third SCM level – $\mathrm{w}_{55\mathrm{m}}$).

The Wave 1 simulations present a large spread, with only a few simulations getting close to the LES (Fig. 4, orange lines). Note here that before introducing the parameters **KOZMIN** and **ZMAX** in the tuning exercise, this variety of behavior was not observed and none of the simulations perform well (not shown). Indeed, with the default setting, this minimum bound for the mixing coefficients was systematically reached and any modifications of the other parameters had very little effect as hidden by the minimum bound. The introduction of **KOZMIN** and **ZMAX** was crucial for this statistical tuning experiment in standard resolution. Table 3 shows the evolution of the metrics for eight waves. From Wave 3, the results are already satisfying for most of the metrics. Only $\theta_{55\mathrm{m}}$ needs a few more waves to achieve reasonable values. The remaining (NROY) space is finally slightly more than 0.1% of the initial space (Fig. 3b).

The θ vertical profiles at 0300 LT of all the Wave 9 simulations are very close to the LES (Fig. 4b, blue lines). The time evolution of the 8-m potential temperature still presents some dispersion, which is consistent with the LES uncertainty. Wave 9 simulations also capture well the sharpened wind speed vertical structure at 0100 LT, and much better than CM6-LR (red line). There is still some significant spread in the jet intensity, but again, it reflects the LES discrepancies (Fig. 4e). All the simulations reproduce well the 25 m wind rotation (Fig. 4d).

Figure 3b presents the remaining space after nine waves. In contrast with the HR version, the **KOZMIN** parameter appears critical. The history matching with iterative refocussing thus shows that the SCM can only behave well over the GABLS4 case for very low values of **KOZMIN**. Given that this parameter is used to maintain significant turbulent fluxes within the first model levels of the model, it was rather expected. For the other parameters, the results are similar to those with the HR configuration: param-

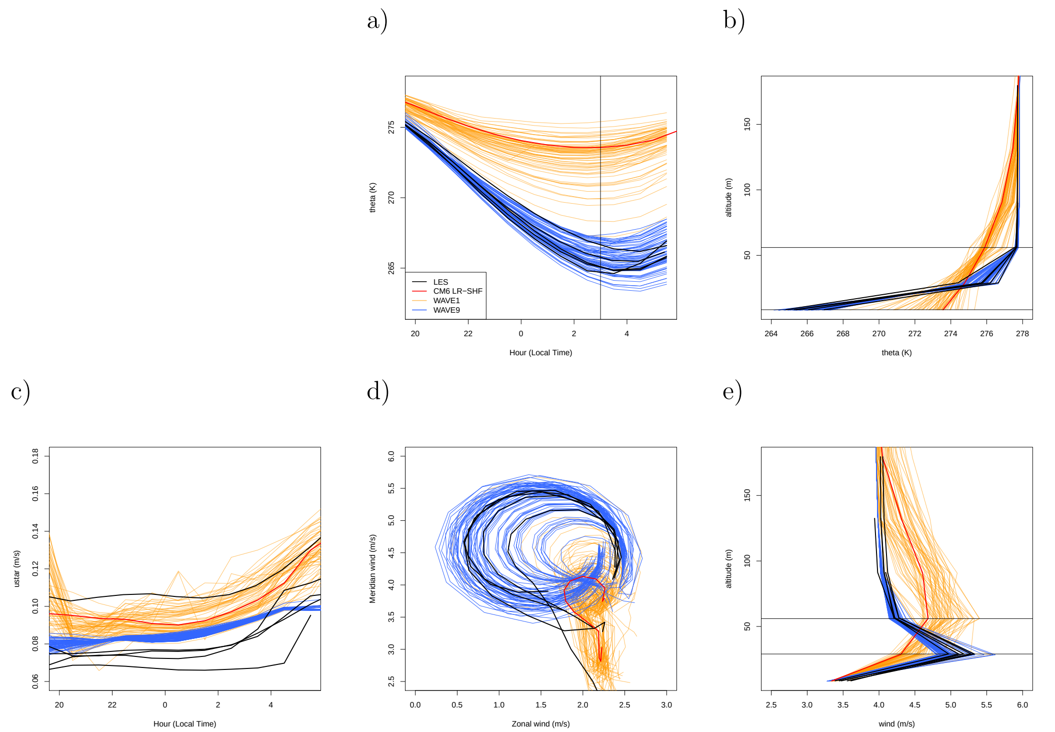


Figure 4. Same as Figure 2 but for the SCM-LR-SHF calibration experiment. The standard calibration simulation is CM6-LR-SHF (red line).

eter values that allow weak turbulent mixing are found the most suitable for adjusting the turbulence parameterization. Note that the CM6 values (black dots) of **KOZMIN** and **LMIN** are well beyond their acceptable range (already eliminated at Wave 1). As for the high-resolution configuration, the SCM behavior is also significantly sensitive to **CE** and **CM** with acceptable values leading to decrease of the turbulent mixing but the CM6 default values are not ruled out. **AE** and **AT** have no significant impact. To summarize, the CM6 turbulence parameterization is able to capture the GABLS4 stable boundary layer as described by the reference LES, even with the standard resolution SCM configuration. Such an acceptable SCM behavior requires to almost remove the two bounds on the turbulent mixing induced by **LMIN** and **KOZMIN**, while the other free parameters of the parameterization can be kept close to their original values, similar to those proposed in Cuxart et al. (2000) and Cheng et al. (2002).

5.3 Standard-resolution SCM configuration forced by surface temperature (SCM-LR-TS)

The iterative refocussing approach is now applied to a less constrained configuration, in which the surface boundary condition is provided by the surface temperature (refered to as CM6-LR-TS), similarly to the LES setup, and as proposed in the GABLS4 intercomparison framework. This adds a degree of freedom as the model now uses its own formulation to compute the surface fluxes.

CM6-LR-TS behaves similarly to CM6-LR-SHF with an underestimated cooling of the atmospheric low levels and a too weak low-level jet (Fig. 1). A notable difference is that the model continues to cool for a longer period (until around 0600 LT). The wind rotation at 25 m remains weakly captured, as in CM6-LR-SHF.

Table 3. Evolution of the different metrics over the successive waves for the SCM-LR-SHF tuning experiment. For each metric, the first column indicates the spread of the metric among the 70 simulations of each wave. The second column gives for each wave, the number of simulations rejected because of an implausibility greater than 3.

	θ_{8m}		θ_{55m}		w_{29m}		w_{55m}	
WAVE	Metric	I>3	Metric	I>3	Metric	I>3	Metric	I>3
1	267.9 – 275.7	67	275.3 – 277.2	65	4.6 – 5.5	4	4.4 – 5.4	59
2	263.4 – 271.5	8	276.5 – 277.9	24	4.7 – 6.0	1	4.1 – 4.7	9
3	262.9 – 269.0	1	277.4 – 277.8	2	4.8 – 5.3	0	4.1 – 4.6	3
4	263.2 – 267.8	0	277.4 – 277.8	4	4.8 – 5.4	0	4.1 – 4.6	1
5	263.2 – 268.8	1	277.4 – 277.8	7	4.8 – 5.5	0	4.1 – 4.3	0
6	263.9 – 267.5	0	277.5 – 277.8	2	4.7 – 5.8	1	4.1 – 4.3	0
7	264.4 – 268.4	0	277.5 – 277.8	0	4.8 – 5.7	1	4.1 – 4.4	1
8	263.6 – 268.0	0	277.5 – 277.8	4	4.8 – 5.5	0	4.1 – 4.3	0
LES	265.7		277.7		5.2		4.2	

The iterative refocussing applied on the present configuration makes use of the same four metrics as for CM6-LR-SHF. Table 4 present the evolution of the metrics and associated implausibility over the successive waves. It only takes two iterations for the procedure convergence with the metrics related to the jet structure. Metrics characterizing the θ vertical profile require three more iterations to ensure convergence. Overall, Wave 8 SCM simulations provide improved and satisfying results over the GABLS4 stable boundary layer (Fig. 5). Nevertheless, it can be noticed that the 8-m potential temperature at 0300 LT is systematically overestimated by 1.1 to 1.3 K (Table 4, Fig. 5c), mostly because its minimum is reached about 2 hours later (Fig. 5b). Regarding the wind vertical structure, the wave 9 simulations are also able to capture the wind vertical structure of LES references and its overnight rotation, within the LES ensemble results (Fig. 5f and 5e).

Figure 3c presents the free parameter remaining space after eight waves. The conclusions are very close to those made with the configuration with prescribed surface sensible heat flux. The role of **KOZMIN**, **LMIN**, **CM** and **CE** is again emphasized and values leading to low turbulent mixing are retained. However, the experiment suggests that the CM6 values for **CM** and **CE** are slightly too high. Thus, while there still exist acceptable sets of parameters guaranteeing a good performance of the SCM, the addition of a partial surface coupling further constrain this parameters, possibly to compensate the surface flux parameterization errors.

6 Discussion

In this section, we discuss in more detail two features of the calibration statistical approach, that are rather subjective. The selection of metrics is first emphasized as a key step that deserves some caution. We discuss the selection we made in Section 2 and analyze the respective role of each selected metrics. Second, we investigate more in depth the convergence criteria, and how we have tackled it. Finally, the SCM computationally-cheap approach provides the opportunity to perform a large simulation ensemble, which can serve as a basis for the evaluation of the full statistical framework.

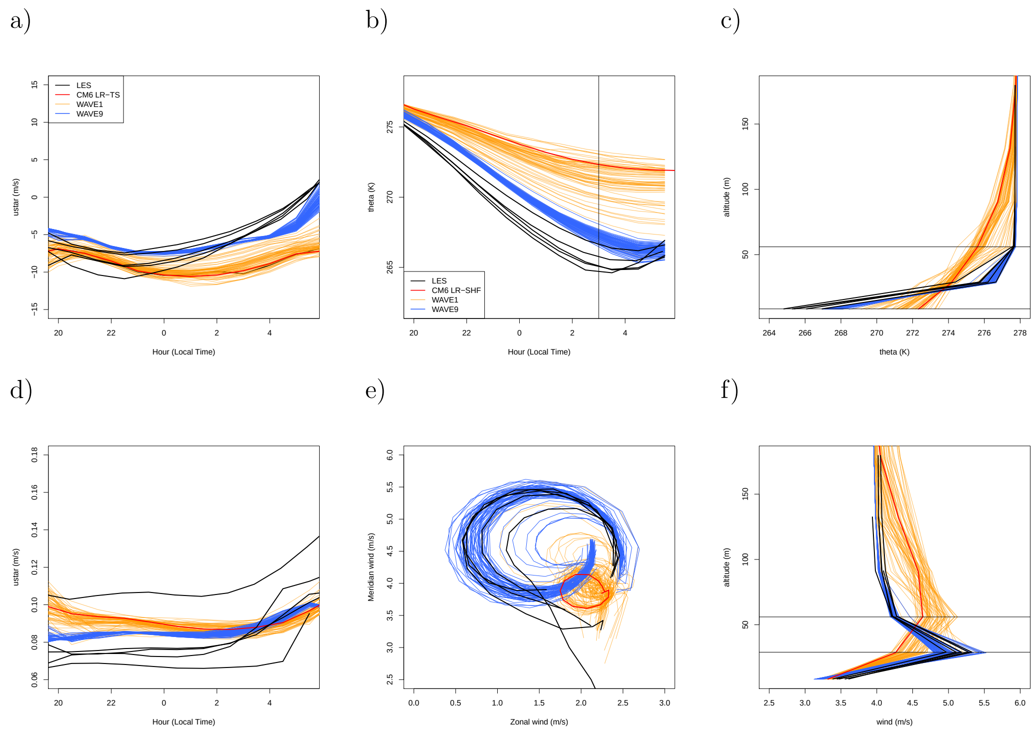


Figure 5. Same as Figure 2 but for the SCM-LR-TS calibration experiment. The standard calibration simulation is CM6-LR-TS (red line).

6.1 Choice of metrics

The choice of metrics in this type of framework is crucial, yet it is difficult to draw up a definitive list in advance. In this section, we illustrate the metric selection procedure in the case of the SCM-LR-SHF experiment and give some guidelines. It might be tempting to choose a large number of metrics to precisely control the model. But for efficiency, especially for the analysis of the results of each wave, the number of metrics should be limited. In particular, redundant metrics should be avoided. Note also that if, in the course of the calibration experiment, biases that were not accounted for emerge, new metrics can be added on the fly at the beginning of a wave.

The first step consists in conducting an evaluation of the model behavior in order to identify the main model biases (and especially those we care about) and then to design the appropriate metrics to quantify them. We have already seen in Section 5.2 that the two main identified biases are a weak nocturnal cooling and an insufficiently marked low-level jet structure. Analysis of the potential temperature profile at 0300 LT (Fig. 4b), the time at which the minimum of θ at 8 m, the first CM6-LR level, is reached in the LES simulations (Fig. 4a) shows that the θ vertical gradient in the nocturnal boundary layer is too weak, with a warm bias (+8 K) at the first level and a cold bias (about -2 K) at the third level. The values of θ at these two levels (respectively $\theta_{8\mathrm{m}}$ and $\theta_{55\mathrm{m}}$) are therefore used as metrics to control the model thermodynamics. Concerning the low-level jet structure at 0100 LT (the time at which it is the sharpest in the LES, Fig. 4e), the main features to be considered are the altitude and the intensity of the wind maximum (which are respectively too high and too low in CM6-LR), and the thickness of the jet, which can be represented by the altitude of the wind return to its geostrophic value. The first two features can be summarized by a single metric, namely the wind intensity at the sec-

Table 4. Evolution of the different metrics over the successive waves for the SCM-LR-TS tuning experiment. For each metric, the first column indicates the spread of the metric among the 70 simulations of each wave. The second column gives for each wave, the number of simulations rejected because of an implausibility greater than 3.

	$\theta_{8\mathrm{m}}$		$\theta_{55\mathrm{m}}$		$\mathrm{w}_{29\mathrm{m}}$		$\mathrm{w}_{55\mathrm{m}}$	
WAVE	Metric	I>3	Metric	I>3	Metric	I>3	Metric	I>3
1	267.8 – 273.4	69	274.9 – 277.6	62	4.5 – 5.2	9	4.5 – 5.1	61
2	266.3 – 270.6	28	277.0 – 277.9	13	4.7 – 5.4	0	4.1 – 4.6	5
3	266.1 – 268.5	4	277.5 – 277.8	19	4.7 – 5.6	0	4.1 – 4.4	1
4	266.9 – 268.3	0	277.6 – 277.8	7	4.8 – 5.6	0	4.1 – 4.3	1
5	267.0 – 268.1	0	277.6 – 277.8	1	4.7 – 5.8	1	4.1 – 4.4	0
6	266.4 – 268.2	0	277.6 – 277.8	1	4.8 – 5.8	1	4.2 – 4.4	1
7	267.2 – 268.2	0	277.6 – 277.8	1	4.8 – 5.4	0	4.1 – 4.3	0
8	266.9 – 268.1	0	277.6 – 277.8	0	4.8 – 5.3	0	4.2 – 4.3	0
LES	265.7		277.7		5.2		4.2	

ond model level ($\mathrm{w}_{29\mathrm{m}}$), which corresponds to the wind maximum in the regridded LES. The third feature can be captured by the wind intensity at the third model level ($\mathrm{w}_{55\mathrm{m}}$). Wind intensity at the fourth level could also be chosen and this choice leads to similar results (not shown).

As explained before, the framework used here takes into account the different sources of uncertainty through the notion of implausibility. Note that the lower the uncertainty on the reference metrics the more constraining the metric: particular attention was devoted to the associated reference uncertainty, and metrics with weak reference uncertainty were preferred. When it is significant, at least compared to the Wave 1 simulation errors, it may rapidly dominate the implausibility and the framework will only weakly constrain the model behavior. For the four metrics defined on the basis of model biases, the uncertainty of the LES regarding these metrics is low in front of the biases of Wave 1 simulations (for $\theta_{8\mathrm{m}}$ and $\mathrm{w}_{29\mathrm{m}}$) or almost negligible (for $\theta_{55\mathrm{m}}$ and $\mathrm{w}_{55\mathrm{m}}$).

To understand how each of the selected metrics constrains the model, four tuning experiments are carried out in which only one of the metrics is considered at a time. For each metric, the evolution of the remaining space throughout the different iterations is presented in Table 5. The $\mathrm{w}_{29\mathrm{m}}$ metric is much less discriminating than the other three. 18% of the original space is compatible with the reference after 9 waves, as opposed to around 0.5 to 1% for the other metrics. The LES uncertainty for this metric explains part of this result. The importance of the reference uncertainty relative to the bias is quantified as the ratio between the two quantities (Table 5). It is thus shown that consideration of this ratio is not sufficient to presuppose the importance of each metric. The results of these experiments are presented in Fig. 6. θ-related metrics and the 55-m wind speed rule out around 99% of the parameter space. θ-related metrics ($\theta_{8\mathrm{m}}$ and $\theta_{55\mathrm{m}}$) lead to simulations with a proper nocturnal cooling (a little more pronounced when considering only $\theta_{55\mathrm{m}}$). They also show a low-level jet significantly better than CM6-LR with a wind maximum at the right altitude. However, the return to geostrophic wind is weakly constrained and for many simulations it occurs at a too high altitude, especially for those resulting from the calibration based on $\theta_{8\mathrm{m}}$ only. The $\mathrm{w}_{55\mathrm{m}}$ metric leads to a jet with a structure at 0100 LT very close to the LES, but with a too weak nocturnal cooling, even if the improvement compared to CM6-LR is still significant. These different metrics reduce the space of the parameters differently (Fig. 7). If high values for **LMIN** and **KOZMIN** are systematically eliminated for each of the three metrics, they differ for the tuning of **CM**, **CE** and **AT**. w_{55m} strongly constrains the sum of **CM** and **CE** but has no influence on **AT** whereas thermodynamic metrics have a less marked sensitivity to **CM** and **CE** but significantly constrain **AT** by eliminating the highest values. The weaker sensitivity of θ_{8m} and θ_{55m} to **CM** is explained as the turbulent heat flux depends on the product of **CM** by **AT** (see Section 2.1). The w_{29m} metric poorly constrains the nocturnal cooling and the altitude of the return to geostrophic wind. The usefulness of this particular metric may be questioned. An experiment conducted without w_{29m} however slightly degrades the low-level jet structure (not shown) and leads to a slower reduction of the parameter space.

In this study we use simple scalar metrics at given times. The time at which the metrics are calculated is chosen according to the maximum intensity of the phenomena they capture. Thus the wind-related metrics are computed at 0100 LT which is the time when the jet is the sharpest. The θ-related metrics are computed at the time of the strongest cooling (0300 LT). More complex (vector) metrics could be used, such as vertical profiles or time series (e.g., D. B. Williamson et al., 2017; Salter et al., 2019), but the results presented in Section 5 emphasize that these very simple metrics are sufficient to constrain significantly the model behavior over the entire duration of the simulation.

We would like to finally stress here that the choices made in section 5 result from a trial-and-error empirical approach, which we advocate for. To help to analyze the results, a small number of metrics were selected from a much larger list of potential metrics. It is certainly possible to use a greater number of metrics, in practice all those that the modeller finds relevant. Nevertheless, we emphasize hereafter a few guidelines on this process of metric selection:

1. The analysis of the default configuration but also of the first wave simulations is an important step, which helps to identify model biases or appropriately-simulated features and build metrics to quantify them.
2. As the method takes into account the different sources of uncertainty, particular attention should be paid to the uncertainty in the LES and to ensure that it does not dominate the biases identified above. Besides, the first wave spread might also give indications about the discriminatory ability of a given metric.
3. Metrics do not need to be taken all at once starting at Wave 1 and sometimes it is preferable to start with few discriminant metrics (with priority metrics being defined by the modeler expertise) in order to ease the construction of the emulators during the following waves as over a smaller parameter space. This ensures by the flexibility of the tool that allows to add new metrics over the waves if necessary.
4. A preference is given to the use of simple metrics, easier to compute and interpret.

6.2 Iterative refocussing convergence and its link with the sources of uncertainty

Deciding when to stop the iterations is not trivial. For the different experiments carried out in this study, we decided to stop the iterations when the size of the NROY space no longer decreases as we perform more waves (e.g., Table 6). Using the SCM-LR-SHF experiment, we illustrate here the NROY convergence and analyze the evolution of the different uncertainty sources that we need to take into account to monitor this convergence.

In the SCM-LR-SHF experiment, the NROY space size has converged from about the fourth wave. The NROY space estimated at the end of Wave N is defined using the

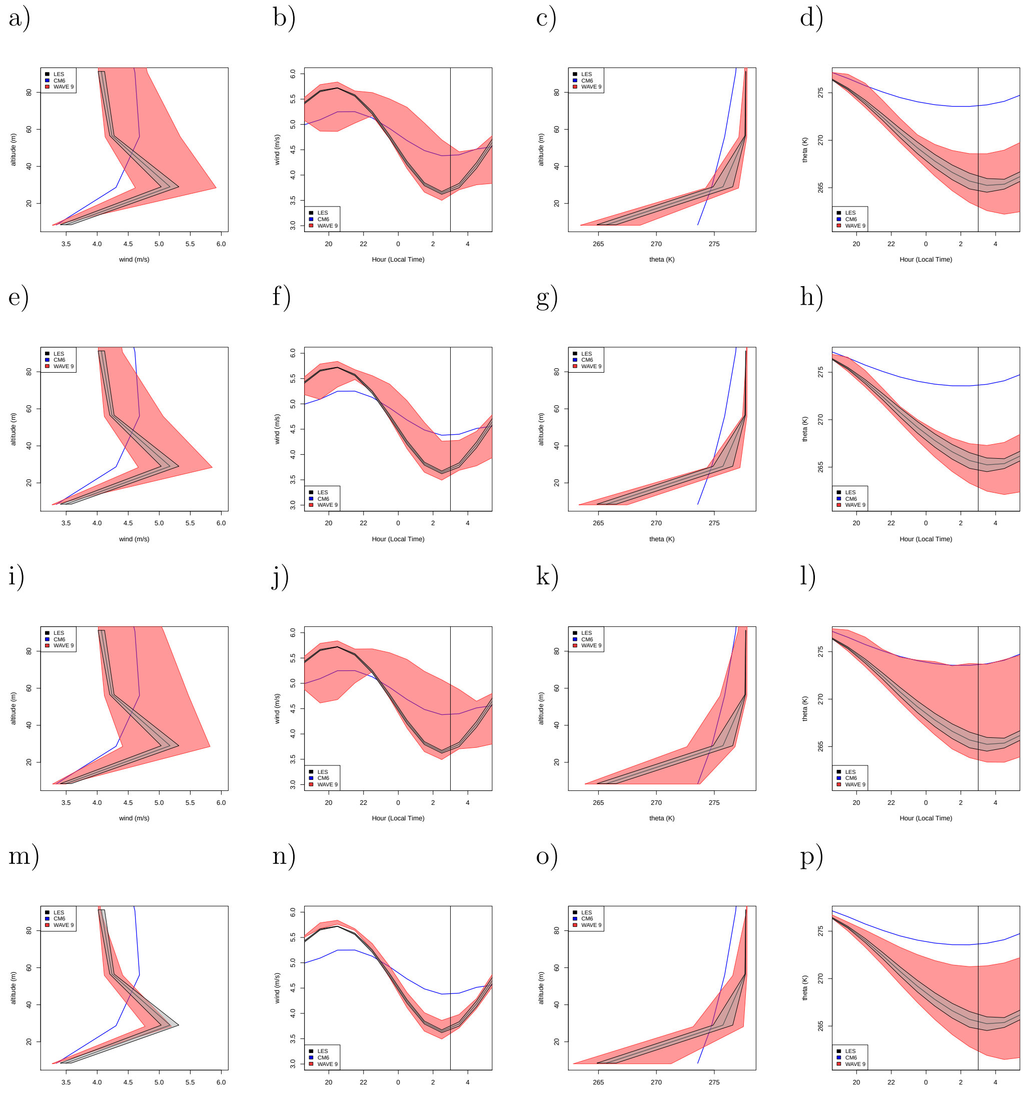


Figure 6. Results of the SCM-LR-SHF calibration experiment, but considering only (a, e, i and m) the $\theta_{8\mathrm{m}}$ metric, (b, f, j and n) the $\theta_{55\mathrm{m}}$ metric, (c, g, k and o) the $\mathrm{w}_{29\mathrm{m}}$ metric and (d, h, l and p) the $\mathrm{w}_{55\mathrm{m}}$. The first column corresponds to the wind profile at 0100 LT (m s^{-1}), the second column to the wind at 55 m (m s^{-1}), the third column to the potential temperature profile at 0300 LT (K) and the fourth column to the 8-m potential temperature. On each panel, the blue line indicates the CM6-LR-SHF simulation, the black line the ensemble mean LES with plus or minus one standard deviation as the grey shading and the red shading the Wave 9 simulation envelope.

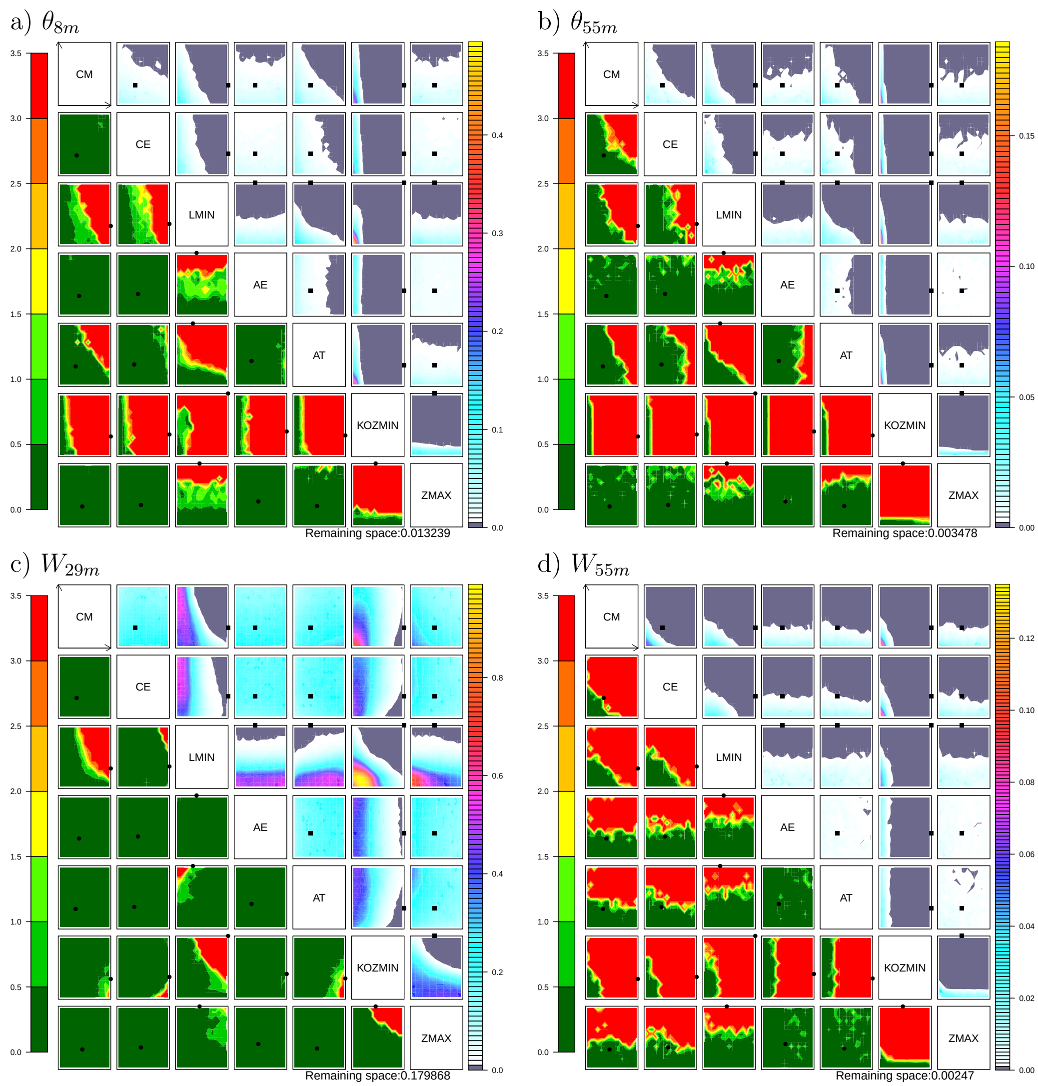


Figure 7. NROY space obtained after Wave 9 of the SCM-LR-SHF calibration experiment, but considering only (a) the $\theta_{8\mathrm{m}}$ metric, (b) the $\theta_{55\mathrm{m}}$ metric, (c) the $\mathrm{w}_{29\mathrm{m}}$ metric and (d) the $\mathrm{w}_{55\mathrm{m}}$ metric.

Table 5. Results of a tuning of ARPEGE-Climat in LR-SHF configuration using one metric at a time. The first line gives the value of each metric computed with CM6-LR-SHF, the second line gives the value of each reference metric computed as the mean of the LES ensemble and the third line gives the LES uncertainty computed as the variance of the LES ensemble. The following lines gives the evolution of the remaining NROY space over waves.

Experiment	θ_{8m}	θ_{55m}	w_{29m}	w_{55m}
CM6	273.5	275.5	4.7	4.7
Obs	265.7	277.7	5.2	4.2
Err Obs	0.69	$7.5\ 10^{-4}$	0.02	$1.2\ 10^{-3}$
NROY 1	2.6 %	2.8 %	25.4 %	6.0 %
NROY 2	1.9 %	0.7 %	22.4 %	0.5 %
NROY 3	1.8 %	0.6 %	21.2 %	0.4 %
NROY 4	1.6 %	0.5 %	20.8 %	0.3 %
NROY 5	1.4 %	0.5 %	19.2 %	0.3 %
NROY 6	1.4 %	0.4 %	19.1 %	0.3 %
NROY 7	1.4 %	0.4 %	18.7 %	0.3 %
NROY 8	1.4 %	0.4 %	18.2 %	0.3 %
NROY 9	1.3 %	0.4 %	18.0 %	0.3 %

implausibility which depends on the metric value estimated using the surrogate model, the different sources of uncertainty and the chosen threshold. Figure 8 presents the evolution through the different waves of the implausibility distribution, computed for each of the four metrics considered here, as well as the different quantities involved in its computation. For Wave N, the implausibility is computed using the points in the Wave N-1 NROY space. From Wave 3, and for each metric, almost all points in the NROY space have an implausibility lower than 3 (the chosen cutoff, cf. Section 4) despite the reduced emulator uncertainty within the NROY space compared to the previous wave.

As the NROY space size reduces during the successive waves, the 70 SCM runs sample a much smaller space, thus leading to improved surrogate models within the NROY space, with reduced uncertainty. This is particularly the case for the metrics $\theta_{55m}$and w_{55m}, for which from Wave 2-3 the surrogate model uncertainty falls mostly below or is of the same order of magnitude as the reference uncertainty (Fig. 8f and 8h). For θ_{8m} the surrogate model uncertainty is also reduced after Wave 1 (Fig. 8e), but there is not a strong decrease of it. It is already significantly lower than the reference uncertainty and thus does not play much in the implausibility for this metrics. Besides, as the NROY space does not anymore reduce dramatically from Wave 2 onwards, the SCM runs mostly sample the same space and thus no surrogate model improvement is expected.

The different metrics estimated by the emulators also converge rapidly and from Wave 3 (Wave 5 for w_{55m}) onwards (Fig. 8a-d), only a few outliers fall outside the range of 3 standard deviations around the reference. In other words, from Wave 5 onwards, the simulations are consistent with the LES given our four metrics, and taking into account the LES uncertainty. To go further in the calibration process, the threshold initially taken to 3 could be changed to a lower value, especially because the emulator uncertainty is not anymore the dominating uncertainty.

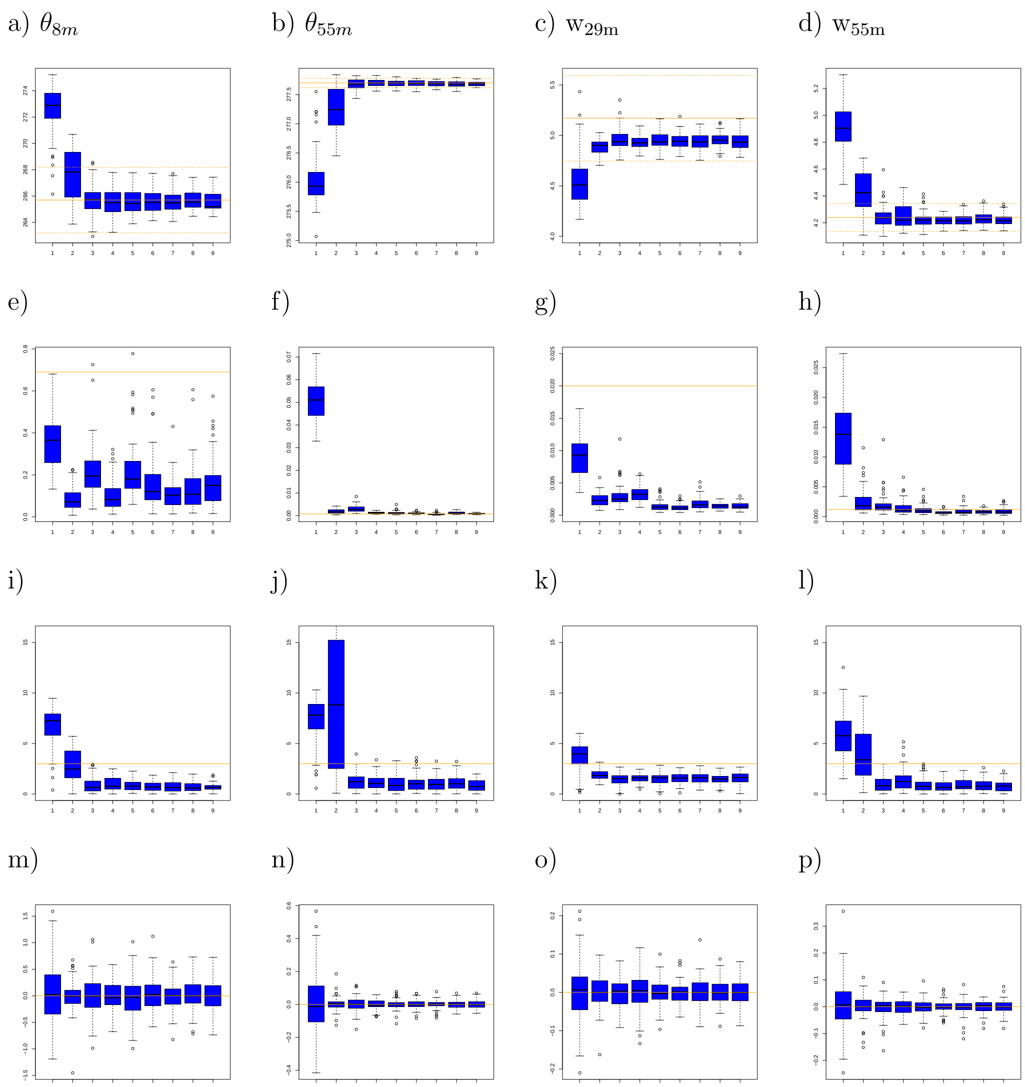


Figure 8. Evolution across the SCM-LR-SHF 9 waves of the distribution of various variables involved in the implausibility computation: (a, b, c and d) metrics values with the LES ensemble mean (orange solid line) and plus or minus three standard deviation (dashed orange lines), (e, f, g and h) surrogate model variance with the LES ensemble variance (orange solid line), (i, j, k and l) implausibility with the cutoff of 3 (orange solid line). The first column correspond to the $\theta_{8\mathrm{m}}$ metric, the second column to the $\theta_{55\mathrm{m}}$ metric, the third column to the $\mathrm{w}_{29\mathrm{m}}$ metric and the fourth column to the $\mathrm{w}_{55\mathrm{m}}$ metric. The distributions are computed using points sampling the NROY space obtained at the end of the previous wave (input space for Wave 1).

The analysis of the metric convergence also emphasizes the question of discrepancy. If for $\theta_{8\mathrm{m}}$, $\theta_{55\mathrm{m}}$ and $\mathrm{w}_{55\mathrm{m}}$, the SCM simulations converge to the mean reference, this is not the case for $\mathrm{w}_{29\mathrm{m}}$. For this metric, the LES ensemble mean is about 5.2 m s^{-1}, while the SCMs struggle to exceed 5.0 m s^{-1}. If a non-null discrepancy is not necessary to avoid an empty space, the latter point advocates for taking a structural error of the model at least equal to 0.2. Our choice for no tolerance to error seems reasonable for exploring a single case study and only a few metrics. It is however not recommended in a more comprehensive calibration of the full model as it is likely to end up with overtuning (D. B. Williamson et al., 2017). When no information can help in suggesting or quantifying this discrepancy, tests during the first wave (when computing the implausibilities and identifying the NROY space) are likely to provide some upper bound on it. The following waves will possibly help to reduce it as the modeler gains a better quantification of his model behavior and can further confront what he wants and what the model can really do.

Finally, we want to stress here the importance of analyzing the comparative importance of the different sources of uncertainty when the procedure seems to converge (i.e. when the space of the parameters no longer reduces significantly from one wave to the next) before reducing the threshold for example.

Table 6. Evolution of the remaining space for the three tuning experiments

Experiment	HR-SHF	LR-SHF	LR-TS
NROY 1	3.0 %	0.59 %	0.21 %
NROY 2	0.97 %	0.17 %	0.12 %
NROY 3	0.52 %	0.14 %	0.07 %
NROY 4	0.39 %	0.12 %	0.05 %
NROY 5	0.27 %	0.11 %	0.04 %
NROY 6	0.21 %	0.10 %	0.04 %
NROY 7	0.16 %	0.09 %	0.03 %
NROY 8	0.14 %	0.08 %	0.03 %
NROY 9	0.12 %	0.08 %	0.03 %

6.3 Evaluation of the statistical framework: comparison with a 100% SCM approach

As SCM low-resolution simulations are computationally cheap, it is possible to perform an independent large ensemble of simulations to assess the overall framework used in Section 4, in particular the quality of the GP-based surrogate models. In the present section, we focus on the SCM-SHF-LR setup for which 10^4 simulations are performed. The parameters are sampled according to a conventional Latin hypercube of the input space. The implausibility of each of these 10^4 simulations is evaluated. As there is no emulator used in this experience, Equation 13 reduces to

$$I_f(\boldsymbol{\lambda}) = \frac{|r_f - f(\boldsymbol{\lambda})|}{\sqrt{\sigma_{r,f}^2 + \sigma_{d,f}^2}} \tag{14}$$

where $f(\boldsymbol{\lambda})$ is the value of the metric f for the set of parameters $\boldsymbol{\lambda}$, directly computed from the SCM simulations. As in the previous tuning experience using iterative refocussing

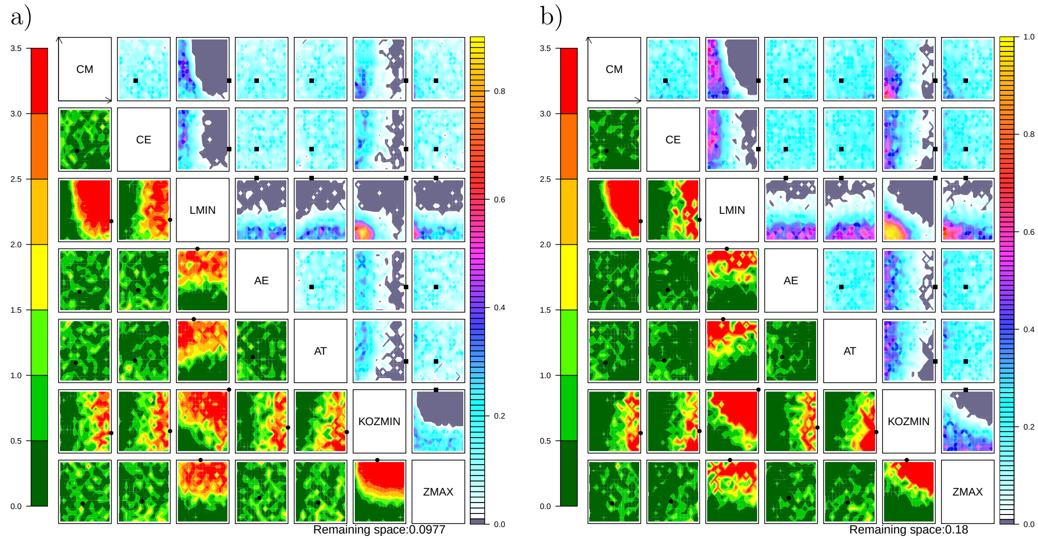


Figure 9. NROY space (a) computed from the 10^4 SCM-LR-SHF simulations and (b) computed from the Wave 9 history matching with iterative refocussing framework applied on SCM-LR-SHF, considering only the w_{29m} metric. See text for details.

method and in order to compare results, the structural error $\sigma^2_{d,f}$ is taken to zero. This formulation is used to estimate the NROY space, directly from the 10^4 SCM ensemble. Note that to appropriately characterize the NROY space, it is necessary that it contains enough points, at least a few hundreds. The NROY space obtained using the four metrics in Section 5.2 is about 0.1% of the initial space. It has been numerically evaluated using the implausibility computed over 10^6 points, so that it contains in the end around 10^3 points. With only 10^4 points in the initial space, only an order of 10 points is expected to remain, which is clearly not sufficient to fully characterize the NROY and make its estimate not too much sensitive to the input space sampling. This further emphasizes the relevance of using surrogate models, even in this SCM framework. Therefore, the statistical framework is evaluated using only the metric w_{29m}, as it keeps a sufficient fraction of the initial space, namely about 18% using emulators (thus about 1,800 points). In parallel, a NROY space is estimated using the emulators built in 5.2 on the same 10^4 points in the parameter space. The use of the same points for both estimates avoids sampling effects in the comparison.

In experiments using emulators, the threshold for implausibility is set at 3. This choice is deliberately conservative and reduces the risk of ruling out a set of parameters that would in fact leads to results consistent with the reference. The threshold of 3 follows the $3-\sigma$ rule of Pukelsheim (1994) which states that 95% of any unimodal distribution lies in the range of $\pm 3\sigma$, σ being the standard deviation of this distribution. To compare the two methods, keeping this threshold of 3 for the direct approach is not relevant, as in our full SCM approach, there is only one source of uncertainty. Assuming a Gaussian distribution for the reference, 95% of the distribution is in the range of $\pm 1.96\sigma$ around the mean. We thus reduce the threshold to 1.96 in the full SCM approach when ruling out parameter values as this provides a more consistent comparison with the GP-based framework.

Figure 9 compares the NROY spaces for the w_{29m} metric, obtained either directly from the 10^4 SCM simulations ($\mathrm{NROY_{SCM}}$) or with Wave 9 emulator ($\mathrm{NROY_{GP}}$). $\mathrm{NROY_{SCM}}$ is significantly smaller than $\mathrm{NROY_{GP}}$ (10.0% against 18.0%), but the two NROY spaces are highly consistent. This comparison clearly validates the statistical framework used in section 5 and the quality of the surrogate models in predicting the metric and its uncertainty.

7 Conclusion

Stable boundary layers are still critical features for weather and climate models. In the present work, we seek to assess whether these model deficiencies reflect calibration choices, or whether they are more deeply rooted in the formulations and implementations of the turbulence parameterization themselves. In the latter case, this would clearly point to intrinsic limits of current parameterizations and possibly to missing processes key for stable boundary layers. To address this question, we took the example of the CNRM atmospheric model, namely ARPEGE-Climat 6.3, which implements the Cuxart et al. (2000) 1.5-order turbulence parameterization, with a few bounds that were historically added to prevent undesirable model behavior under certain circumstances (e.g., runaway cooling over Antarctica). At this stage, our example solely makes use of a single-column model framework, based on the very stable boundary layer regime of GABLS4 (Bazile et al., 2015). The Large-Eddy Simulation (LES) ensemble analyzed by Couvreux et al. (2020a) serves as reference for evaluating and constraining the model behavior. A statistical approach, based on history matching and Gaussian-Process-based surrogate models, is then used to identify whether there exists or not some calibration of the free parameters of the turbulence parameterization, which can provide satisfying results on this GABLS4 case. More precisely, our framework follows the process-based tuning advocated by Couvreux et al. (2020b) to characterize the part of the free parameter space, which leads to SCM simulations compatible with the LES references, given the various sources of uncertainty.

We have addressed this experience using two vertical resolutions, namely the standard ARPEGE-Climat 6.3 vertical resolution and a LES-type vertical resolution (2 m), and using two configurations for the interaction with the surface (prescribed surface fluxes and prescribed surface temperatures). Using four metrics (two characterizing the temperature profile, and two characterizing the wind profile), sampled at a given time of the GABLS4 nighttime stable boundary-layer regime, we proved that for each SCM configurations, there exist calibrations of the Cuxart et al. (2000) turbulence parameterization, as implemented in ARPEGE-Climat 6.3, which provide results consistent with the LES reference. This indicates in an unambiguous manner that this turbulence parameterization contains sufficient physics to capture strongly-stable boundary layers. As expected, such acceptable model behavior requires calibration that allows to weaken turbulent mixing. This is mostly achieved when strongly reducing the impact of lower bounds historically introduced for maintaining a minimum turbulent mixing (mixing length and minimum flux close to the surface). In contrast, and even though it can be revisited, the calibration of other turbulent parameters can broadly remain consistent with the previous proposals of Cuxart et al. (2000) and Cheng et al. (2002). This importance of lower bounds in the turbulence parameterization clearly echoes similar results obtained by Vignon, Hourdin, et al. (2017) for the climate model LMDZ.

The present work is also the opportunity to gather and formalize our experience with the statistical tools used here and borrowed from the Uncertainty Quantification community (history matching, GP-based surrogate models). As such, we attempt to provide guidance for their use in the context of parameterization and atmospheric model calibration. The importance of the different sources of uncertainty is emphasized,. The choice of metrics is an important step that is case-dependent: if the analysis of the reference simulation is important by highlighting the main biases of the standard version of the model, the analysis of the model behavior over the first wave simulations is equally important. It allows to explore, to some extent, what the model can do and where the uncertainties lie. We also illustrate how to understand and tackle the convergence question of the framework by comparing the emulator uncertainties to the other sources of uncertainty.

The present first step based on an SCM/LES framework will serve as a basis for further calibration of the ARPEGE-Climat atmospheric model in its more complete configurations, while attempting to keep an acceptable (physical) model behavior on the GABLS4 stable boundary layer regimes. The Not-Ruled-Out-Yet space will be used and further explored with 3D model configurations and other metrics to identify parameter calibration that are both compatible with stable boundary layers and the constraints of a climate model (e.g., global energy budget, climatological mean state). This requires a revisit of the present work with the introduction of a tolerance to error, and with the addition of other parameterizations involved in such regimes (e.g., radiation, surface fluxes, snow).

Acknowledgments

This work received funding from grant HIGH-TUNE ANR-16-CE01-0010. It was supported by the DEPHY2 project, funded by the French national program LEFE/INSU and the GDR-DEPHY. Daniel Williamson was funded by NERC grant: NE/N018486/1 and by the Alan Turing Institute project Uncertainty Quantification of multi-scale and multiphysics computer models: applications to hazard and climate models as part of the grant EP/N510129/1 made to the Alan Turing Institute by EPSRC. The LES simulations of the GABLS4-stg3-10hr have been kindly provided by G. Matheou, B Maronga, C Van Heerwaarden, J Edwards.

All the programs, scripts, results of the GCM (SCM) as well as the reference LES used will be made available together with the paper.

References

Abdel-Lathif, A. Y., Roehrig, R., Beau, I., & Douville, H. (2018). Single-Column Modeling of Convection During the CINDY2011/DYNAMO Field Campaign With the CNRM Climate Model Version 6. *Journal of Advances in Modeling Earth Systems*, *10*(3), 578-602.

Acevedo, O. C., Mahrt, L., Puhales, F. S., Costa, F. D., Medeiros, L. E., & Degrazia, G. A. (2016). Contrasting structures between the decoupled and coupled states of the stable boundary layer. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 693-702. doi: 10.1002/qj.2693

Baas, P., van De Wiel, B., van der Linden, S., & Bosveld, F. (2018). From Near-Neutral to Strongly Stratified: Adequately Modelling the Clear-Sky Nocturnal Boundary Layer at Cabauw. *Boundary-Layer Meteorology*, *166*, 217–238.

Bazile, E., Couvreux, F., Le Moigne, P., & Genthon, C. (2015). First Workshop on the GABLS-4 Intercomparison. *GEWEX News*, *25*(3).

Bazile, E., Couvreux, F., Le Moigne, P., Genthon, C., Holtslag, A. A. M., & Svensson. (2014). GABLS4: An intercomparison case to study the stable boundary layer over the Antarctic Plateau. *GEWEX News*, *24*(4).

Beare, R. J., Macvean, M. K., Holtslag, A. A., Cuxart, J., Esau, I., Golaz, J.-C., ... others (2006). An intercomparison of large-eddy simulations of the stable boundary layer. *Boundary-Layer Meteorology*, *118*(2), 247–272.

Beljaars, A., & Viterbo, P. (1998). Role of the boundary layer in a numerical weather prediction model. *Clear and cloudy boundary layers*, 287–304.

Blackadar, A. K. (1957). Boundary layer wind maxima and their significance for the growth of nocturnal inversions. *Bull. Amer. Meteor. Soc*, *38*(5), 283–290.

Bosveld, F. C., Baas, P., Steeneveld, G.-J., Holtslag, A. A., Angevine, W. M., Bazile, E., ... others (2014). The third GABLS intercomparison case for evaluation studies of boundary-layer models. part B: results and process understanding.

Boundary-Layer Meteorology, *152*(2), 157–187.

Bougeault, P., & Lacarrere, P. (1989). Parameterization of Orography-Induced Turbulence in a Mesobeta–Scale Model. *Monthly Weather Review*, *117*(8), 1872–1890.

Cheng, Y., Canuto, V., & Howard, A. (2002). An improved model for the turbulent PBL. *Journal of the Atmospheric sciences*, *59*(9), 1550–1565.

Couvreux, F., Bazile, E., Rodier, Q., Maronga, B., Matheou, G., Chinita, M. J., … Vignon, E. (2020). Intercomparison of Large-Eddy Simulations of the Antarctic Boundary Layer for Very Stable Stratification. *Boundary-Layer Meteorology*.

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., … Rodier, Q. (2020). Process-based climate model development harnessing Uncertainty Quantification. Part I: a calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*. Retrieved from `https://doi.org/10.1002/essoar.10503597.1` (submitted) doi: 10.1002/essoar.10503597.1

Cuxart, J., Bougeault, P., & Redelsperger, J.-L. (2000). A turbulence scheme allowing for mesoscale and large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society*, *126*(562), 1–30.

Cuxart, J., Holtslag, A. A., Beare, R. J., Bazile, E., Beljaars, A., Cheng, A., … others (2006). Single-column model intercomparison for a stably stratified atmospheric boundary layer. *Boundary-Layer Meteorology*, *118*(2), 273–303.

Derbyshire, S. H. (1999). Boundary-layer decoupling over cold surfaces as a physical boundary-instability. *Boundary-Layer Meteorology*, *90*(2), 297–325.

Gallée, H., Barral, H., Vignon, E., & Genthon, C. (2015, June). A case study of a low-level jet during OPALE. *Atmospheric Chemistry and Physics*, *15*(11), 6237-6246. Retrieved from `https://hal-insu.archives-ouvertes.fr/insu-01204565` doi: 10.5194/acp-15-6237-2015

Gettelman, A., Truesdale, J. E., Bacmeister, J. T., Caldwell, P. M., Neale, R. B., Bogenschutz, P. A., & Simpson, I. R. (2019). The Single Column Atmosphere Model Version 6 (SCAM6): Not a Scam but a Tool for Model Evaluation and Development. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1381-1401. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001578` doi: 10.1029/2018MS001578

Holtslag, A., Svensson, G., Baas, P., Basu, S., Beare, B., Beljaars, A., … others (2013). Stable atmospheric boundary layers and diurnal cycles: challenges for weather and climate models. *Bulletin of the American Meteorological Society*, *94*(11), 1691–1706.

Hourdin, F., Grandpeix, J.-Y., Rio, C., Bony, S., Jam, A., Cheruy, F., … others (2013). LMDZ5B: the atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection. *Climate Dynamics*, *40*(9-10), 2193–2222.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., … Williamson, D. (2017). The Art and Science of Climate Model Tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589-602. Retrieved from `https://doi.org/10.1175/BAMS-D-15-00135.1` doi: 10.1175/BAMS-D-15-00135.1

Hourdin, F., Rio, C., Grandpeix, J.-Y., Madeleine, J.-B., Cheruy, F., Rochetin, N., … Ghattas, J. (2020). LMDZ6A: The Atmospheric Component of the IPSL Climate Model With Improved and Better Tuned Physics. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS001892. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001892` (e2019MS001892 10.1029/2019MS001892) doi: 10.1029/2019MS001892

Mahrt, L. (1998). Stratified atmospheric boundary layers and breakdown of models. *Theoretical and computational fluid dynamics*, *11*(3), 263–279.

Mascart, P., Noilhan, J., & Giordani, H. (1995). A modified parameterization of flux-profile relationships in the surface layer using different roughness length values for heat and momentum. *Boundary-layer meteorology*, *72*, 331–344.

Mauritsen, T., & Svensson, G. (2007). Observations of stably stratified shear-driven atmospheric turbulence at low and high Richardson numbers. *Journal of the atmospheric sciences*, *64*(2), 645–655.

Neggers, R. A. J. (2015). Exploring bin-macrophysics models for moist convective transport and clouds. *Journal of Advances in Modeling Earth Systems*, *7*(4), 2079-2104. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015MS000502 doi: 10.1002/2015MS000502

Noilhan, J., & Mahfouf, J.-F. (1996). The ISBA land surface parameterisation scheme. *Global and Planetary Change*, *13*(1), 145 - 159. Retrieved from http://www.sciencedirect.com/science/article/pii/0921818195000437 doi: https://doi.org/10.1016/0921-8181(95)00043-7

Paulson, C. A. (1970). The Mathematical Representation of Wind Speed and Temperature Profiles in the Unstable Atmospheric Surface Layer. *Journal of Applied Meteorology*, *9*(6), 857–861.

Pukelsheim, F. (1994). The Three Sigma Rule. *The American Statistician*, *48*(2), 88–91. Retrieved from http://www.jstor.org/stable/2684253

Randall, D., Krueger, S., Bretherton, C., Curry, J., Duynkerke, P., Moncrieff, M., ... others (2003). Confronting models with data: The GEWEX cloud systems study. *Bulletin of the American Meteorological Society*, *84*(4), 455–469.

Randall, D. A., Xu, K.-M., Somerville, R. J., & Iacobellis, S. (1996). Single-column models and cloud ensemble models as links between observations and climate models. *Journal of Climate*, *9*(8), 1683–1697.

Redelsperger, J., & Sommeria, G. (1982). Method of representing the turbulence associated with precipitations in a three-dimensional model of cloud convection. *Boundary-Layer Meteorology*, *24*(2), 231–252.

Redelsperger, J.-L., & Sommeria, G. (1986). Three-dimensional simulation of a convective storm: Sensitivity studies on subgrid parameterization and spatial resolution. *Journal of the atmospheric sciences*, *43*(22), 2619–2635.

Roehrig, R., Beau, I., Saint-Martin, D., Alias, A., Decharme, B., Gurmy, J.-F., ... Snsi, S. (2020). The CNRM global atmosphere model ARPEGE-Climat 6.3: description and evaluation. *Journal of Advances in Modeling Earth Systems*, *n/a*(n/a), e2020MS002075. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002075 doi: 10.1029/2020MS002075

Salter, J. M., & Williamson, D. (2016). A comparison of statistical emulation methodologies for multi-wave calibration of environmental models. *Environmetrics*, *27*(8), 507-523. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2405 doi: 10.1002/env.2405

Salter, J. M., Williamson, D. B., Scinocca, J., & Kharin, V. (2019, October). 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

Sandu, I., Beljaars, A., Bechtold, P., Mauritsen, T., & Balsamo, G. (2013). Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models? *Journal of Advances in Modeling Earth Systems*, *5*(2), 117–133.

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., ... Madec, G. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4182-4227. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001791 doi: 10.1029/2019MS001791

Simmons, A. J., & Burridge, D. M. (1981). An Energy and Angular-Momentum Conserving Vertical Finite-Difference Scheme and Hybrid Vertical Coordinates. *Monthly Weather Review*, *109*(4), 758-766.

Steeneveld, G.-J., Holtslag, A., Nappo, C., Van de Wiel, B., & Mahrt, L. (2008). Exploring the possible role of small-scale terrain drag on stable boundary layers over land. *Journal of applied meteorology and climatology*, *47*(10), 2518–2530.

Steeneveld, G.-J., Van de Wiel, B., & Holtslag, A. (2006). Modeling the evolution of the atmospheric boundary layer coupled to the land surface for three contrasting nights in CASES-99. *Journal of the atmospheric sciences*, *63*(3), 920–935.

Svensson, G., Holtslag, A., Kumar, V., Mauritsen, T., Steeneveld, G., Angevine, W., … others (2011). Evaluation of the diurnal cycle in the atmospheric boundary layer over land as represented by a variety of single-column models: the second GABLS experiment. *Boundary-layer meteorology*, *140*(2), 177–206.

Tsiringakis, A., Steeneveld, G. J., & Holtslag, A. A. M. (2017). Small-scale orographic gravity wave drag in stable boundary layers and its impact on synoptic systems and near-surface meteorology. *Quarterly Journal of the Royal Meteorological Society*, *143*(704), 1504–1516.

Van de Wiel, B. J., Moene, A., & Jonker, H. (2012). The cessation of continuous turbulence as precursor of the very stable nocturnal boundary layer. *Journal of the Atmospheric Sciences*, *69*(11), 3097–3115.

van Hooft, A., van Heerwaarden, C., Popinet, S., de Roode, S., van de Wiel, B., et al. (2017). Adaptive grid refinement for atmospheric boundary layer simulations. In *Egu general assembly conference abstracts* (Vol. 19, p. 7784).

Vernon, I., Goldstein, M., & Bower, R. G. (2010, 12). Galaxy formation: a Bayesian uncertainty analysis. *Bayesian Analysis*, *5*(4), 619–669. Retrieved from https://doi.org/10.1214/10-BA524

Vignon, E., Hourdin, F., Genthon, C., Gallee, H., Bazile, E., Lefebvre, M.-P., … Van de Wiel, B. J. (2017). Antarctic Boundary-Layer Parametrization in a General Circulation Model: 1D simulations facing summer observations at Dome C. *Journal of Geophysical Research: Atmospheres*.

Vignon, E., Hourdin, F., Genthon, C., Van de Wiel, B. J. H., Galle, H., Madeleine, J.-B., & Beaumet, J. (2018). Modeling the Dynamics of the Atmospheric Boundary Layer Over the Antarctic Plateau With a General Circulation Model. *Journal of Advances in Modeling Earth Systems*, *10*(1), 98-125. doi: 10.1002/2017MS001184

Vignon, E., van de Wiel, B. J. H., van Hooijdonk, I. G. S., Genthon, C., van der Linden, S. J. A., van Hooft, J. A., … Casasanta, G. (2017). Stable boundary-layer regimes at Dome C, Antarctica: observation and analysis. *Quarterly Journal of the Royal Meteorological Society*, *143*(704), 1241-1253.

Voldoire, A., Saint-Martin, D., Snsi, S., Decharme, B., Alias, A., Chevallier, M., … Waldman, R. (2019). Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2177-2213. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001683 doi: 10.1029/2019MS001683

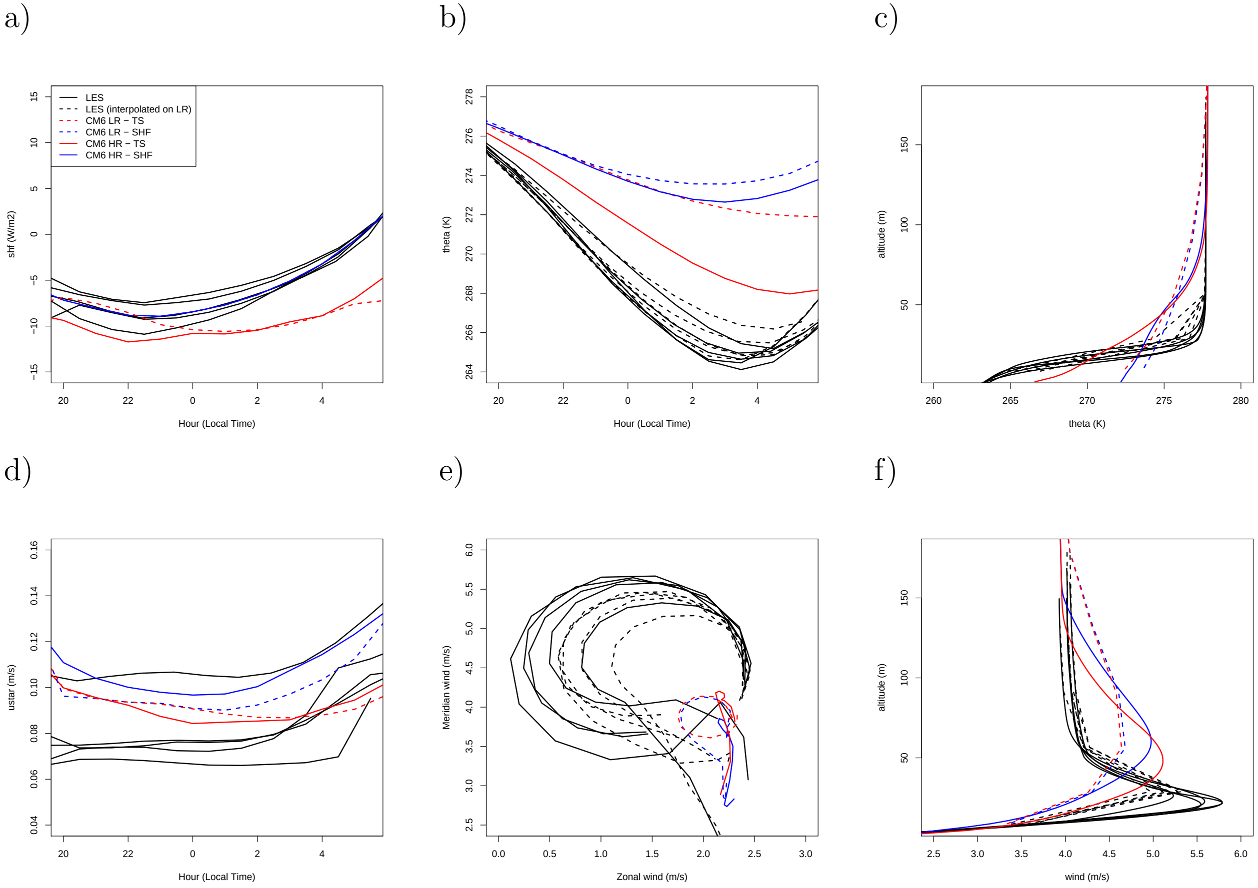
Williamson, D. (2015). Exploratory ensemble designs for environmental models using k-extended Latin Hypercubes. *Environmetrics*, *26*(4), 268-283. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2335 doi: 10.1002/env.2335

Williamson, D., Blaker, A., Hampton, C., & Salter, J. (2015, 9). Identifying and removing structural biases in climate models with history matching. *Climate Dynamics*, *45*, 1299-1324.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., & Yamazaki, K. (2013, 10). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics*, *41*, 1703-1729. doi: 10.1007/s00382-013-1896-4

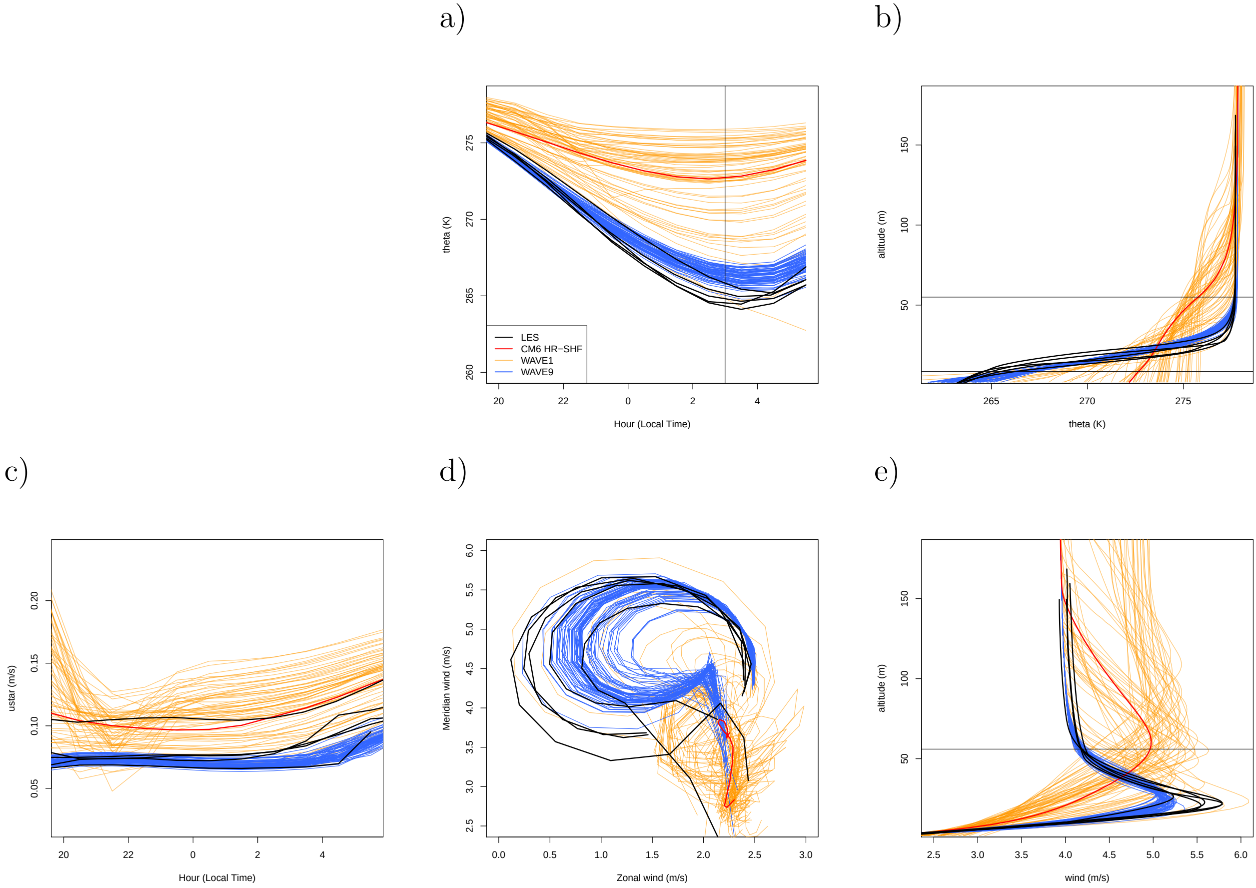
Williamson, D. B., Blaker, A. T., & Sinha, B. (2017). Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model. *Geoscientific Model Development*, *10*(4), 1789–1816. Retrieved from `https://www.geosci-model-dev.net/10/1789/2017/` doi: 10.5194/gmd-10-1789-2017

Figure 1.



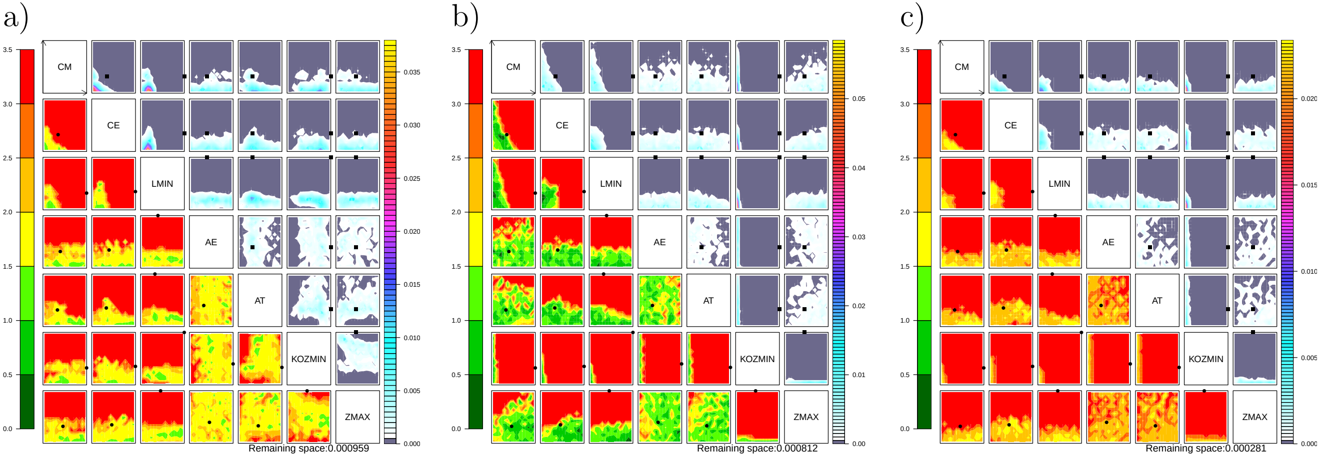
a)
b)
c)
d)
e)
f)
LES
LES (interpolated on LR)
CM6 LR − TS
CM6 LR − SHF
CM6 HR − TS
CM6 HR − SHF
shf (W/m2)
Hour (Local Time)
theta (K)
Hour (Local Time)
altitude (m)
theta (K)
ustar (m/s)
Hour (Local Time)
Meridian wind (m/s)
Zonal wind (m/s)
altitude (m)
wind (m/s)

Figure 2.



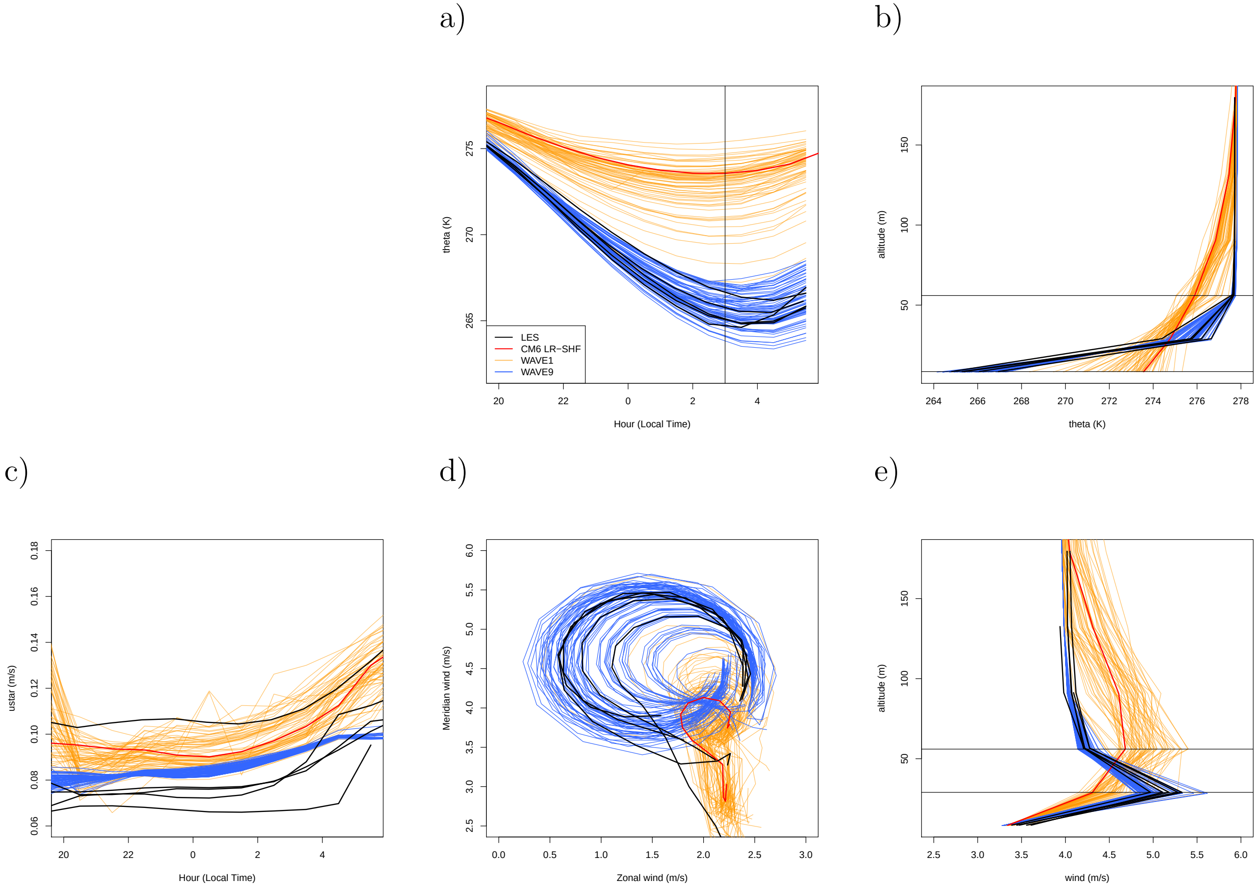
a)
b)
c)
d)
e)
theta (K)
Hour (Local Time)
LES
CM6 HR−SHF
WAVE1
WAVE9
altitude (m)
theta (K)
ustar (m/s)
Hour (Local Time)
Meridian wind (m/s)
Zonal wind (m/s)
altitude (m)
wind (m/s)

Figure 3.



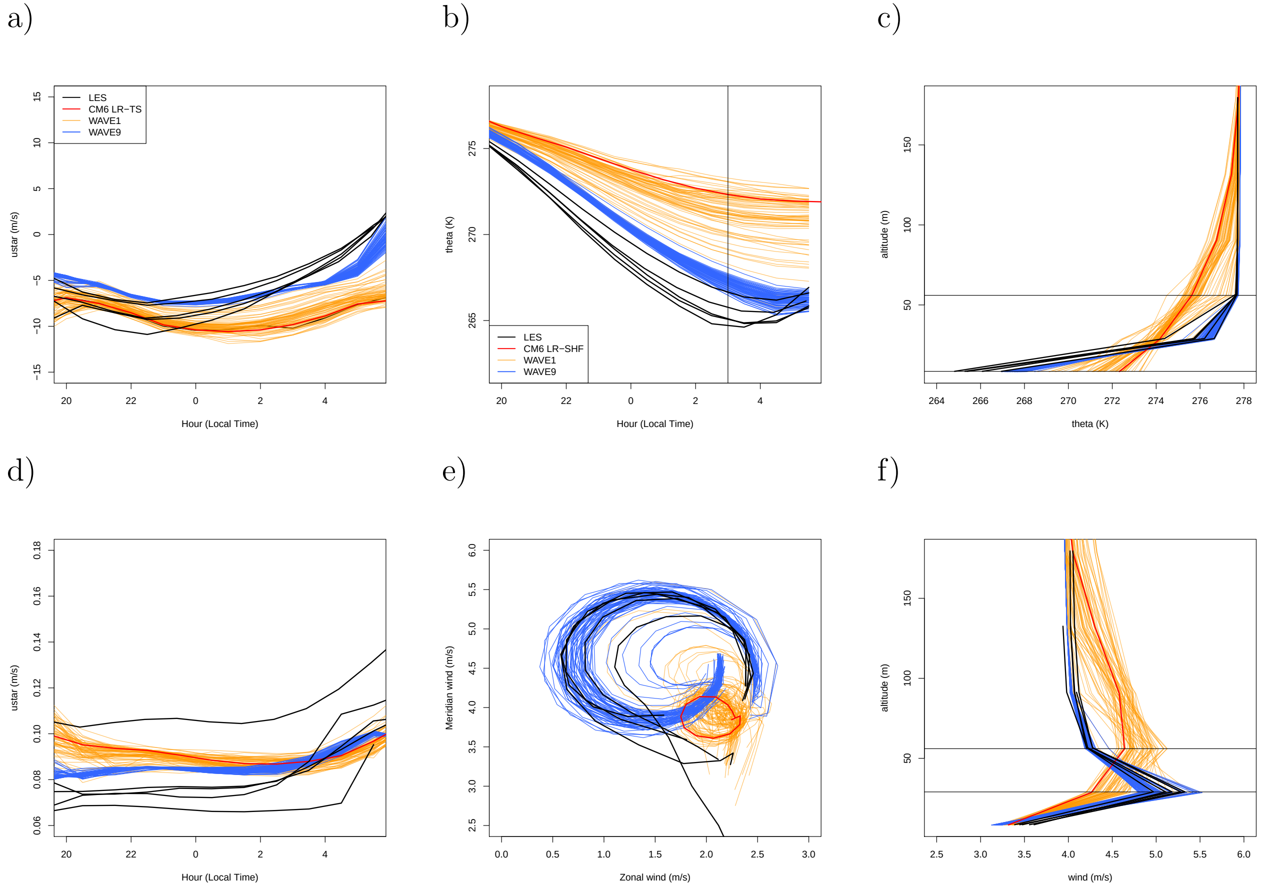
a)
b)
c)
CM
CE
LMIN
AE
AT
KOZMIN
ZMAX
Remaining space:0.000959
Remaining space:0.000812
Remaining space:0.000281

Figure 4.



a)
b)
c)
d)
e)
LES
CM6 LR−SHF
WAVE1
WAVE9
theta (K)
Hour (Local Time)
altitude (m)
ustar (m/s)
Meridian wind (m/s)
Zonal wind (m/s)
wind (m/s)

Figure 5.



a)
b)
c)
d)
e)
f)
LES
CM6 LR−TS
WAVE1
WAVE9
ustar (m/s)
Hour (Local Time)
LES
CM6 LR−SHF
WAVE1
WAVE9
theta (K)
altitude (m)
Meridian wind (m/s)
Zonal wind (m/s)
wind (m/s)

Figure 6.

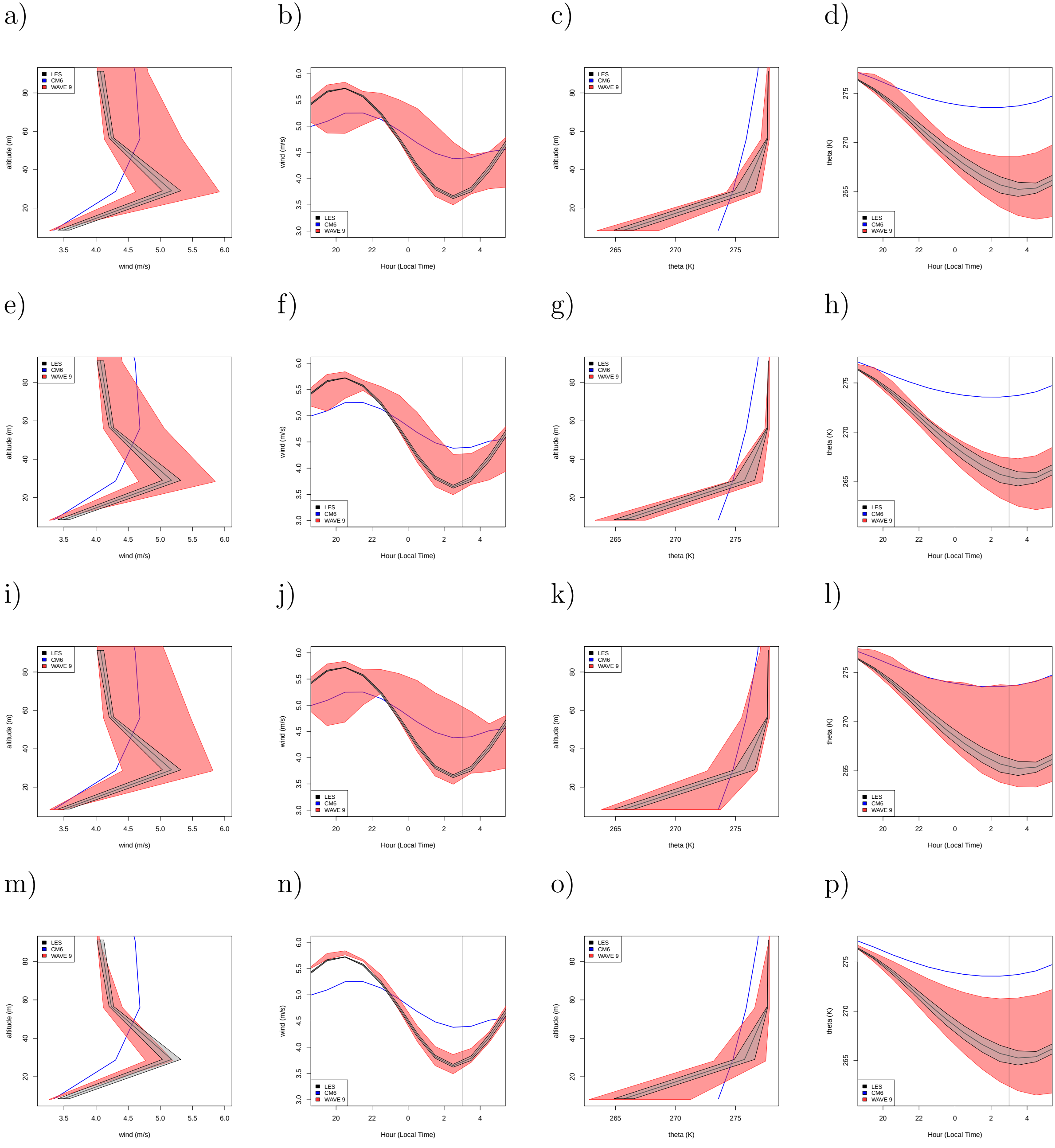

a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
m)
n)
o)
p)
LES
CM6
WAVE 9
altitude (m)
wind (m/s)
Hour (Local Time)
theta (K)

Figure 7.

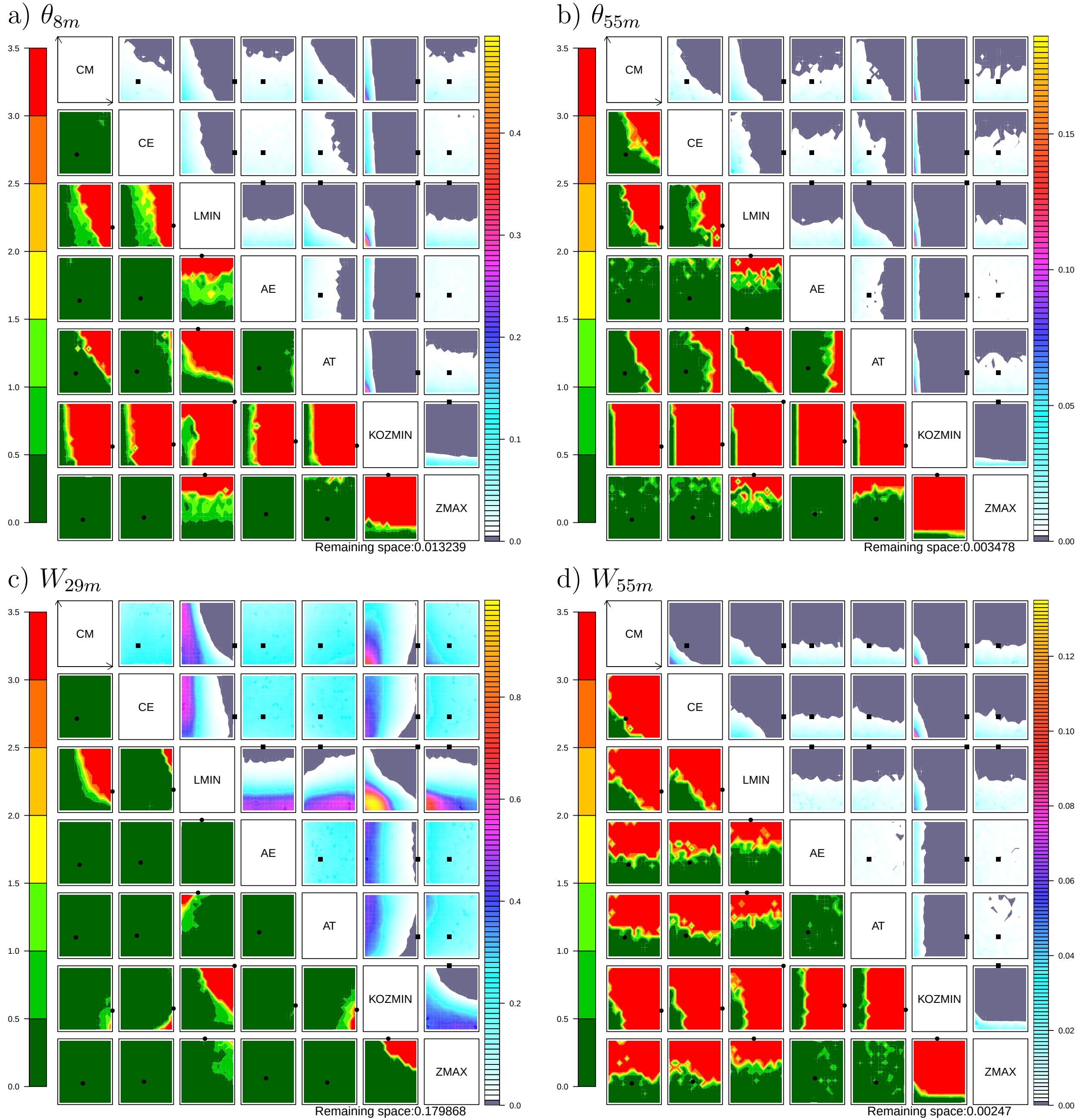

a) θ_{8m}
b) θ_{55m}
c) W_{29m}
d) W_{55m}
CM
CE
LMIN
AE
AT
KOZMIN
ZMAX
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Remaining space:0.013239
Remaining space:0.003478
Remaining space:0.179868
Remaining space:0.00247
0.0 0.1 0.2 0.3 0.4
0.00 0.05 0.10 0.15
0.0 0.2 0.4 0.6 0.8
0.00 0.02 0.04 0.06 0.08 0.10 0.12

Figure 8.

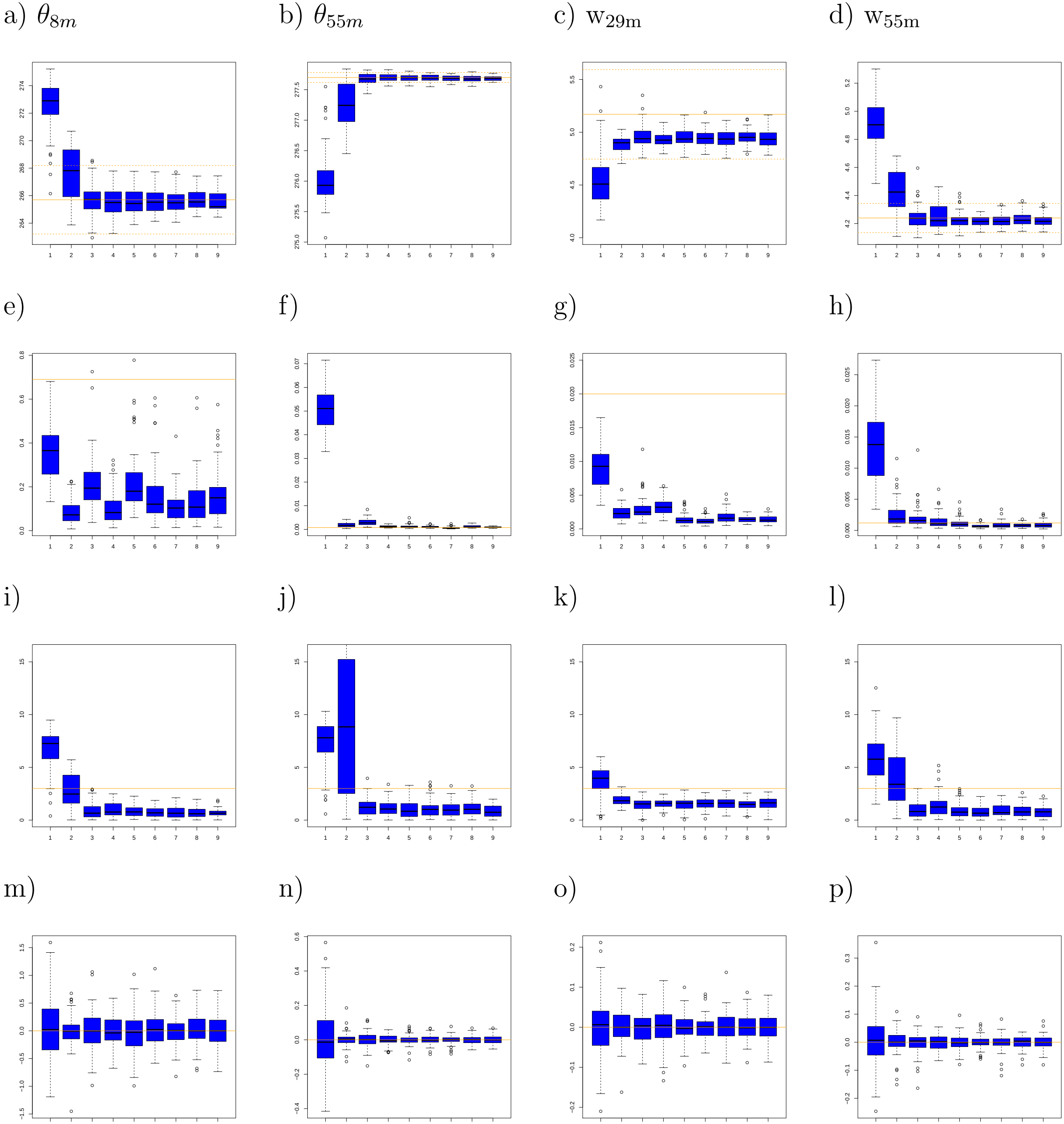
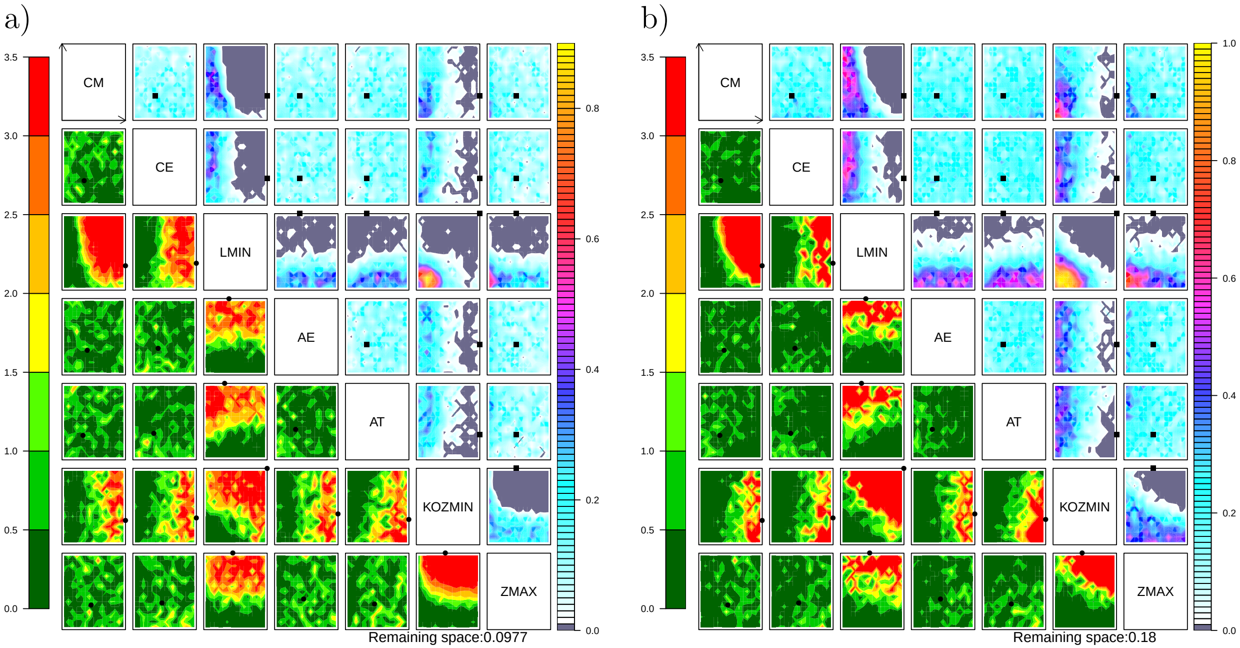

a) θ_{8m}
b) θ_{55m}
c) w_{29m}
d) w_{55m}
e)
f)
g)
h)
i)
j)
k)
l)
m)
n)
o)
p)

Figure 9.



a)
b)
CM
CE
LMIN
AE
AT
KOZMIN
ZMAX
Remaining space:0.0977
Remaining space:0.18
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1.0
0.8
0.6
0.4
0.2
0.0