



Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe

M. García-Díez^{1,2}, J. Fernández^{*2}, L. Fita³ and C. Yagüe⁴

¹ Instituto de Física de Cantabria, IFCA (CSIC-UC), Santander, Spain

² Dept. of Applied Mathematics and Comp. Sci., Universidad de Cantabria, Santander, Spain

³ Climate Change Research Center, University of New South Wales, Sydney, Australia

⁴ Departamento de Geofísica y Meteorología, Universidad Complutense de Madrid, Madrid, Spain

*Correspondence to: Dept Matemática Aplicada y CC Comp., ETSI de Caminos, Canales y Puertos, Avenida los Castros s/n, Santander, 39005, Spain
jesus.fernandez@unican.es

The seasonal dependence of Weather Research and Forecasting (WRF) model surface temperature biases and sensitivity to PBL schemes are jointly explored. For this purpose, the year 2001 was simulated using three different PBL schemes in a domain covering all Europe. The simulations were compared with gridded observations, upper air data and high-frequency station data. Seasonal and daily cycles were analysed, aiming at providing a link between long term biases and restricted case studies. The results show that the model mean bias significantly depends on the season, being warm in winter and cold in summer. The winter warm bias is related to misrepresented cold extremes, while a systematic cold bias dominates the whole temperature range in summer. Regarding PBL schemes, an overall underestimation of the entrainment is found, with the non-local YSU scheme producing systematically warmer temperatures. It is shown that the opposite seasonal biases and the systematic behaviour of the PBL schemes along the year lead to a different best-performing scheme in winter and summer. Moreover, the best-performing PBL scheme in an average sense is a result of the compensation of errors. The average summer results can be partially explained by a detailed case study. It is concluded that short term studies should be used with caution to decide on the parametrizations to be used in long term simulations. Copyright © 2011 Royal Meteorological Society

Key Words: WRF, numerical modelling, model bias, parametrization schemes, planetary boundary layer

Received ...

Citation: ...

1. Introduction

Parametrization of sub-grid scale phenomena persists as one of the most challenging problems in numerical modelling of the atmosphere. Parametrization schemes describe the contribution of unresolved atmospheric phenomena in terms of variables resolved at the model discrete grid. These schemes rely on the detailed observation of variables which are seldom observed (e.g. very high-frequency winds, heat fluxes, detailed vertical profiles of the atmosphere, size

distribution of microphysical species, etc.) and need specific field campaigns in order to acquire the data to validate the physically-plausible model and adjust its parameters. Field campaigns are expensive and, thus, they are usually available for short time periods and over small areas. This low availability (in time and space) of adequate observations to validate the inner assumptions of physical parametrizations, contrasts with their global application scope (Stensrud 2007). As a result of this duality, the comparison of different parametrization approaches for

the same sub-grid phenomenon gives rise to, at least, two different kind of studies: (1) those devoted to assess the validity of the assumptions involved in the physical schemes, determining the role of the different components of the physical model and their adequacy to represent the observed reality (see e.g. Alapaty *et al.* 1997; Braun and Tao 2000; Kotroni and Lagouvardos 2001; Bright and Mullen 2002; Stensrud and Weiss 2002; Zhang and Zheng 2004; Deng and Stauffer 2006; Hong *et al.* 2006; Otkin and Greenwald 2008; Hu *et al.* 2010; Shin and Hong 2011; Steeneveld *et al.* 2011, among others) and (2) those devoted to the comparison of the parametrization schemes in a more general way, beyond specific events or meteorological situations (see e.g. Fernández *et al.* 2007; Flaounas *et al.* 2011; Argüeso *et al.* 2011; Awan *et al.* 2011; Jerez *et al.* 2012; Jerez 2011; Evans *et al.* 2011, among others). The studies of the first kind, aim at contributing to the advance of parametrization schemes by (among many others): detecting missing or under-represented processes, obtaining more accurate parameters or distinguishing situations where regimes change and a different model or parameter set is required. These “case studies” require accurate, high-frequency, low-available data. On the other hand, studies of the second kind have a more pragmatic approach and usually aim at selecting the best scheme for a given purpose (weather forecast, regional climate studies, etc.) or at combining multiple schemes in order to provide a probabilistic forecast. These studies usually focus on their target variables of interest (most commonly screen level variables) and make use of averaged values to summarize the skill of the different schemes across space and/or time. Thus, these “statistical studies” rely on well observed variables over relatively long periods and large areas. Model validation in terms of average surface variables suffers from several drawbacks, though. The main one is the chance that the model may provide the right average value for the wrong reason, as a result of a compensation of errors. A recent example is provided by Nikulin *et al.* (2012), where an ensemble of regional climate models is shown to represent daily average precipitation relatively well, while the daily precipitation cycle (i.e. trigger time and rainfall duration) is badly captured by most models. In the present work, we perform a “statistical study”, characterizing the differences introduced by different planetary boundary layer (PBL) schemes on the average behaviour of a mesoscale atmospheric model over a large region (Europe) and an extended period of time (1 year). However, in order to overcome some of the limitations of these kind of studies, we analyse not only average standard surface variables (such as 2-meter temperature), but also additional variables (such as specific humidity) in terms of process-related statistics, such as average daily cycles and vertical profiles. Finally, a focus is made over a short period of time at a specific site (a “case study”), in order to compare the general conclusions from averaged results with those related to a particular meteorological situation (clear-sky summer days over flat terrain).

Planetary Boundary Layer (PBL) schemes deal with the problem of parametrizing the turbulent layer that develops over earth surface due to surface heating, wind shear and friction. Vertical transport of heat, moisture, momentum and other physical properties of the lower troposphere are driven by PBL processes, and also low level clouds. Therefore, a correct parametrization of this layer is essential to achieve realistic simulations, especially regarding surface

variables. For example, a low turbulent mixing at the capping inversion layer leads to PBLs which are too cold, moist and shallow. This, in turn, produces incorrect values of important parameters like convective available energy (Bright and Mullen 2002). Excessive turbulent mixing leads to too warm, dry and thick PBLs, and this has large effects on the simulation of important meteorological systems, such as hurricanes (Braun and Tao 2000).

A PBL scheme appears naturally in numerical models when a continuous variable is discretized by considering a grid-point averaged value plus a perturbation (Reynolds decomposition) in the Navier-Stokes equations. Unresolved perturbation terms appear in the set of differential equations, which becomes incomplete. The problem of closing this equation system is called the turbulence closure problem (Stull 1988), and is the main issue a PBL scheme must face.

PBL parametrizations can be classified depending on how they approach the turbulence closure problem. Local closure schemes use variables and parameters that are defined at each model level or its neighbours (e.g. local gradients), while non-local closure schemes use parameters that can depend on the whole vertical profile, or on relationships between separated levels (e.g. diffusion coefficients dependent on the PBL thickness). This is often summarized by referring to them as local and non-local schemes.

Since they are not able to directly represent transport between non-consecutive levels, local schemes are not expected to behave correctly under fully developed turbulence (Stull 1991), but they still can compete with non-local schemes by adding higher order terms. As the top of the PBL is usually statically stable, non-local transport is important to represent the entrainment of air from the free atmosphere into the mixed layer. Entrainment, as the surface fluxes, is the factor dominating the main characteristics of the diurnal PBL. Local schemes using a Turbulent Kinetic Energy (TKE) closure, based on Mellor and Yamada (1982), have become popular, since they can represent the entrainment with the diffusion of the TKE and the so-called counter-gradient terms, that allow diffusion against the local gradients (Deardorff 1966). This is, in fact, an indirect way to represent non-local transport, so the notion of “local scheme” can be argued to have an unclear physical interpretation. Nevertheless, previous works show that, in general, local schemes tend to produce unrealistically shallow and moist boundary layers (Alapaty *et al.* 1997; Bright and Mullen 2002; Stensrud and Weiss 2002; Hong *et al.* 2006). This is directly related to their inability to directly represent large-scale turbulence and the underestimation of entrainment. Other works show that non-local schemes are not always more skillful (Deng and Stauffer 2006), and that they can produce too deep boundary layers in windy conditions (Braun and Tao 2000; Persson *et al.* 2001). It is worth to note that the results of the PBL schemes are also affected by the development of boundary layer clouds by other physical parametrization schemes (microphysics or convection), as these play also a relevant role in the budgets at the inversion layer. The present study compares both local (TKE closure), and non-local approaches over Europe (i.e. a large region encompassing a great variety of climatic conditions) using a state-of-the-art numerical modelling system.

The Weather Research and Forecasting (WRF) model (Skamarock *et al.* 2008) is a limited area model widely

used in many fields, such as local weather forecasting, air-quality studies, regional climate research, and others. The reason behind this popularity is that it is an open source model that can be freely used and modified by the user and also because of the modularity of its components. WRF components (dynamical core, parametrization schemes for the different sub-grid processes, nudging options, etc) can be combined to create many different model configurations. This flexibility makes choosing the optimal configuration for each particular study a commonly found problem. There is no consensus about the set of schemes that performs better for each application. Frequently, the performance depends on the variables, season or time of the day considered and one cannot identify a best model configuration in a general sense, even for a particular region (Zhang and Zheng 2004; Fernández *et al.* 2007; Awan *et al.* 2011; Jerez *et al.* 2012). The usual approach to deal with this problem is to carry out "statistical studies" over a sensible set of model configurations and look for the one that reproduces better the observations in an average sense. A number of sensitivity studies to PBL schemes have been carried out with WRF, and with its predecessor MM5 (Bright and Mullen 2002; Zhang and Zheng 2004; Deng and Stauffer 2006; Hong *et al.* 2006; Weisman *et al.* 2008; Shin and Hong 2011). Most of them focus on short periods of few days or weeks, and relatively small domains, mostly flat and homogeneous (Stensrud 2007). Since models are used over many different domains and meteorological conditions, this kind of restricted experiments may not be representative enough.

Awan *et al.* (2011) performed an analysis over a broad set of parametrization schemes, including cumulus, microphysics and also PBL schemes. Their analysis spans an annual cycle but they focus on a very specific and complex region, the Alps. They found that the different schemes interact non-linearly with each other, i.e. the resulting biases are not combinations of the biases caused by each individual parametrization. This makes the problem of evaluating scheme performance even harder.

Hu *et al.* (2010) analysed the performance of 3 PBL schemes in a summer season over Texas (US). Their study has some similarities with the present one, including the three PBL schemes chosen and most of the experimental setup (Section 2.1). Thus, their results can be directly compared with our work. They found that the non-local scheme showed significantly less bias than the other two (one local and the other one a combination of local and non local). Our work extends that of Hu *et al.* (2010) by analysing the seasonal dependence of the biases and performing an analysis over a larger region including many different meteorological conditions.

In this work, we explicitly sacrifice the advantages of detailed "case studies" in favour of testing whether the most important results found in previous (case) studies do hold in a large domain, over a full seasonal cycle. In particular, we mainly focus on temperatures and the main goals pursued are:

1. Assess model biases in the reproduction of the observed seasonal cycle, daily cycle and probability distribution.
2. Assess the average skill of 3 different PBL schemes in reproducing the above features in different European regions and seasons.

3. Bridge the gap with "case study" works by comparing the statistical results with a single site over a short period, in order to check the consistency of point and averaged results.

The layout of the paper is as follows. Section 2 describes the experimental design and the observational datasets used. The results (Section 3) analyze the biases found when comparing with surface and upper-air observations and also focuses on a summer clear-sky case study. Section 4 summarizes the main conclusions found.

2. Data and model configuration.

2.1. Model configuration

The domain used covers all Europe with a horizontal resolution of 15 km (Figure 1) and 39 vertical levels. The model has been used in reforecast mode, performing 42-hour simulations, started daily at 6 UTC. The approach is similar to that followed in Lenderink *et al.* (2009) or Hu *et al.* (2010). This leads to a total of 1095 (365×3) individual simulations. Such a large number of simulations was handled by using WRF4G (Fernández-Quiruelas *et al.* 2010), a framework that allows to manage the WRF workflow easily in multi-parametric experiments. The initial and boundary data are taken 6-hourly from the ERA-Interim reanalysis (Simmons *et al.* 2007). Daily simulations were concatenated to create a pseudo-continuous simulation, leaving the first 12 hours out as spin-up. This method reduces the bias, keeping the model close to the forcing fields but still free enough to generate mesoscale features. This methodology enables the hour-by-hour comparison shown in Section 3.6.

Advanced research WRF (WRF-ARW) version 3.1.1 has been used. The common set of parametrizations chosen is the following: Microphysics, WRF Single Moment 5-class scheme (Hong and Lim 2006); cumulus, Grell-Devenyi ensemble scheme (Grell and Devenyi 2002); long wave radiation scheme, Rapid and accurate Radiative Transfer Model (RRTM) (Mlawer *et al.* 1997); short wave radiation scheme by Dudhia (1989). The land surface model (LSM) used was Noah (Chen and Dudhia 2001).

In order to explore the differences between the local and non-local approach to the PBL modelling, the three schemes chosen for this work are a non-local one, a local one and a non-local one that turns to local under stable conditions. Namely, the schemes used were:

Yonsei University scheme

The Yonsei University PBL scheme (YSU) (Hong *et al.* 2006) is a first order scheme that uses non-local eddy diffusivity coefficients to compute the turbulent fluxes. It is based on the Medium Range Forecast (MRF) model PBL scheme, and improves it with an explicit treatment of the entrainment. The turbulent diffusion equations are given by equation (1), where C is the diagnosed variable, K_c the eddy diffusivity, and γ_c a correction to the local flux to account for the effect of the large eddies. The term on the right is the entrainment flux, which in the case of sensible heat is taken proportional to the surface buoyancy flux. For the other variables, the flux across the inversion layer is considered proportional to the jump of the variable itself at this layer.

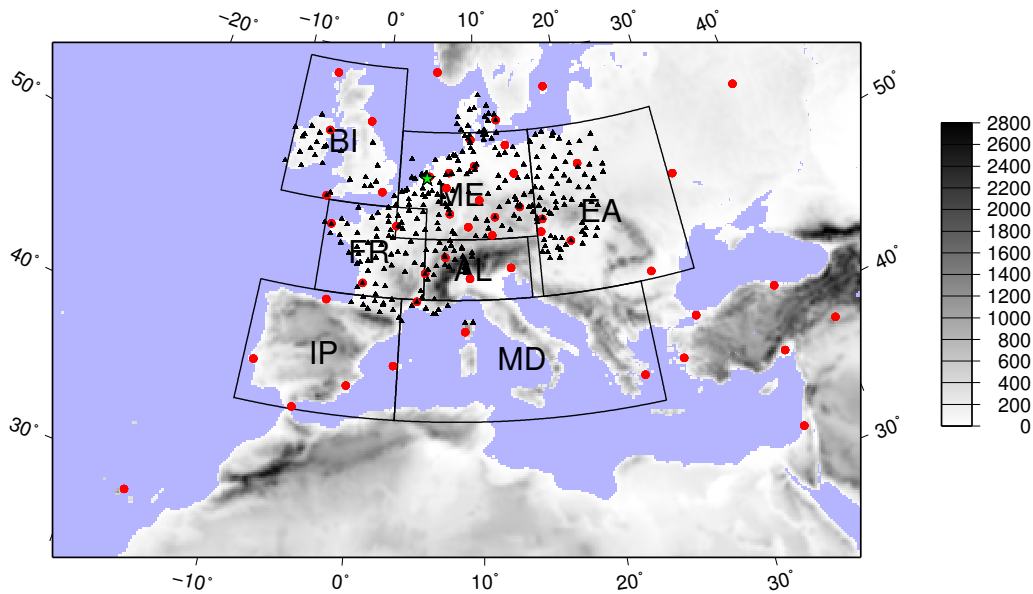


Figure 1. Domain of the simulations, and limits of the PRUDENCE regions used in the paper. Also shown: SYNOP stations used in Section 3.3 (black triangles), upper-air soundings used in Section 3.4 (red circles), and the CESAR observatory (star). The projection used is a Lambert Conformal Conic projection.

The PBL height, h , is diagnosed using a critical Richardson number.

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z} \left[K_c \left(\frac{\partial C}{\partial z} - \gamma_c \right) - (\overline{w'c'})_h \left(\frac{z}{h} \right)^3 \right] \quad (1)$$

Mellor-Yamada-Janjic scheme

The Mellor-Yamada-Janjic PBL scheme is a local closure scheme of order 1.5 and level 2.5. It is based on Mellor and Yamada (1982) and its implementation is described on Janjić (1990), Janjić (1994) and Janjić (2002). It solves a prognostic equation for the TKE. The equations for the heat and moisture fluxes include a term that allows them to go against the local gradient, so counter-gradient fluxes caused by large eddies can be represented. This, along with the TKE diffusion outside of the mixed layer, leads to an improved representation of the entrainment. PBL height is diagnosed using a TKE threshold. A maximum value, that depends on the stability, is imposed to the master length scale.

Asymmetric Convective Model 2

The Asymmetric Convective Model 2 (ACM2) (Pleim 2007a,b) uses a combination of local (first order) and non-local transport that switches off smoothly to local eddy diffusion at stable environments. At unstable conditions, local transport is used for subsidence, while upward fluxes are modelled combining local eddy diffusion with a non-local approach that computes the transition probability between non-consecutive levels. The latter approach, based on Blackadar (1978); Stull (1984) and Pleim and Chang (1992), can represent rising thermals much larger than the grid spacing. The combination between local and transilient approaches is weighted with a parameter that depends on stability.

The Surface-Layer (SL) parametrizations are paired with the PBL parametrizations, and all use similarity

theory (Monin and Obukhov 1954) to compute the surface turbulent fluxes, though they do not share the same similarity functions. SL scheme paired with YSU (called MM5) uses the functions proposed by Paulson (1970), Dyer and Hicks (1970), and Webb (1970), while SL paired with MYJ (ETA) uses those proposed by Zilitinkevich (1995) and Beljaars (1995). ACM2 has its own approach to similarity functions, proposed by Pleim (2006). It also should be noted that the PBL schemes themselves do use different similarity profiles to compute some parameters. The assumptions made in the different similarity functions can have relevant effects in the simulated surface variables, specially under stable environments (Louis 1979; Yagüe et al. 2006).

In the following, the three 1-year simulations resulting from the use of these PBL schemes will be referred to as: YSU, MYJ and ACM2.

2.2. Data

2.2.1. The E-OBS dataset

The E-OBS dataset (Haylock et al. 2008) is a daily high resolution grid derived from observed precipitation, maximum, minimum and mean 2-meter temperatures. It was developed within the ENSEMBLES EU-project (Hewitt and Griggs 2004), and covers Europe with different spatial resolutions, intended to match those used by the RCMs of the project, for the period 1950-2006. In the present study the 0.25° regular lon-lat grid, version 4.0, has been used. Data from about 2000 stations were used for developing this grid, and the series are available through the European Climate Assessment and Data set (ECA&D; <http://eca.knmi.nl>). To enable comparison, WRF data has been interpolated to this grid using bilinear interpolation.

In spite of being the best European observation-based dataset available, E-OBS presents several problems, and it should be used with caution for RCM validation. Comparisons with regional datasets elaborated with much denser station networks reveal important regional biases

in the E-OBS dataset (Hofstra *et al.* 2009; Herrera *et al.* 2010; Kysely and Plavcova 2010). This especially affects areas with sparse station data and complex orography, where the stations entering the interpolation might not be representative of their surrounding region. The bias was also found to be larger for extreme values. The interpolation process smooths the data, and this leads to an important underestimation of the extremes, especially for precipitation. A grid of standard errors is provided with the dataset, but they are generally smaller than the real errors, as they do not take into account the previously mentioned issues (Hofstra *et al.* 2009). Therefore, the results of the validation of a limited area model against this dataset should be interpreted with caution, and possibly even discarded over some areas with very poor station density, like north Africa or eastern Turkey. On the other hand, there are many areas where E-OBS has a good coverage, like Benelux, England, and Germany.

In order to analyze the spatial variability of the skill, the domain has been divided into regions, following those defined in the PRUDENCE Christensen and Christensen (2007) project. These are shown in Figure 1. Table I shows the limits for each region, their area, station density (as a measure of quality of the dataset) and standard deviation of the terrain height (as a measure of orographic complexity).

2.2.2. Atmospheric soundings

The soundings of the Earth System Research Laboratory (ESRL) Radiosonde Database have been used to check the model behaviour on levels other than the surface. The quality of the data, the levels available and the frequency depends on the station considered. Most stations only provide two daily soundings, taken at 00 and 12 UTC, and a few also at 6 UTC and 18 UTC. The variables used in this work are temperature and dew point depression. The specific humidity used in Section 3.4 has been derived from temperature, pressure and dew point using the Magnus formula (Lawrence 2005).

In general, sounding data suffers from several deficiencies. The ESRL radiosonde database has not been subjected to a quality control, and many missing and incorrect data are present. Thus, only stations with 90% or more data have been considered (a total of 51 sites, see Figure 1), and the values departing from the model more than 2.5 the interquartile range of the full-year series have been rejected as outliers.

2.2.3. SYNOP data

Surface synoptic observations (SYNOP) are observational data recorded by a large number of meteorological stations around the world. This data are encoded and broadcast by radio, following an international format established by the World Meteorological Organization (WMO). The frequency of the data depends on the station, ranging from one hour to two days or more. For this work hourly European SYNOP codes of the year 2001 have been compiled, decoded and filtered from outliers and meaningless values. Only stations with hourly data and less than 20% missing data has been used, after filtering values departing from the 2.5 the interquartile range as outliers (see figure 1 for a map with the station used). The goal of this process is to compare the daily cycle produced by WRF with the observed one, for temperature and specific humidity. The latter has been

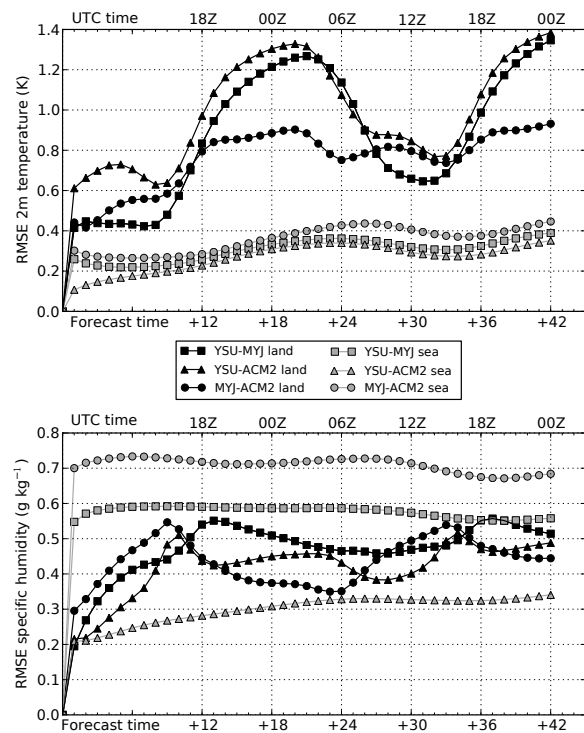


Figure 2. Root mean square difference between simulations, computed for each forecast time (in hours), for 2 m temperature (top) and for 2 m specific humidity (bottom). Note that the 0 corresponds to the initial condition, taken at 06 UTC every day.

computed with the Magnus (Lawrence 2005) formula using the surface pressure and the screen level standard and dew point temperatures.

2.2.4. CESAR tower data

Observations from the Cabauw Experimental Site for Atmospheric Research (CESAR) tower located in the Netherlands (shown with a star in Figure 1) have been used to carry out a clear-sky case study. This station provides data with high temporal resolution and non-standard variables such as surface turbulent fluxes. Upper-air data have been taken from the nearby (20 km) De Bilt sounding station. The characteristic flat and homogeneous terrain of the Netherlands makes the nearest grid point approximation a reasonable assumption.

3. Results

3.1. Model sensitivity to forecast time

In order to analyse the impact of the daily reforecast approach, the root mean square difference (RMSD) between the simulations is computed for each forecast time, from 0 to 42 hours, for screen level temperature and specific humidity. The spread associated to the use of different PBL schemes is expected to increase as forecast time increases. For temperature (Figure 2, top) over land grid points, a strong diurnal cycle can be observed, so the model is more sensitive to the PBL scheme at night, when stable conditions are more frequent. Over the sea, the sensitivity is smaller and there is no diurnal cycle. This is expected given that the three simulations share the

Table I. Limits of the PRUDENCE regions used for validation, their area and the number and density of stations used for E-OBS. The area corresponds only to that occupied by the E-OBS grid points falling inside the region.

ID	Description	W	E	S	N	Area (1000 km ²)	Station density (St. per 1000 km ²)	Height S. Dev. (m)
BI	British Isles	-10	2	50	59	305	0.53	109
ME	Mid Europe	2	16	48	55	600	0.30	196
AL	Alps	5	15	44	48	308	0.63	667
FR	France	-5	5	44	50	345	0.12	237
MD	Mediterranean	3	25	36	44	573	0.21	362
EA	Eastern Europe	16	30	44	55	1210	0.14	231
IP	Iberian Peninsula	-10	3	36	44	645	0.11	378

same Sea Surface Temperature (SST). Since the growth of the sensitivity as forecast time increases is small (especially when disregarding the initial 12h), we conclude that the pseudo-continuous simulations are temporally homogeneous enough for the analysis.

The behaviour of specific humidity (Figure 2 bottom) is different, the RMSD between local (MYJ) and non-local (YSU, ACM2) schemes is larger over the sea than inland, and it appears almost instantaneously after the beginning of the run. This points to a systematic differences in how this variable is diagnosed over the sea by the schemes.

3.2. Comparison with E-OBS.

A first look at model biases (Figure 3) shows relevant differences among the different PBL schemes, seasons and variables (maximum and minimum temperatures). DJF minimum temperature (Figure 3, first column) shows a warm bias in most regions, larger in the eastern half of the domain. The smaller bias is found in areas with good station density and small orographic complexity (such as ME, BI or FR). Regions with scarce data (North Africa, Eastern Turkey) and/or steep orography (Alps, Pyrenees) can be distinguished in the spatial patterns. The large spatial variability of winter minimum temperatures, very dependent on local cloud cover, winds and orography favourable to inversions, might be affecting the accuracy of the interpolation used in E-OBS.

Maximum temperature bias (Figure 3, right) shows smoother spatial patterns, suggesting better data quality (better performance of the interpolation) due to the smaller spatial variability of maximum temperatures. A generalized cold bias is found for the three PBL schemes in JJA, which also extends to MAM and SON (not shown), reaching -3°C in regions with good data quality. This cold bias is smaller for ACM2, but still important. For DJF maximum temperature, YSU is close to E-OBS over a large part of the domain, while MYJ and ACM2 show small cold biases. Finally, a large area of warm bias, located at the north-eastern corner of the E-OBS domain, can be observed for the three schemes. North Africa and Eastern Turkey shows large biases that are probably related to the excessive interpolation of scarce data (see Figure 1 in Hofstra *et al.* (2009) for a map of the station network used in E-OBS).

Mean bias for each PRUDENCE region and season has been plotted in Figure 4. An annual cycle on the bias is apparent for the three schemes. This bias cycle is always cold for maximum temperatures and becomes warm in

winter for minimum temperatures in most regions. Regions with good coverage (BI and ME) show a warm bias of $0.5\text{--}2.5^{\circ}\text{C}$ for winter minimum temperatures in YSU, while ACM2 and MYJ stay closer ($< 0.5^{\circ}\text{C}$) to E-OBS, except at northern BI. JJA minimum temperatures show very different biases, mostly cold, except over Eastern Europe in YSU and MYJ, where the bias is small with warm areas. ACM2 shows the coldest summer minimum temperatures, with a general cold bias reaching -2.5°C even in regions with good station coverage. MAM and SON biases are generally half way between those in summer and winter. ACM2 is the scheme showing less bias for maximum temperatures, but it is still too cold. For minimum temperatures, ACM2 and MYJ are closer to observations in winter than YSU, but the non-local scheme outperforms them in summer and in other seasons. Mean temperature (not shown), computed as $\frac{T_{max}+T_{min}}{2}$, is affected by the warm winter minimum temperature bias in YSU, so the best scheme, and the best general approach (local or non-local), would reverse depending on the season, and studies referring only to shorts periods of time would not hold for the whole year. This fact is a consequence of the previously addressed annual cycle in the bias, which is generally greater than the model sensitivity to the PBL scheme. As this bias is not necessarily related to the parametrization of the PBL, following this criteria to determine the best PBL scheme would be meaningless.

Seasonal dependence of the bias suggests that it is related to absolute temperatures, i.e., that different biases exist in different parts of the temperature range. To check this, quantile-quantile plots (q-q plots) have been computed for each region and for both maximum and minimum temperatures (Figure 5). Q-q plots compare the probability distribution of two variables, by representing on a Cartesian plane some quantiles of a variable against those of another variable or a theoretical distribution. In our case, we compared the distribution of simulated temperatures (Y-axis) against the observations (X-axis) dividing the probability range into 40 pieces (i.e. taking a quantile every 2.5%). These representations allow to easily identify deviations in the probability distribution (as departures from a straight diagonal line), biases (as shifts), differences in the variability (as straight lines with a different slope) or asymmetries (as curved lines). In order to compute the quantiles no averaging was carried out, pooling time and space for a given region and season.

DJF minimum temperatures distributions differ on lower quantiles, and the differences are larger for lower absolute

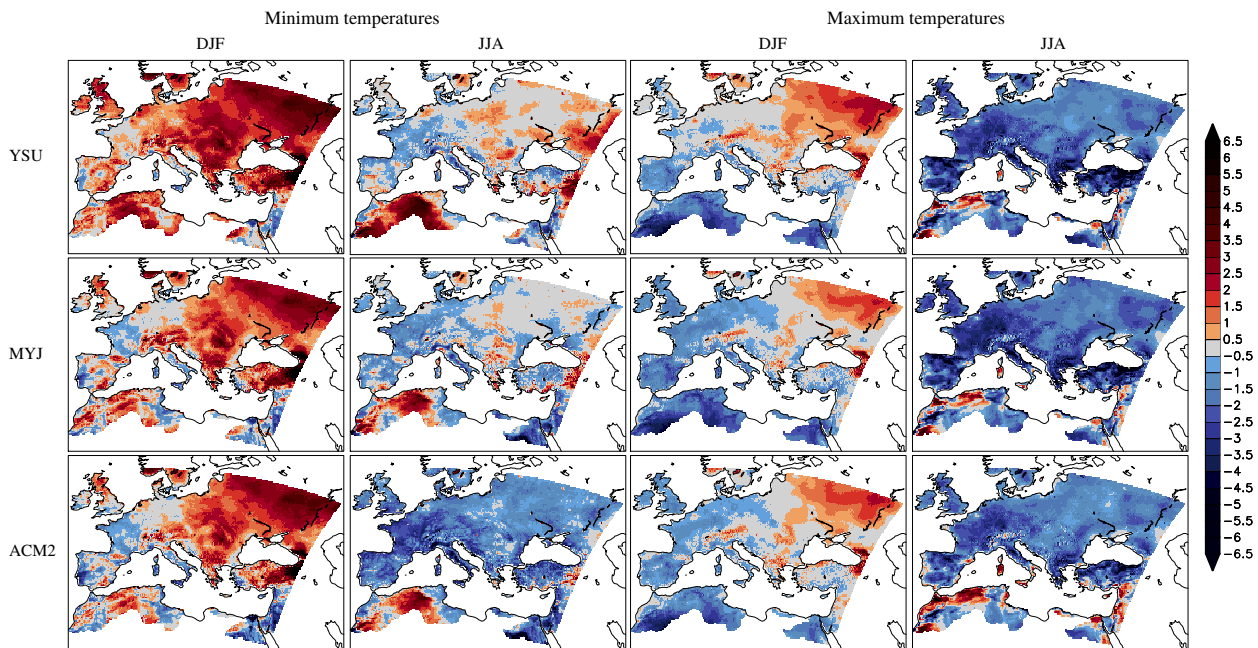


Figure 3. Minimum and maximum temperature bias of WRF compared with E-OBS. Rows are the three PBL schemes used, and columns the seasons, excluded MAM and SON.

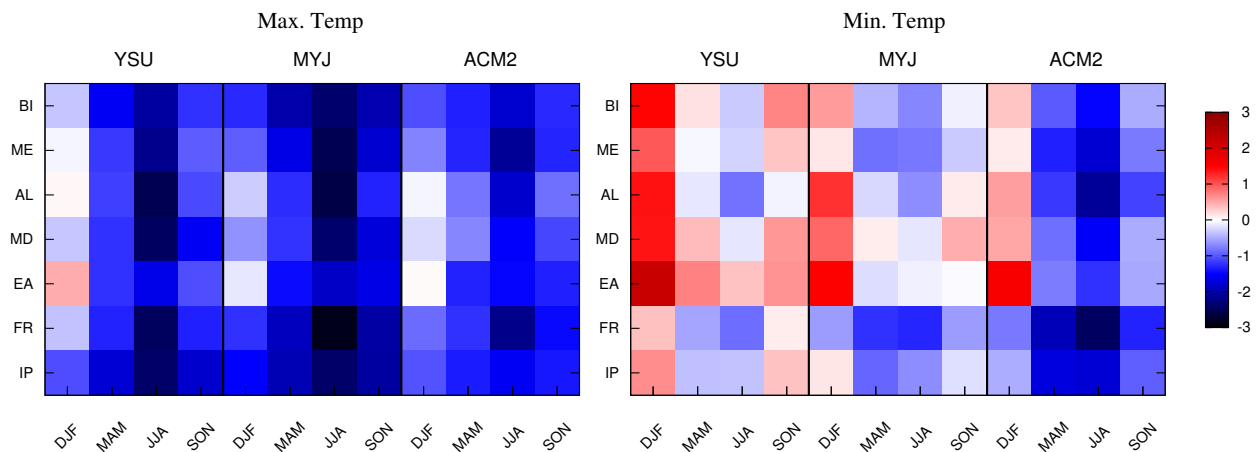


Figure 4. Maximum (left) and minimum (right) temperature bias for each region, season and PBL scheme in °C.

temperatures, being the largest at EA region (note that the plots do not share the same axes, due to the large differences between the regional temperature ranges). In this region the lowest 2.5% quantiles differ by $\approx 7^\circ\text{C}$ which is probably caused by deficient simulation of strong surface temperature inversions. This problem has been addressed before by Mölders and Kramm (2010), who related it to the similarity profiles assumed by the Surface Layer (SL) parametrizations. According to this work, these profiles would not be suitable for extremely stable situations, and SL schemes are in fact imposing a lower limit to the stability.

DJF maximum temperature behaviour differs depending on the region considered. The shape of the q-q plots is similar to the winter minimum temperatures in the cold region EA, but similar to summer maximum temperatures in the warm IP. Again, the deviations between E-OBS and WRF seem to depend on the temperature. JJA maximum temperatures are systematically biased. They follow the same probability distribution both in E-OBS and in WRF, since the points of the qq-plots follow approximately a straight line, and all the quantiles of WRF are displaced

1–2°C to cooler temperatures. JJA minimum temperature behaviour again largely differs from winter. Lower quantiles of WRF tend to be cold biased (opposite to winter) in all regions except for YSU and MYJ at MD. This bias disappears at the upper quantiles in some regions, when minimum temperatures reach 20°C, and the warmer scheme YSU even shows warm biases in 5 of the 7 regions.

3.3. Daily cycle

To obtain a more detailed view of the behaviour of the different PBL schemes, average daily cycles have been computed from SYNOP station data, for both winter and summer. WRF comparison has been performed by seeking for the nearest neighbour. As the domain is large, daily cycles are not synchronized. To address this, they have been shifted in time, by adding or subtracting one hour for each 15 degrees of displacement from the 0 meridian, so the time axis corresponds to the local solar time (Note that the original time of the SYNOP reports is UTC). The difference reaches 2 hours between Ireland and Poland.

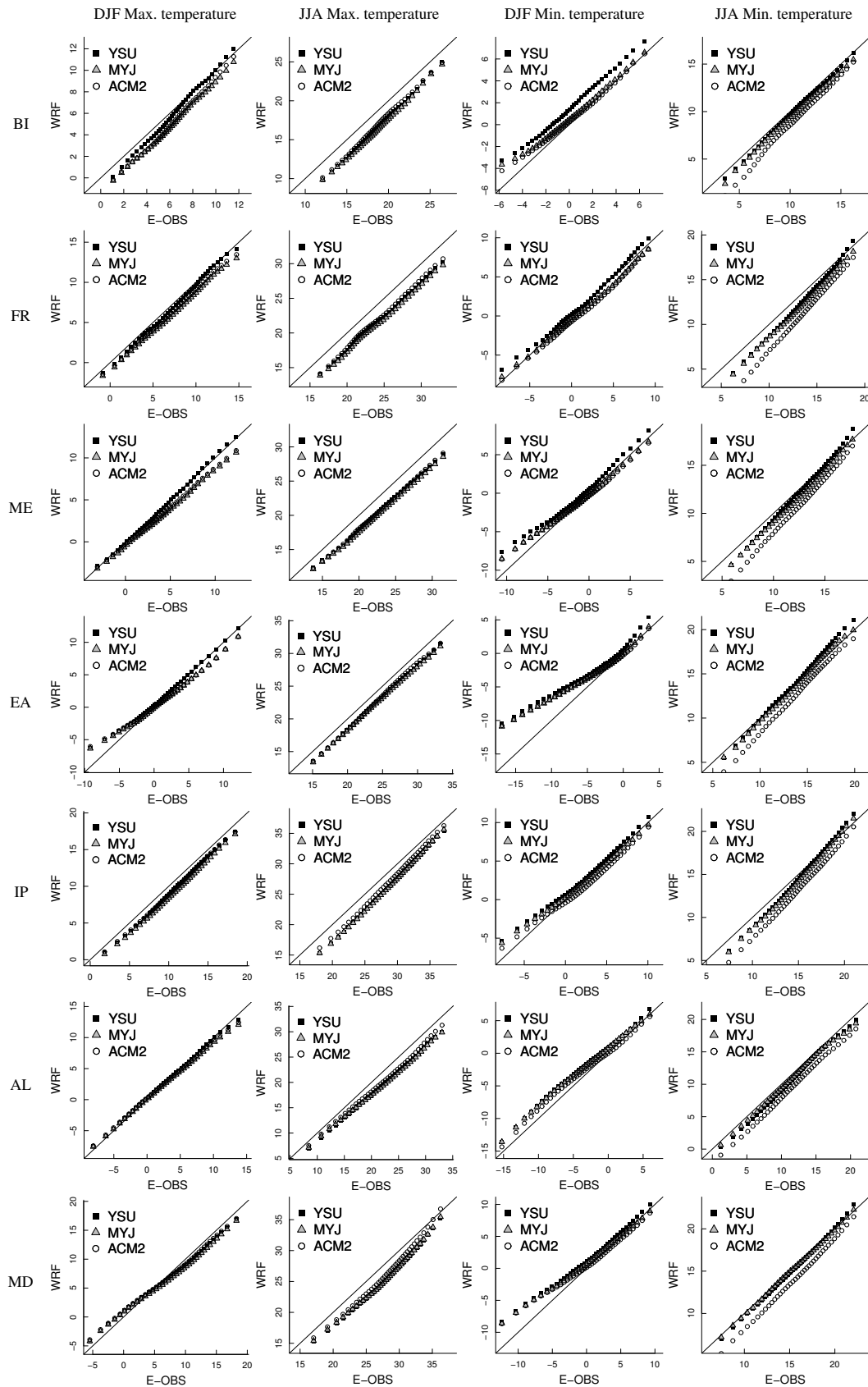


Figure 5. Quantile-quantile plots computed for the 7 regions and for maximum and minimum temperatures.

Figure 6 shows average daily cycles for both screen level temperature and specific humidity, and for winter and summer. As seen with E-OBS, a large cold bias of around 1.8°C is observed for JJA maximum temperatures,

being somewhat larger for MYJ. In contrast, summer night time temperatures are very well reproduced by YSU, and underestimated by around 0.5°C by MYJ and by 1°C by ACM2. As expected from Section 3.2, winter biases are

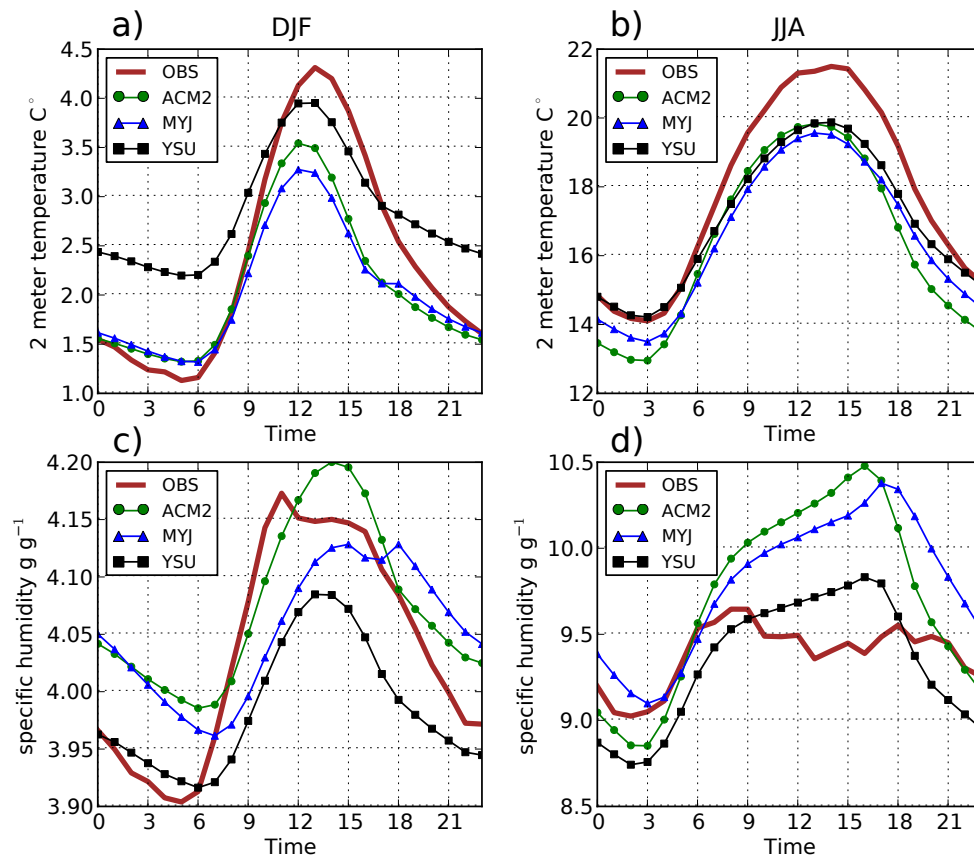


Figure 6. Mean daily cycle of the SYNOP stations compared with WRF for a) winter temperature, b) summer temperature, c) winter specific humidity and d) summer specific humidity.

different. For DJF maximum temperatures YSU shows the smaller bias, followed at some distance by ACM2 and MYJ. But in winter night temperatures, YSU shows a warm bias of 1°C , while MYJ and ACM2 are matching the observations pretty close. Again, results are consistent with chapter 3.2, this was expected, because SYNOP and E-OBS are not independent datasets, as they share the records of many stations, and gives confidence about the additional results obtained with the SYNOP data.

Temperature daily cycles help to interpret the specific humidity daily cycles. Figure 6 shows that WRF overestimates diurnal specific humidity in summer. Moisture budget in the PBL has two main contributions, evapotranspiration, which is a source, and entrainment, which is a sink. Often both contributions cancel each other, and the amplitude of the daily cycle of specific humidity is rather small. This is consistent with the observed amplitudes in Figure 6, of around 0.6 g g^{-1} (JJA) and 0.15 g g^{-1} (DJF). During the summer daytime, WRF tends to overestimate the moisture at the screen level, specially ACM2 and MYJ. This could be related either to excessive evaporation or to too weak entrainment flux. As shown in the case study with Cabauw data (Section 3.6), excessive evaporation seems unlikely, since WRF is not overestimating the latent heat flux. Apart from the different amplitude, observed cycle in summer does not show the peak that the model shows in the evening, on the contrary, observed humidity peaks at early morning. Differences found in winter are rather small ($\approx 0.1\text{ g g}^{-1}$), and fall inside the error of the observations.

The averaging carried out here could be discussed, because very different stations (coastal, continental, etc.) are

being mixed. Again, it must be stressed that the goal of this work is to deal with the gap existing between the detailed analysis of case studies and the generic validation of large datasets such as ENSEMBLES runs.

3.4. Comparison with atmospheric soundings.

Figure 7 shows the mean and the standard deviation of the temperature differences between the model and the soundings for 4 mandatory levels (925, 850, 700 and 500 hPa) and the screen level, in summer and winter. During both seasons warm biases prevail at high levels, 500 hPa and 700 hPa, while differences are found at lower levels. 925 hPa DJF temperature biases are small, distributed around zero, but cold biases of around 0.7°C are found in JJA. This bias is a little more pronounced in MYJ scheme. This implies that the model is generating too cold boundary layers in summer, and it is consistent with the surface cold bias that was found at JJA when comparing with E-OBS and SYNOP data, and its relationship with the lack of entrainment. As winter boundary layers are shallower, 925 hPa temperature may not be representative of its true temperature, or they may not suffer from this bias. Also 850 hPa temperature biases tend to have opposite signs, being positive in winter and negative in summer. Again, the 850 hPa level is probably falling inside the PBL in many of the summer soundings, while it is well above the PBL top in winter.

Regarding specific humidity (Figure 8), as expected, 925 hPa level is found to be too moist, specially in summer and with the local MYJ scheme, reaching $+0.5\text{ g kg}^{-1}$. On

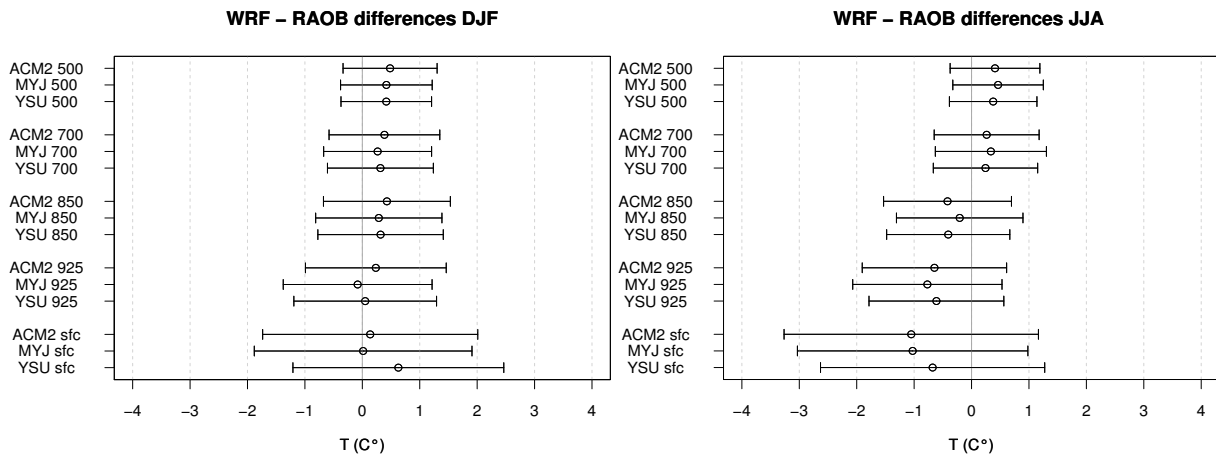


Figure 7. Mean bias and standard deviation computed with the instantaneous temperatures of WRF and atmospheric soundings in winter (left), and in summer (right).

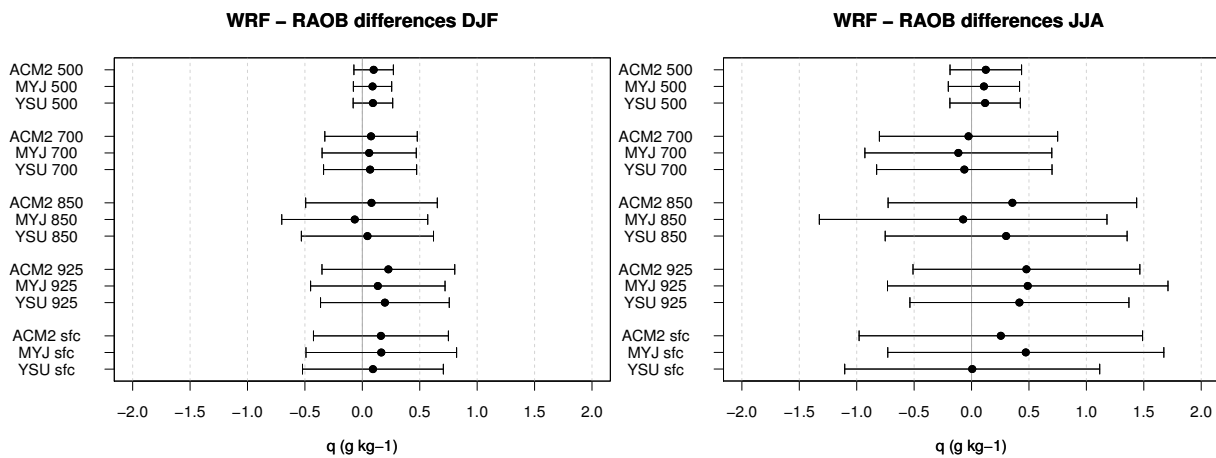


Figure 8. Mean bias and standard deviation computed with the instantaneous specific humidity biases between WRF and soundings in winter (left), and in summer (right).

the other hand, no dry bias is found in upper levels, but it could be hidden between the 850 and 700 hPa levels. As stated before, too cold and moist PBLs can be caused by an underestimation of vertical mixing and entrainment, in the next Section we show that this is probably the case, since WRF also tends to underestimate the PBL thickness.

The analysis of the soundings is affected by some inhomogeneities. Mainly the different height of the stations, and the different solar times for the same UTC time. To check the impact of this point in the previous results, the same plots have been computed using only the sounding stations located in the ME region, and below 200 meters of elevation, but no significant differences were found (not shown).

3.5. PBL top height

Boundary layer height is a good indicator of the turbulent mixing strength and the entrainment [Stull \(1988\)](#). Each scheme computes the PBL top height in very different ways, and this makes them to be not comparable. To solve this, PBL top height has been diagnosed following a simple method both for WRF and for the De Bilt soundings (The Netherlands). The method consists in searching for the first level where potential temperature exceeds the minimum potential temperature reached in the mixed layer by more than 1.5 K. This simple algorithm is shown

to have a good skill compared with others in [Nielsen-Gammon et al. \(2008\)](#), although it may not be well suited for stable boundary layers. Nevertheless, it enables a fair comparison of the 3 schemes among themselves and with observations, at least for well formed mixed layers. [Fig 9](#) shows scatter-plots (green circles), with quantile-quantile plots superimposed (white squares), comparing the PBL tops of WRF and the De Bilt soundings diagnosed using this method. De Bilt was chosen because it has 4 daily soundings available for the year 2001, excepting some missing values. The scatter-plots are noisy, which is expected for a parameter like this, since soundings are frequently unrepresentative of the average air column due to being an instantaneous measurement ([Stull 1984](#)). Thus, quantile-quantile plots are again very useful to study systematic errors in the model. The q-q plots show that MYJ and ACM2 tend to produce lower PBL tops than observed, especially when PBL tops exceed 700 m. In contrast, YSU scheme shows a small overestimation of the PBL tops lower than 1000 m and, above that it also underestimates the observed values, but still gets closer than the other 2 schemes. This supports the hypothesis of a general underestimation of the entrainment by ACM2 and MYJ, while YSU would only suffer from that problem for thicker PBLs. Again, different days are being mixed, and the results could be different, for example, for thick PBLs driven by strong winds than those driven by strong

diabatic heating. But the goal of the paper is to see how the signal of the different biases emerges when applying generic statistical analysis to large amount of data.

3.6. Cabauw case study

Finally, a few days with stable clear sky are studied with data from the CESAR observatory. The goal here is to check how the observed general biases are reproduced by a single station, and to analyse variables not usually available, such as the surface turbulent fluxes. Although a profound analysis is out of the scope of the paper, this could help to confirm the role of the entrainment in the biases, and to see if there is any deficiency in the simulation of the fluxes behind. In Figure 10 (top) simulated and observed temperatures are plotted in the CESAR tower location the days 3-7 of July (note that the land surface scheme used here is the Noah-LSM, as mentioned in Section 2). This period has been chosen because it was a warm, dry and stable period, dominated by the presence of a high pressure system, being a good example of the cold bias that we want to analyse. At night time YSU scheme stays close to observed temperatures, while ACM2 and MYJ still show cold biases. During daytime YSU is slightly warmer than ACM2 and MYJ, but the three schemes stay well below the observed temperature. In general, the best match is achieved by YSU, and the results are similar to those found in Section 3.3 with SYNOP data, and those obtained by Hu *et al.* (2010).

Noah LSM is known to cause cold bias in the last versions of WRF-ARW. According to Manning *et al.* (2010), this cold bias is caused by problems with the seasonal variation of parameters like albedo. Figure 10 (middle) shows the simulated and observed surface turbulent fluxes in CESAR for the period of interest. We find that, during daytime, both fluxes tend to be underestimated by 50-100 W m^{-2} . This is partly explained because the observed fluxes are in fact underestimated by $\sim 30\%$ (Braam 2008). Thus, fluxes do not seem to contribute to the cold bias observed in the screen level temperature. Other candidates are the radiative balance and the entrainment. We have seen various results pointing to an underestimation of the entrainment (Sections 3.4 and 3.3), but a stronger link with surface temperatures is missing. Fig 10 (bottom) shows the PBL height computed with the algorithm explained in Section 3.5. Although WRF underestimates it at noon, YSU produces thicker PBLs than ACM2 and MYJ, but very similar screen level temperatures. This can also be observed in Figure 6. Thus, we conclude that the lack of entrainment is not playing a mayor role in the observed cold bias. What is left is the radiative balance, and there are a few works pointing that the impact of this factor in the mixed layer temperature is important (Steenefeld *et al.* 2010; Lemone *et al.* 2002). A detailed analysis of this balance is complex, and falls out of the scope of this paper, but it is recommended for the community of WRF users.

Vertical profiles at 12 UTC of the July 5th (Figure 11a) show that the observed boundary layer is warmer and drier and more mixed than the simulated one. The same happens at night hours, (Figure 11b) shows the modelled temperatures together with the De Bilt sounding at 00 UTC of July 5th. Again, the observed profile is warmer and more mixed than the simulated one, also in the YSU scheme, that has a very small bias at surface. This is a good example of the underestimation of the entrainment found before in this

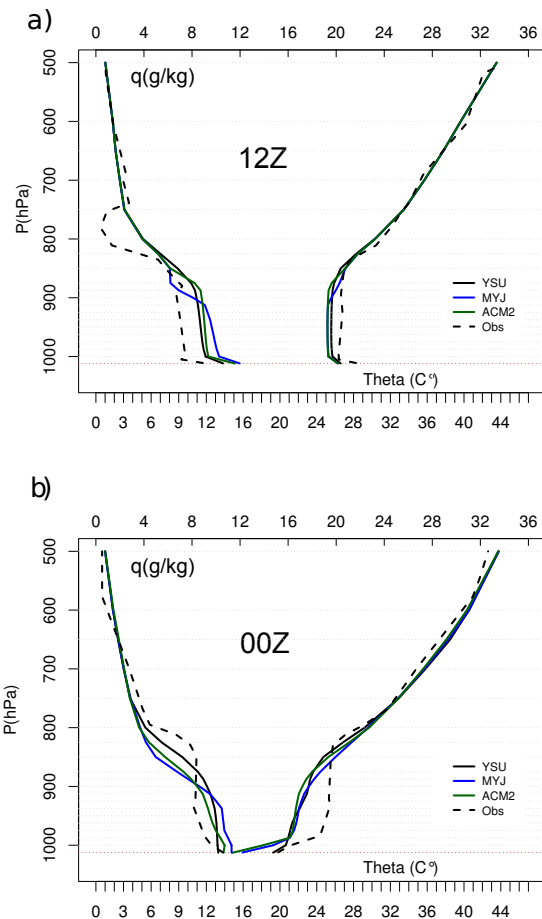


Figure 11. Simulated and observed potential temperature (lines on the right) and specific humidity (lines on the left) profiles in De Bilt, The Netherlands, at 7-5-2001. At a) 12 UTC with Noah and b) 00 UTC with Noah.

study. It also shows how a model can precisely reproduce surface temperatures, but at the same time give a wrong vertical profile (YSU).

4. Conclusions

A set of three 1-year simulations varying the PBL scheme have been carried out over Europe. The regional model has been restarted daily to keep it close to the driving reanalysis (ERA-Interim). It has been shown, by comparing with the gridded, SYNOP and sounding data, that model biases are dependent on the season, geographical location and time of the day. Furthermore, this work is an example of how model biases found in "statistical studies" (where, often, physical causes are hard to uncover due to the mixture of different processes or due to the lack of detailed observations for a long enough period) can be traced down to their seasonal and daily timescales, and connected with the results of case studies.

A systematic cold bias, as compared to the E-OBS dataset, has been found during the warm season. This bias affects equally all the range of maximum temperatures. Contrarily, a warm bias is apparent during the cold season, as a result of the overestimation of the lowest minimum temperatures. This warm bias in winter cold extremes is probably related to unresolved strong temperature inversions and to the limits imposed to the stability existing in the surface layer parametrizations. The classical

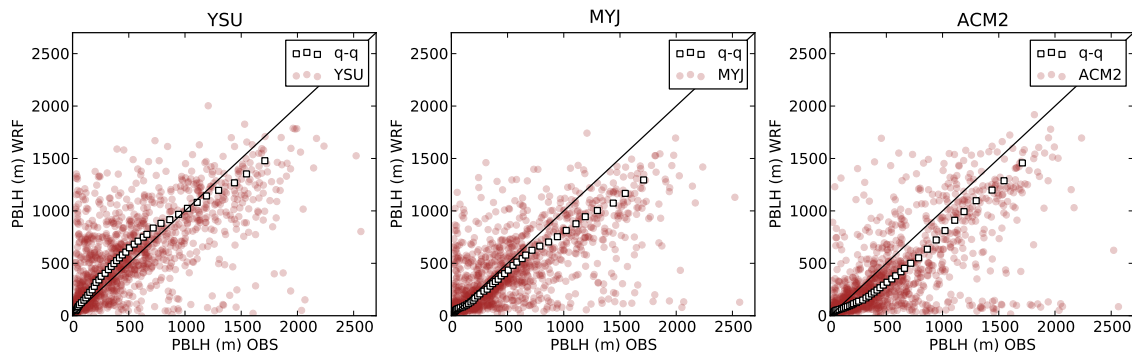


Figure 9. Scatter-plots comparing the PBL height observed in the De Bilt soundings with the one simulated by WRF.

similarity functions used in these surface layer schemes tended to excessive cooling in very stable environments, and the adjustments made to solve this (e.g. in ACM2; Pleim 2006) seem to be excessive for the case of strong winter inversions. The observed dependence of the bias with absolute temperature and, therefore, with the seasonal cycle, is not particular to WRF. A similar behaviour has been reported in other regional models. For example, the RCMs used in the EU-funded project ENSEMBLES (Christensen *et al.* 2008). Plavcová and Kyselý (2011) also found this dependence with the same set of models in a study focused over the Czech Republic. These common deficiencies are not surprising, since many models share common building blocks (Fernández *et al.* 2009; Knutti 2010), especially regarding parametrization schemes.

Model biases are usually larger than the spread of the ensemble of PBL schemes tested. The differences among the schemes show little change along the year, with YSU consistently showing warmer temperatures. Thus, in statistical terms, the scheme performing closer to the observations depends on the season and is probably a result of error cancellation. On the other hand, the presence of the same signatures on the bias of regions with very different land cover, orography, climate, and observation density/coverage, indicates that the results concerning to the model biases and the relative temperature of the different PBL schemes are robust to these features.

Simulated daily cycles have been compared using SYNOP data (Section 3.3), and it has been shown that the most relevant differences among the schemes occur at night. As reported by Hu *et al.* (2010), the YSU scheme has been found to be close to observations during summer nights but, interestingly, the opposite does occur in winter. This points to YSU as the best choice for studies focused on summer periods, but ACM2 or MYJ would be better suited in winter. Thus, results of “case studies” must be used with caution when choosing the model configuration for longer term simulations.

Entrainment has been indirectly studied through the daily cycle of specific humidity, the comparison with sounding data and the diagnosis of the PBL top height. Results show that WRF tends to underestimate the entrainment for thick PBLs (above 1000 m), as it produces a too large specific humidity daily cycle, and a too moist and cold 925 hPa level. However, YSU does produce significantly thicker PBLs than the other 2 schemes. This lack of entrainment does not seem to be the cause of the cold bias of summer daily temperatures because, despite of the increased entrainment, YSU does show very similar screen level temperatures.

Finally, the Cabauw case study shows that model-simulated turbulent sensible and latent heat fluxes are overestimated (see also Steeneveld *et al.* 2011), although observed fluxes suffer from underestimation themselves (Braam 2008). Given that the overestimation of the surface fluxes should shift the model results towards warmer temperatures, errors in the estimation of the fluxes are also discarded as the cause of the summer cold bias. The only explanation left is that the errors are coming from the radiative balance, as proposed by Manning *et al.* (2010). More research in this area is encouraged to confirm this hypothesis, but it is out of the scope of our analyses.

The main conclusion of the present work is that model systematic errors are very dependent on the seasonal and daily cycles, and thus on the different atmospheric conditions that prevail on the different seasons and times of the day. This must be considered when selecting the most adequate the model configuration, when improving the parametrization schemes and when performing statistical bias correction for regional climate modelling studies.

Acknowledgement

The E-OBS dataset was produced within the EU-FP6 project ENSEMBLES (<http://www.ensembles-eu.org>) and the data was provided through the ECA&D project (<http://eca.knmi.nl>). We thank Earth System Research Laboratory (<http://www.esrl.noaa.gov>) for making available the sounding data and Cabauw Experimental Site for Atmospheric Research (<http://www.cesar-database.nl>) for their surface temperature data. The authors acknowledge financial support from the Spanish Ministry of Science and Innovation (CORWES CGL2010-22158-C02-01 and EXTREMILES CGL2010-21869 projects) and *Ministerio de Medio Ambiente, Rural y Marino* (ESCENA project, 200800050084265).

References

- Alapaty K, Pleim J, Raman S, Niyogi D, Byun D. 1997. Simulation of atmospheric boundary layer processes using local-and nonlocal-closure schemes. *Journal of Applied Meteorology* **36**(3).
- Argüeso D, Hidalgo-Muñoz J, Gámiz-Fortis S, Esteban-Parra M, Dudhia J, Castro-Díez Y. 2011. Evaluation of WRF parameterizations for climate studies over southern Spain using a multi-step regionalization. *Journal of Climate* doi:10.1175/JCLI-D-11-00073.1.
- Awan N, Truhetz H, Gobiet A. 2011. Parameterization induced error-characteristics of MM5 and WRF operated in climate mode over the Alpine region: An ensemble based analysis. *Journal of Climate* **24**(12): 3107–3123, doi:10.1175/2011JCLI3674.1.

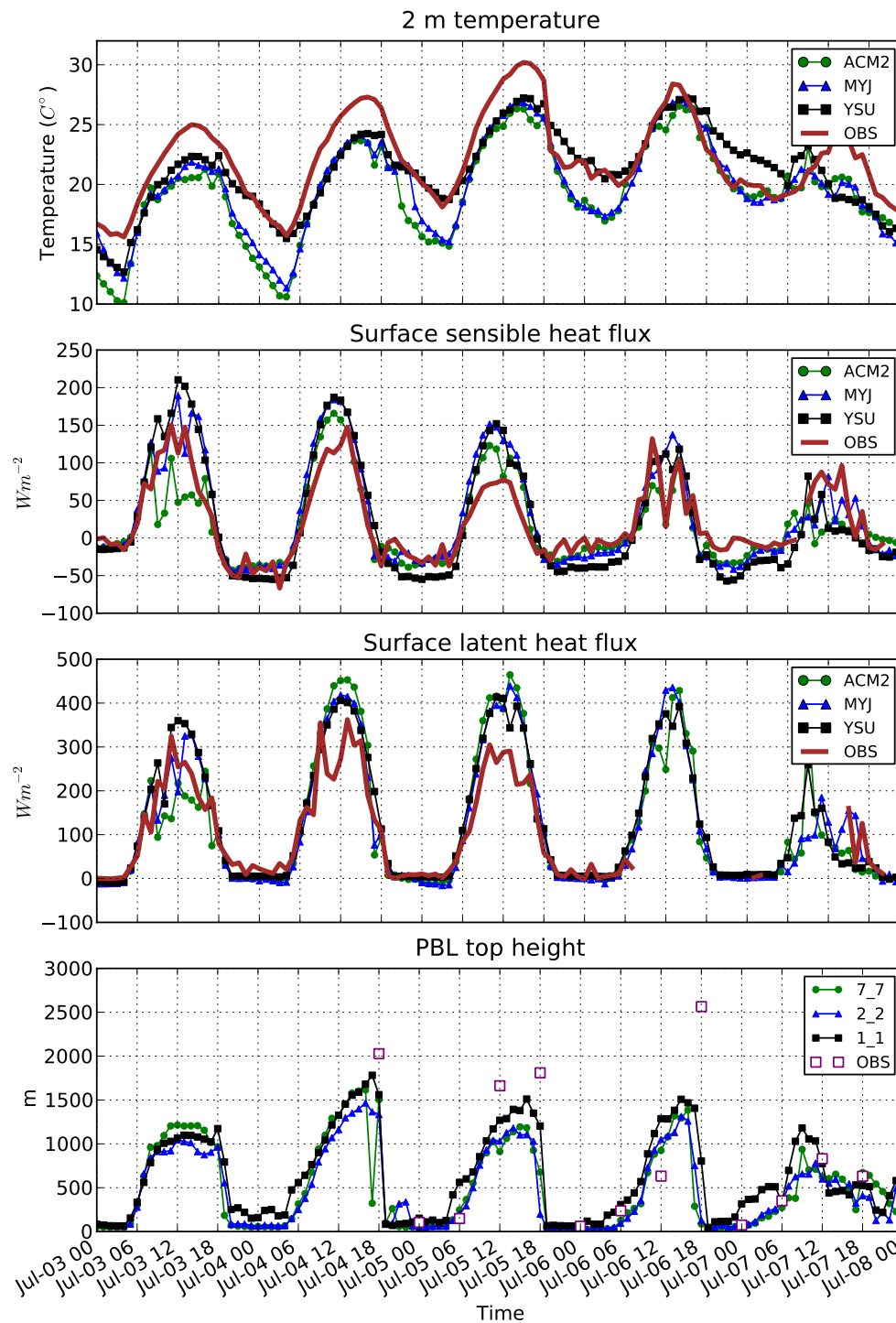


Figure 10. Hourly time series of the measured and simulated 2 meter temperatures, sensible and latent heat fluxes and PBL top height at the CESAR location, for the July 3-7th 2001. Sounding data used to compute PBL top height is taken from De Bilt.

Beljaars A. 1995. The parametrization of surface fluxes in large-scale models under free convection. *Quarterly Journal of the Royal Meteorological Society* **121**(522): 255–270, doi:10.1002/qj.49712152203.

Blackadar A. 1978. Modeling pollutant transfer during daytime convection. In: *Symposium on Turbulence, Diffusion, and Air Pollution*, 4 th, Reno, Nev. pp. 443–447.

Braam M. 2008. *Determination of the surface sensible heat flux from the structure parameter of temperature at 60 m height during day-time.* Koninklijk Nederlands Meteorologisch Instituut.

Braun S, Tao W. 2000. Sensitivity of high-resolution simulations of hurricane Bob (1991) to planetary boundary layer parameterizations. *Monthly weather review* **128**(12): 3941–3961, doi:10.1175/1520-0493(2000)129<3941:SOHRSO>2.0.CO;2.

Bright DR, Mullen SL. 2002. The sensitivity of the numerical simulation of the southwest monsoon boundary layer to the choice of PBL turbulence parameterization in MM5. *Weather and Forecasting* **17**(1): 99–114.

Chen F, Dudhia J. 2001. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review* **129**(4): 569–585.

- Christensen J, Boberg F, Christensen O, Lucas-Picher P. 2008. On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters* **35**(20): L20 709, doi:10.1029/2008GL035694.
- Christensen J, Christensen O. 2007. A summary of the prudence model projections of changes in European climate by the end of this century. *Climatic Change* **81**: 7–30, doi:10.1007/s10584-006-9210-7.
- Deardorff J. 1966. The counter-gradient heat flux in the lower atmosphere and in the laboratory. *Journal of Atmospheric Sciences* **23**: 503–506, doi:10.1175/1520-0469(1966)023<0503:TCGHFI>2.0.CO;2.
- Deng A, Stauffer D. 2006. On improving 4-km mesoscale model simulations. *Journal of Applied Meteorology* **45**(3): 361–381.
- Dudhia J. 1989. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences* **46**(20): 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.
- Dyer A, Hicks B. 1970. Flux-gradient relationships in the constant flux layer. *Quarterly Journal of the Royal Meteorological Society* **96**(410): 715–721, doi:10.1002/qj.49709641012.
- Evans J, Ekström M, Ji F. 2011. Evaluating the performance of a WRF physics ensemble over south-east Australia. *Climate Dynamics* : 1–18, doi:10.1007/s00382-011-1244-5.
- Fernández J, Montávez J, Sáenz J, González-Rouco J, Zorita E. 2007. Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: Annual cycle. *Journal of geophysical research* **112**(D4): D04 101, doi:10.1029/2005JD006649.
- Fernández J, Primo C, Cofino A, Gutiérrez J, Rodríguez M. 2009. MVL spatiotemporal analysis for model intercomparison in EPS: application to the DEMETER multi-model ensemble. *Climate dynamics* **33**(2): 233–243, doi:10.1007/s00382-008-0456-9.
- Fernández-Quiruelas V, Fita L, Fernández J, Cofiño A. 2010. WRF workflow on the Grid with WRF4G. In: *11th WRF Users' Workshop. Boulder (CO), USA*.
- Flaounas E, Bastin S, Janicot S. 2011. Regional climate modelling of the 2006 west african monsoon: sensitivity to convection and planetary boundary layer parameterisation using wrf. *Climate Dynamics* **36**(5): 1083–1105, doi:10.1007/s00382-010-0785-3.
- Grell G, Devenyi D. 2002. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters* **29**(14): 38–1, doi:10.1029/2002GL015311.
- Haylock M, Hofstra N, Tank A, Klok E, Jones P, New M. 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research* **113**(D20): D20 119.
- Herrera S, Fita L, Fernández J, Gutiérrez J. 2010. Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel simulations over Spain. *Journal of Geophysical Research* **115**(D21): D21 117, doi:10.1029/2010JD013936.
- Hewitt C, Griggs D. 2004. Ensembles-based predictions of climate changes and their impacts (ENSEMBLES). *Eos* **85**(52): 566.
- Hofstra N, Haylock M, New M, Jones PD. 2009. Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *Journal of Geophysical Research* **114**, doi:10.1029/2009JD011799.
- Hong S, Lim J. 2006. The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc* **42**(2): 129–151.
- Hong S, Noh Y, Dudhia J. 2006. A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review* **134**(9): 2318–2341.
- Hu XM, Nielsen-Gammon JW, Zhang F. 2010. Evaluation of three planetary boundary layer schemes in the WRF model. *Journal of Applied Meteorology and Climatology* **49**(9): 1831–1844, doi:10.1175/2010JAMC2432.1.
- Janjić Z. 2002. Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model. *NCEP Office Note* **437**: 61.
- Janjić ZI. 1990. The step-mountain coordinate: Physical package. *Mon. Weather Rev.* **118**: 1429–1443.
- Janjić ZI. 1994. The step-mountain eta coordinate model: Further developments of the convection. *Mon. Weather Rev.* **122**: 927–945.
- Jerez S. 2011. *Climate simulations over the Iberian peninsula: study of the role of the parameterisation schemes and characterisation of climate change patterns*. PhD Thesis. Universidad de Murcia. Spain.
- Jerez S, Montavez J, Jimenez-Guerrero P, Gomez-Navarro J, Lorente R, Zorita E. 2012. Assessment of the role of the parameterization schemes from a multi-physics ensemble of present-day climate simulations over the Iberian Peninsula. *Climate Dynamics* : Submitted.
- Knutti R. 2010. The end of model democracy? *Climatic change* **102**(3): 395–404, doi:10.1007/s10584-010-9800-2.
- Kotroni V, Lagouvardos K. 2001. Precipitation forecast skill of different convective parameterization and microphysical schemes: Application for the cold season over Greece. *Geophysical Research Letters* **28**(10): 1977–1980, doi:10.1029/2000GL012705.
- Kysely J, Plavcova E. 2010. A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations. *Journal of Geophysical Research-Atmospheres* **115**, doi:10.1029/2010JD014123.
- Lawrence M. 2005. The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bulletin of the American Meteorological Society* **86**: 225–233, doi:10.1175/BAMS-86-2-225.
- Lemone M, Grossman R, Mcmillen R, Liou K, Ou S, Mckeen S, Angevine W, Ikeda K, Chen F. 2002. CASES-97: Late-morning warming and moistening of the convective boundary layer over the Walnut River watershed. *Boundary-Layer Meteorology* **104**(1): 1–52, doi:10.1023/A:1015569104180.
- Lenderink G, van Meijgaard E, Selten F. 2009. Intense coastal rainfall in the Netherlands in response to high sea surface temperatures: analysis of the event of August 2006 from the perspective of a changing climate. *Climate Dynamics* **32**(1): 19–33, doi:10.1007/s00382-008-0366-x.
- Louis JF. 1979. A parametric model of vertical eddy fluxes in the atmosphere. *Boundary-Layer Meteorology* **17**(2): 187–202, doi:10.1007/BF00117978.
- Manning K, Tewari M, Barlage M, Chen F, Salamanca F, Martill A. 2010. Updates to the Noah LSM in WRF version 3.2. In: *11th WRF Users' Workshop. Boulder (CO), USA*.
- Mellor GL, Yamada T. 1982. Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.* **20**: 851–875.
- Mlawer E, Taubman S, Brown P, Iacono M, Clough S. 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research* **102**(D14): 16 663–16, doi:10.1029/97JD00237.
- Mölders N, Kramm G. 2010. A case study on wintertime inversions in interior Alaska with WRF. *Atmospheric Research* **95**(2-3): 314–332, doi:10.1016/j.atmosres.2009.06.002.
- Monin A, Obukhov A. 1954. Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci., USSR, (151)* : 163–187.
- Nielsen-Gammon J, Powell C, Mahoney M, Angevine W, Senff C, White A, Berkowitz C, Doran C, Knupp K. 2008. Multisensor estimation of mixing heights over a coastal city. *Journal of Applied Meteorology and Climatology* **47**(1): 27–43, doi:10.1175/2007JAMC1503.1.
- Nikulin G, Jones C, Samuelsson P, Giorgi F, Sylla M, Asrar G, Büchner M, Christensen O, Déqué M, Fernandez J, et al. 2012. Precipitation climatology in an ensemble of CORDEX-Africa regional climate simulations. *Journal of Climate* : In print.
- Otkin J, Greenwald T. 2008. Comparison of WRF model-simulated and MODIS-derived cloud data. *Monthly Weather Review* **136**(6): 1957–1970, doi:10.1175/2007MWR2293.1.
- Paulson C. 1970. The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *Journal of Applied Meteorology* **9**: 857–861, doi:10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2.
- Persson P, Walter B, Bao J, Michelson S. 2001. Validation of boundary-layer parametrizations in a marine storm using aircraft data. *Ninth Conf. on Mesoscale Processes* : 117–121.
- Plavcová E, Kysely J. 2011. Evaluation of daily temperatures in Central Europe and their links to large-scale circulation in an ensemble of regional climate models. *Tellus A* doi:10.1111/j.1600-0870.2011.00514.x.

- Pleim J. 2006. A simple, efficient solution of flux-profile relationships in the atmospheric surface layer. *Journal of applied meteorology and climatology* **45**(2): 341–347, doi:10.1175/JAM2339.1.
- Pleim J, Chang J. 1992. A non-local closure model for vertical mixing in the convective boundary layer. *Atmospheric Environment. Part A. General Topics* **26**(6): 965–981, doi:10.1016/0960-1686(92)90028-J.
- Pleim JE. 2007a. A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *Journal of Applied Meteorology and Climatology* **46**(9): 1383–1395.
- Pleim JE. 2007b. A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. *Journal of Applied Meteorology and Climatology* **46**(9): 1396–1409.
- Shin H, Hong SY. 2011. Intercomparison of planetary boundary-layer parametrizations in the WRF model for a single day from CASES-99. *Boundary-Layer Meteorology* **139**: 261–281, doi:10.1002/qj.606. 10.1007/s10546-010-9583-z.
- Simmons A, Uppala S, Dee D, Kobayashi S. 2007. ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter* **110**: 25–35.
- Skamarock W, Klemp J, Dudhia J, Gill D, Barker D, Duda M, Wang W, Powers J. 2008. A description of the advanced research wrf version 3. Technical report, NCAR.
- Steenefeld G, Tolck L, Moene A, Hartogensis O, Peters W, Holtslag A. 2011. Confronting the WRF and RAMS mesoscale models with innovative observations in the Netherlands. Evaluating the boundary-layer heat budget. *Journal of Geophysical Research. D, Atmospheres* doi:10.1029/2011JD016303.
- Steenefeld G, Wokke M, Groot Zwaafink C, Pijlman S, Heusinkveld B, Jacobs A, Holtslag A. 2010. Observations of the radiation divergence in the surface layer and its implication for its parameterization in numerical weather prediction models. *J. Geophys. Res* **115**: D06 107, doi:10.1029/2009JD013074.
- Stensrud D. 2007. *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge Univ Pr.
- Stensrud D, Weiss S. 2002. Mesoscale model ensemble forecasts of the 3 may 1999 tornado outbreak. *Weather and Forecasting* **17**(3).
- Stull R. 1984. Transient turbulence theory. I—the concept of eddy-mixing across finite distances. *Journal of the atmospheric sciences* **41**: 3351–3367, doi:10.1175/1520-0469(1984)041<3351:TTTPIT>2.0.CO;2.
- Stull R. 1988. *An introduction to boundary layer meteorology*. Springer.
- Stull RB. 1991. Static stability—an update. *Bulletin of the American Meteorological Society* **72**(10): 1521–1529.
- Webb E. 1970. Profile relationships: The log-linear range, and extension to strong stability. *Quarterly Journal of the Royal Meteorological Society* **96**(407): 67–90, doi:10.1002/qj.49709640708.
- Weisman ML, Davis C, Wang W, Manning KW, Klemp JB. 2008. Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Weather and Forecasting* **23**(3): 407–437, doi:10.1175/2007WAF2007005.1.
- Yagüe C, Viana S, Maqueda G, Redondo J. 2006. Influence of stability on the flux-profile relationships for wind speed, ϕ_m , and temperature, ϕ_h , for the stable atmospheric boundary layer. *Nonlinear Processes in Geophysics* **13**: 185–203.
- Zhang D, Zheng W. 2004. Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations. *Journal of Applied Meteorology* **43**(1): 157–169.
- Zilitinkevich S. 1995. Non-local turbulent transport: Pollution dispersion aspects of coherent structure of convective flows. *Air Pollution III* **1**: 53–60.