Report on the objective calibration of COSMO 5.0 within the COPAT project.

Daniel Lüthi, Institute for Atmospheric and Climate Science, ETH Zürich

Silje Sørland, Institute for Atmospheric and Climate Science, ETH Zürich

Katherine Osterried, Center for Climate System Modelling, ETH Zürich

Omar Bellprat, Institut Catala de Ciènces del Clima (IC3), Barcelona, Spain

Overview

The objective calibration was the last step taken in the COPAT project to find a recommended configuration of the COSMO-CLM regional climate model that was based on COSMO 5.0 after, in earlier phases of the project, model options had been tested systematically to come to a set of options that produced the most promising results, i.e. climatology with the smallest biases.

The objective calibration with the method of Bellprat et al. (2012) consists of the 5 main steps listed here:

- Choice of the sensitive tuning parameters which will be used in the calibration
- Performing the simulations over the range of the tuning parameters.
- Building of a meta-model
- Definition of a performance metric
- Exploration of parameter space for optimal set of parameters

The individual steps of calibration will be addressed separately in the sections below

Choice of sensitive tuning parameters

The choice of the sensitive tuning parameters is based on the work of Bellprat et al. (2016). Four of the parameters used for the calibration were preexisting tuning parameters, two were model internal parameters and two were introduced especially for the calibration by changing the code of the COSMO model. This new model version became COSMO5.0_clm5. All the parameters can be set in the “TUNING” block of the namelist input file INPUT_ORG.

As can be seen the tuning parameters are from different parameterization schemes and should therefore be independent.

The following table lists the parameters are used for the calibration, the values given in the Values column are [minimum value; default value; maximum value]:

Parameter name	Description of the parameter	Values
rlam_heat	Scalar resistance for sensible and latent heat fluxes in the laminar surface layer	[0.1;**1**;5]
entr_sc	Entrainment rate for shallow convection [m^{-3}]	[$3*10^{-5}$;$\mathbf{3*10^{-4}}$;$3*10^{-3}$]
tkhmin	Minimal vertical turbulent diffusion rate [m^2s^{-1}]	[0.1; **1**; 2]
qi0	Threshold for conversion cloud ice to snow	[**0** ;$5*10^{-5}$;10^{-4}]
uc1	Parameter controlling the vertical variation of critical relative humidity for sub-grid cloud formation	[0; **0.8**; 1.6]
radfac	Fraction of cloud water and ice considered by radiation scheme	[0.3; **0.5**; 0.9]
fac_rootdp2	Uniform factor for root depth field	[0.5; **1**; 1.5]
soilhyd	Factor for the hydraulic conductivity and diffusivity	[0.1;**1**;10]

Performing the simulation with changed parameters

The simulation necessary for the calibration consist of a reference simulation, simulations where 1 parameter is changed to either the minimum or maximum value of the physical meaningful range and simulations where two parameters are changed to either of the extreme values. Each of those simulation is ERAinterim driven and covers the years 1979-1983. All simulations use the same initial and boundary data. Initial soil moisture is taken from a twenty-year long climatology from a previous simulation with the reference configuration. The total number of simulations that need to be performed for such a setup with N parameters is as follows :

1. Reference simulation with all parameter at the default value (1 simulation)
2. Sensitivity simulations with one of the parameters set to either extreme value (2*N simulations)
3. Sensitivity simulations where 2 different parameters are set to either extreme value (N*(N-1)/2 pairs with 4 possible combinations => 2*N*(N-1) simulations)
4. A total of 1 + 2*N*N simulations (for N=8: **129 simulations**)

For each of these simulations the target values (monthly mean values for the climate parameters **T2M**, **TOT_PREC** and **CLCT** averaged over the 8 PRUDENCE domains) are determined in a post-processing step. That is 5(years) * 12(months) * 8(regions) * 3(climate parameters) = 1440 values.

Building of a meta-model

The next step is to build a meta-model that estimates the values of the climate parameters for each month and each PRUDENCE region as a function of the values of the calibration parameters, where the values of the calibration parameters are within the predefined range. In fact there are 1440 individual meta-models that have the same functional structure. They are all calculated as a quadratic deviation from the reference state (all parameters assume their uncalibrated default values). For a calibration with 3 tuning parameters $\mu 1$, $\mu 2$ and $\mu 3$ this looks as follows:

$$\theta(\mu 1, \mu 2, \mu 3) = \theta_{ref} + a1 \cdot \mu 1 + a2 \cdot \mu 2 + a3 \cdot \mu 3 + b1 \cdot \mu 1^2 + b2 \cdot \mu 2^2 + b3 \cdot \mu 3^2 + c1 \cdot \mu 1 \cdot \mu 2 + c2 \cdot \mu 1 \cdot \mu 3 + c3 \cdot \mu 2 \cdot \mu 3$$

The coefficients a1, a2 and a3 (linear terms), b1, b2 and b3 (quadratic terms) and c1, c2 and c3 (interaction terms) need then to be calculated based on the values of the simulations with the changed parameters. For N tuning parameters the number of unknown coefficients is 2*N + N*(N-1)/2 . For 8 calibration parameters this amounts to 44 unknown coefficients. With the 128 simulations with changed tuning parameters we have an over-determined linear system that needs to be solved. A least square error method is used for this purpose.

Performance metric

To evaluate the individual simulations a performance metric, which delivers a single value for each simulation, is necessary. We use here the same performance metric that was also applied in the work of Bellprat et al. (2016) for the European continent. However, we are evaluating temperatures and precipitation against the E-OBS version 10 0.44deg rotated data-set, which matches the grid of the CORDEX simulation. For the evaluation of the cloud-cover data we use the same gridded data-set originating from CRU as was used in Bellprat et al. (2016). The difficulty is to bring together model bias values of different physical quantities and for different seasons. The approach taken here is, for each month of the simulation and each PRUDENCE region, to normalize the squared difference model - observation by the total uncertainties given by an estimate of the observation error, the interannual variability and an estimate of the internal variability of the model simulation. In mathematical language:

$$PI^2 = \frac{1}{12 \cdot R \cdot CV \cdot Y} \sum_{m=1}^{12} \sum_{r=1}^{R} \sum_{c=1}^{CV} \sum_{y=1}^{Y} \frac{(mod(y,m,r,c) - obs(y,m,r,c))^2}{(\sigma_{iv}(m,r,c) + \sigma_{iav}(m,r,c) + \sigma_{oerr}(m,r,c))^2}$$

Where r runs over all PRUDENCE regions, y over all 5 simulated years, m over the 12 months of the year and cover all 3 evaluated climate parameters T_2M, TOT_PREC and CLCT. The uncertainty estimates that appear in the normalization term are as follows:

$\sigma_{iv}(m,r,c)$: estimate of internal variability based on 6 member ensemble covering 180 years

$\sigma_{iav}(m,r,c)$: interannual variability based on E-OBS data for corresponding month, region and climate parameter.

$\sigma_{oerr}(m,r,c)$: estimate of observational uncertainty based on estimated values given by E-OBS or differences between independent data sets (CLCT).

To cast the score to the range 0 to 1, where 1 would be the perfect score we use for the final performance score

$$PS = e^{-0.5 \cdot PI^2}$$

Exploration of the parameter space to find optimal settings

The Matlab function ***lhsdesign*** with the criterion "correlation" was used to produce a set of 5 million combinations for the tuning parameters to sample the parameter space for the optimal combination of parameters. Figure 1 shows the distribution of the achieved PS score for all of the 5 million sampled combinations. As can be seen in Figure 1 only a relatively small fraction (0.15%) of the simulations achieve a better score than the reference simulation. The skill score of the reference simulation has a value of 0.8790, whereas the highest predicted value by the meta-model in the sample amounts to 0.9047.

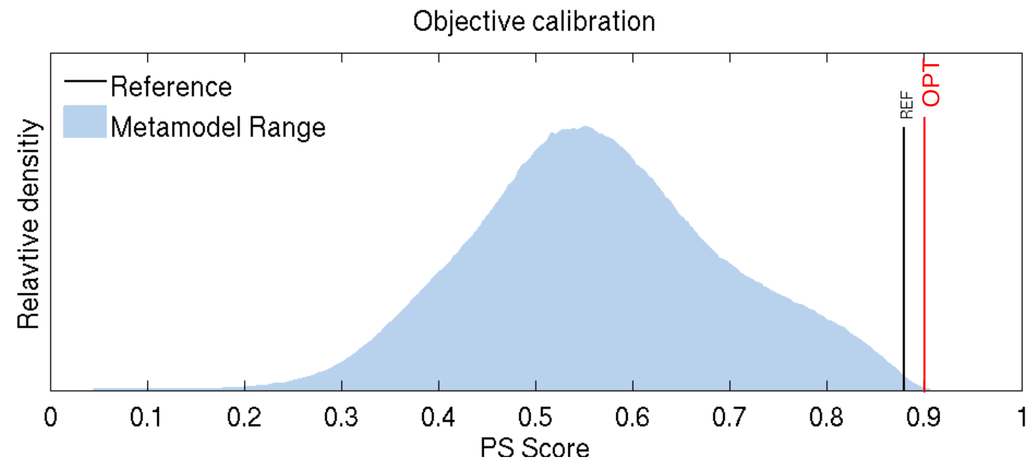

Figure 1: relative density of PS skill scores for all 5 million sampled calibration parameter combinations. The line labelled REF shows the value for the default setting of the parameters. The red line labelled OPT denotes the highest score achieved in the whole sample.

The sampled combination of tuning parameters which achieves the highest skill score, is as follows:

Calibration parameter	Optimized value
rlam_heat	**0.5249**
entr_sc	**$1.86e^{-4}$**
qi0	**0.0**
uc1	**0.0626**
tkhmin	**0.35**
fac_rootdp2	**0.9**
radfac	**0.5**
soilhyd	**1.62**

In Figure 2 the distribution of the values of the calibrated values with respect to the reference values can be seen. For qi0 and radfac the calibrated values coincide with the default values. For uc1 the default value seems to be at the high end of the range of values with high scores. The calibrated value of tkhmin is at the low end of the range of values with high scores.

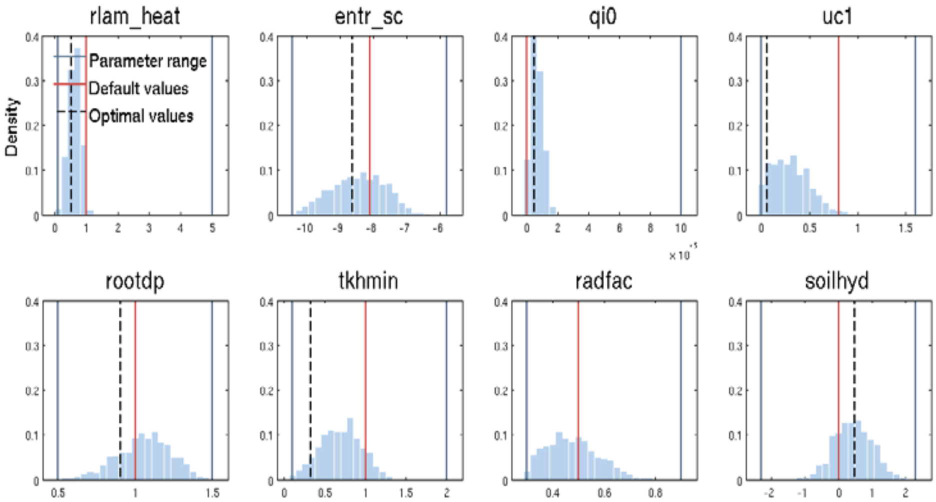

Figure 2: parameter density for the 0.1% combinations with the highest PS skill score. The gray vertical lines show the extreme values of the parameter range, the red line the default values used in the reference simulation. And the dashed black line the values for the parameter combination with the highest skill score. Note that for entr_sc and soilhyd the logarithm of the actual values is shown.

Control simulation with calibrated parameters

In the next step a simulation covering the period 1979-2000 with all calibration parameters set to their calibrated values was performed.

Figure 3 shows the effect of the calibration with respect to the bias of the climatological summer (JJA) 2 meter temperature based on E-OBS observation data for the period 1981-2000. It is evident that the calibration causes an overall cooling over the whole domain. This cooling is beneficial to all regions which overestimated the summer temperature in the uncalibrated simulation. However, the cooling is detrimental to the region north of 60° latitude, which was already too cold in the uncalibrated simulation.

In contrast to temperature the effect of the calibration on JJA total precipitation is minor. The bias remains almost unchanged. A slight improvement may be visible over Eastern Europe. The same is also true for total cloud cover, which is not shown here.

The reduction of the bias shows also in the PS skill score, which for the calibrated simulation evaluated again for the period 1979-1983 amounts now to **0.8958**. This is slightly lower than value that was estimated by the meta-model (**0.9047**) but still 1.5% higher than the skill score of the uncalibrated configuration (**0.8790**).

The improvement is also visible in the seasonal cycles of the biases shown in Figures 5, 6 and 7. Please note that for total precipitation and total cloud cover the differences between the calibrated and the uncalibrated simulation mostly remain in the range of the internal variability or the observational uncertainty.

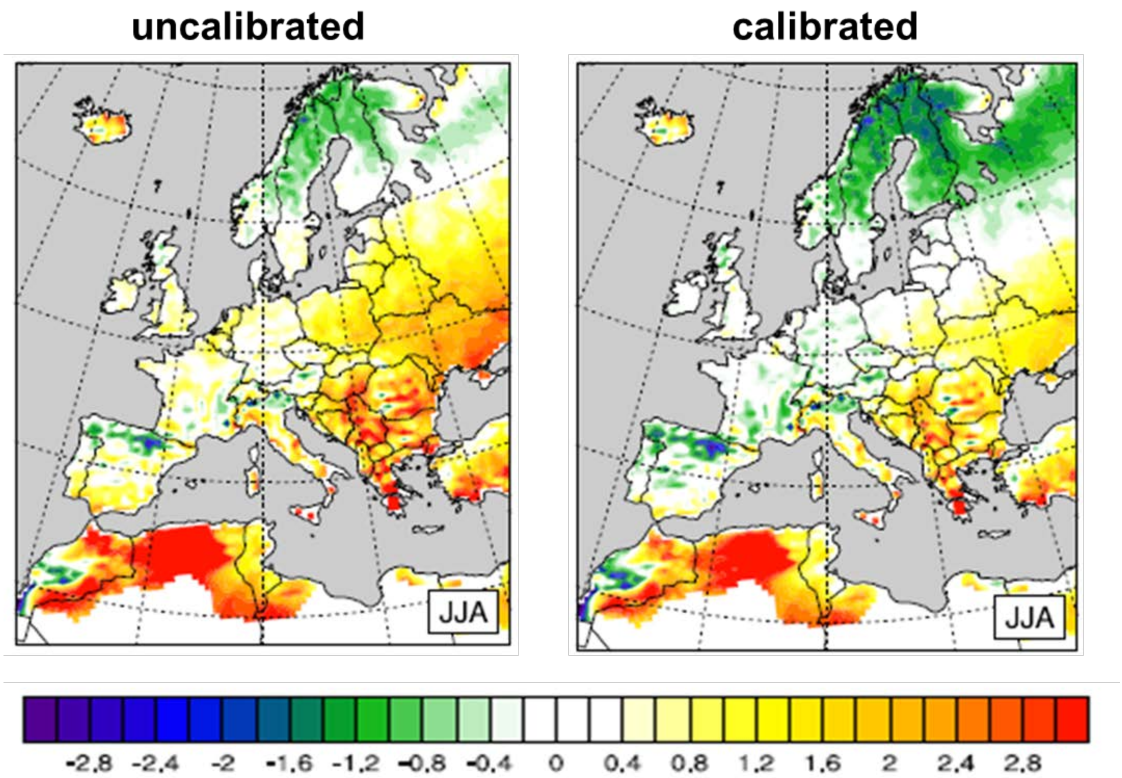

Figure 3: model bias of the 2 meter temperature with respect to E-OBS observation values for the summer seasonal mean (JJA) for the average of the years 1981-2000. Values are given in °K.

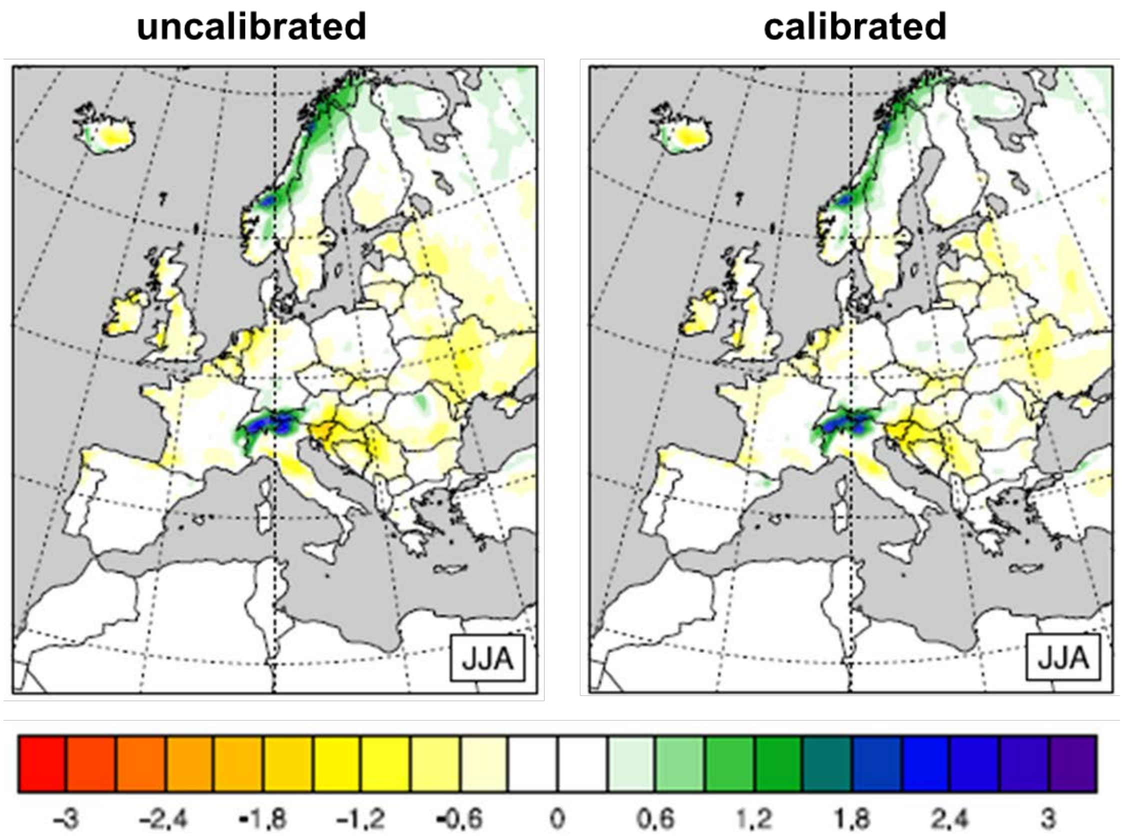

Figure 4: model bias of the total precipitation with respect to E-OBS observational values for the summer seasonal mean (JJA) averaged over the years 1981-2000. Values are given in mm/day.

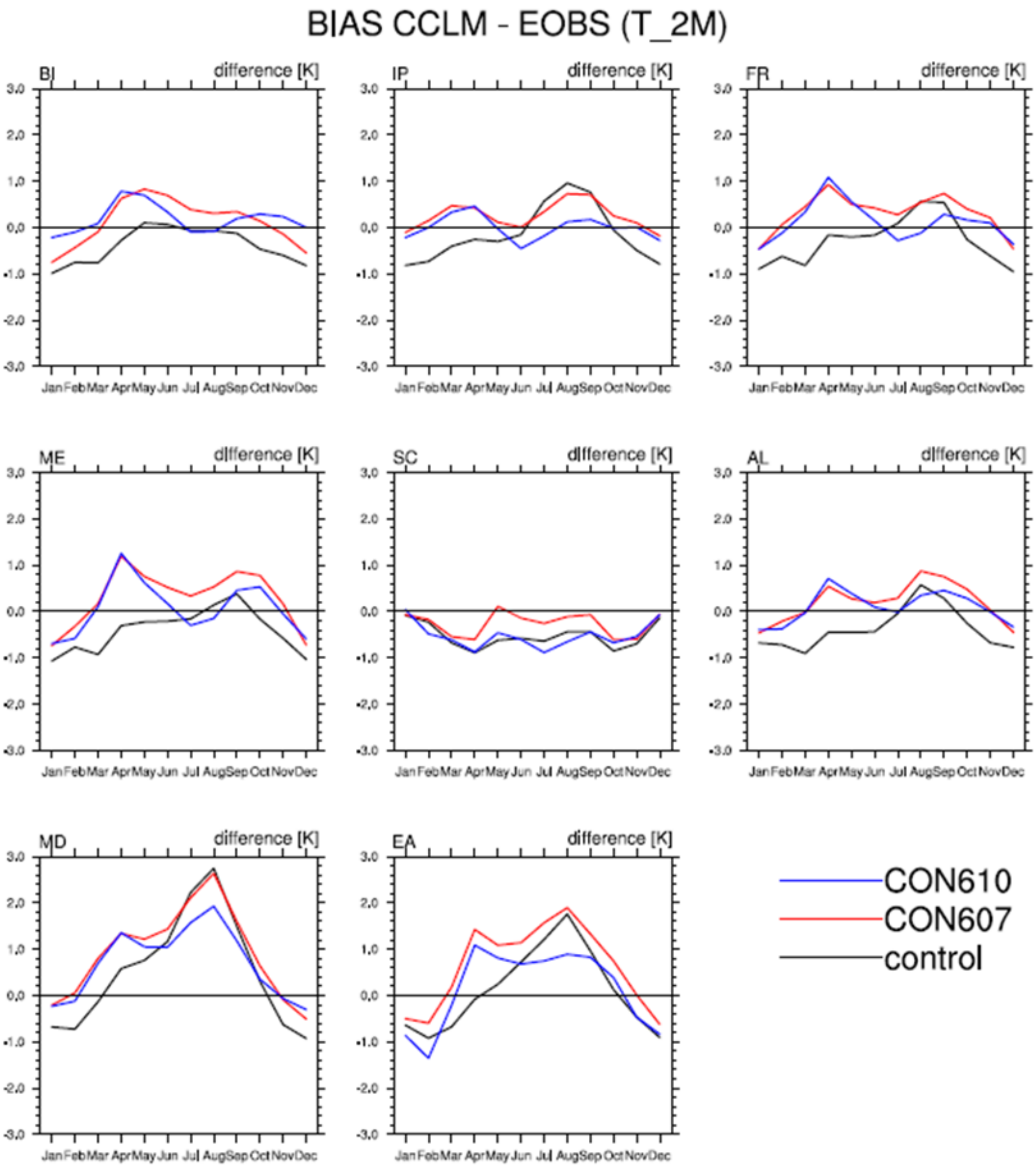

Figure 5: seasonal cycle of 2 meter temperature bias with respect to the E-OBS data averaged over the 8 PRUDENCE regions (British Islands(BI), Iberian Peninsula (IP), France (FR), Middle Europe (ME), Scandinavia (SC), Alpine region (AL), Mediterranean area (MD) and Eastern Europe (EA)). The blue lines show the results for the calibrated simulation, the red line the results for the uncalibrated simulation and the black line results for the configuration which was the starting point of the COPAT experiment.

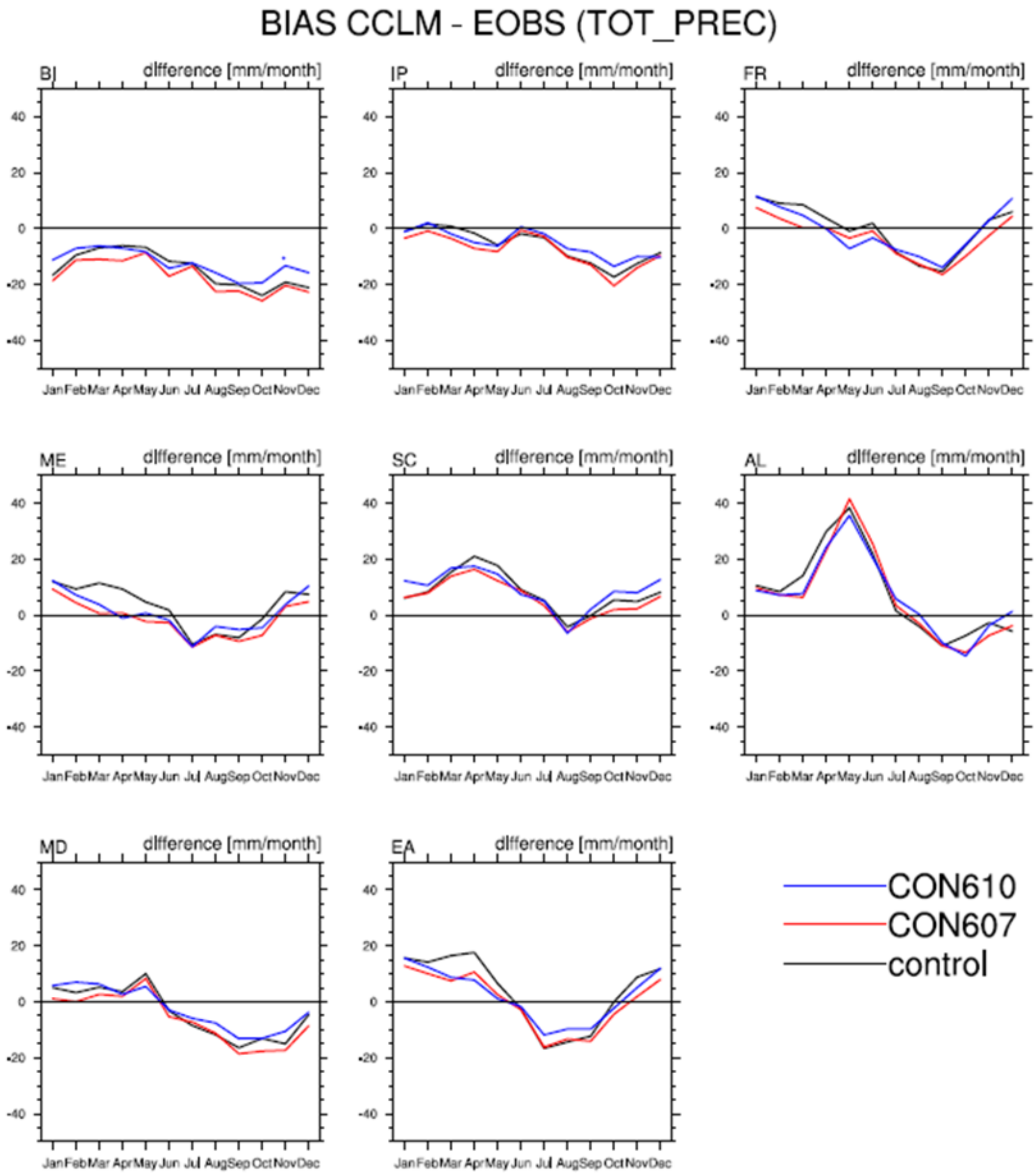

Figure 6: As Figure 5 but for total precipitation.

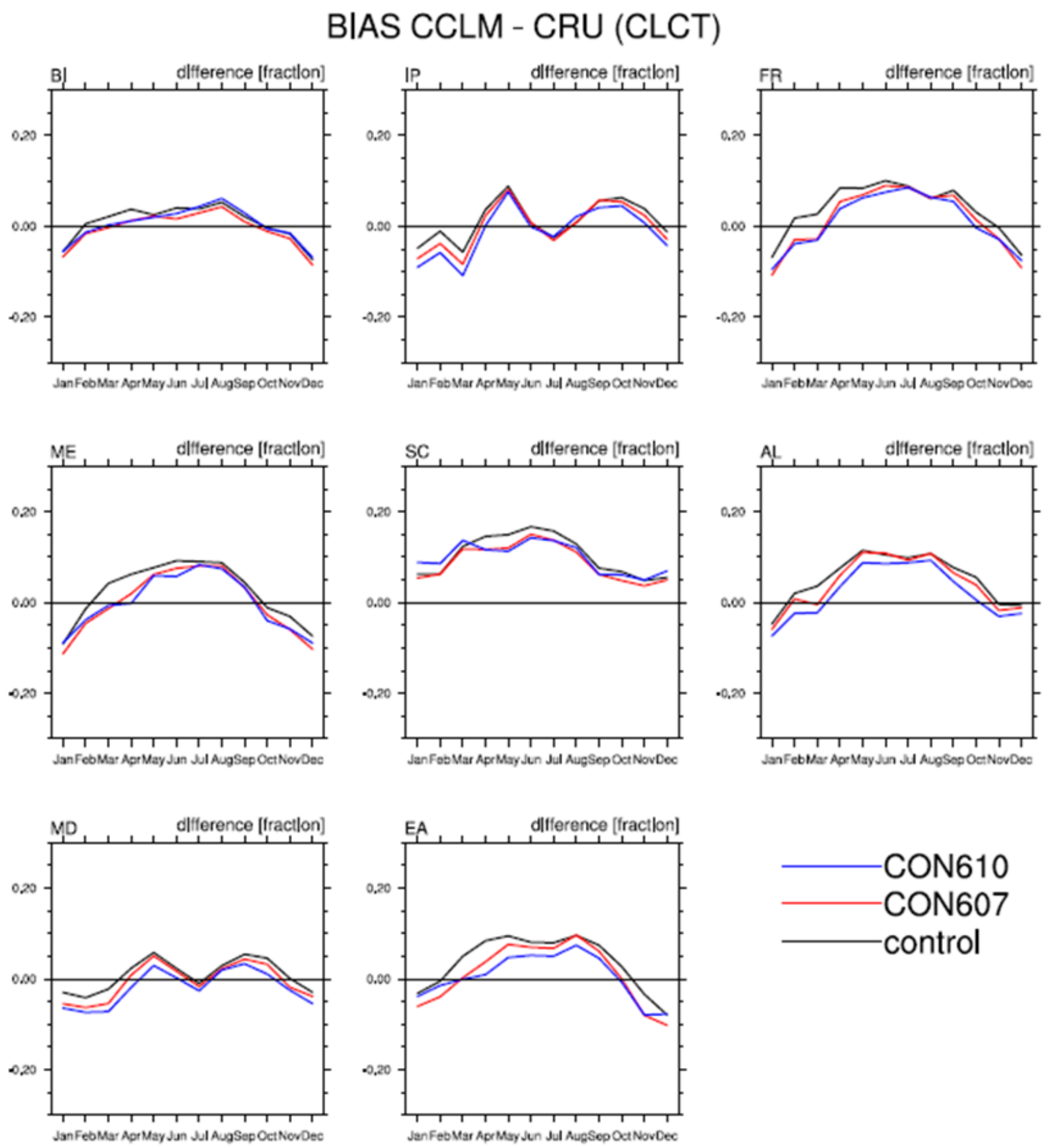

Figure 7: As Figure 5 but for total cloud cover bias as compared to the CRU gridded climatology.

Final remarks

It should be noted here that the configuration which was presented at the CLM Assembly 2015 to be accepted as the recommended configuration at hydrostatic scales for the European domain (EURO-CORDEX domain) was slightly different to the configuration used in the calibration process. It was suggested by Jan-Peter Schulz that itype_root=2 should be used in combination with itype_evsl=3. itype_root=2 then also requires a time invariant root depth, i.e. itype_rootdp=4 in int2lm. This suggestion was only made after the calibration work had been started and it was too late to include it in the process. But since the change of this switch alone showed a beneficial impact on the Brier Skill Score at the end of calibration an additional simulation was performed with itype_root=2 and all calibration parameters set to the calibrated values except for fac_rootdp2, which remained at the default value of 1. This additional simulation achieved an even higher PS skill score of **0.8984** than the simulation with itype_root=1.

Even though the daily temperature range (DTR) is not used in the evaluation of the simulations during the calibration process, the resulting values in the calibrated simulation showed an improved skill for representing this quantity compared to the uncalibrated simulation.

Bellprat et al. (2016) found a significant improvement for the interannual variability of the summer mean temperature when using the calibrated set of parameters (see their Figure 5). In Figure 8 it is shown that also as a result of the calibration for COSMO 5.0 this quantity is improved. However, this improvement is not quite as significant since the bias of the reference simulation is already considerably smaller than the bias of the reference in their paper.

It was also tested whether DTR could be directly included in the performance skill score PS as a 4th climate variable. From a technical point of view this can be achieved rather easily since T2m_max and T2m_min are also available from E-OBS. However since the biases of DTR are much larger than those for T2m and the normalization values are essentially of the same size the quadratic term in the numerator of the PI score favors improvements of DTR also at the expense of a deterioration of T2m. Thus it was decided to stay with the 3 climate parameters used in Bellprat et al. (2012).

qi0 was used as a tuning parameter in this work. But as the results of the calibration procedure show, any change from the default value of 0.0 deteriorates the results. Similar results were also found in the CALMO project for NWP at the 1km scale. Omitting this calibration parameter and also the parameter for rooting depth (which is set to the default value in the recommended configuration) would reduce the computational costs substantially (73 instead of 129 simulations needed).

The whole calibration was performed for a grid resolution of 0.44 degrees. However we also performed a 30 year long simulation with the same calibration parameters for the EURO-CORDEX configuration at 12 km resolution. The biases for this simulation were very similar to the biases seen with the coarser resolution simulation. Thus the calibrated values should also be usable for higher resolution simulations as long as they use parameterized convection.

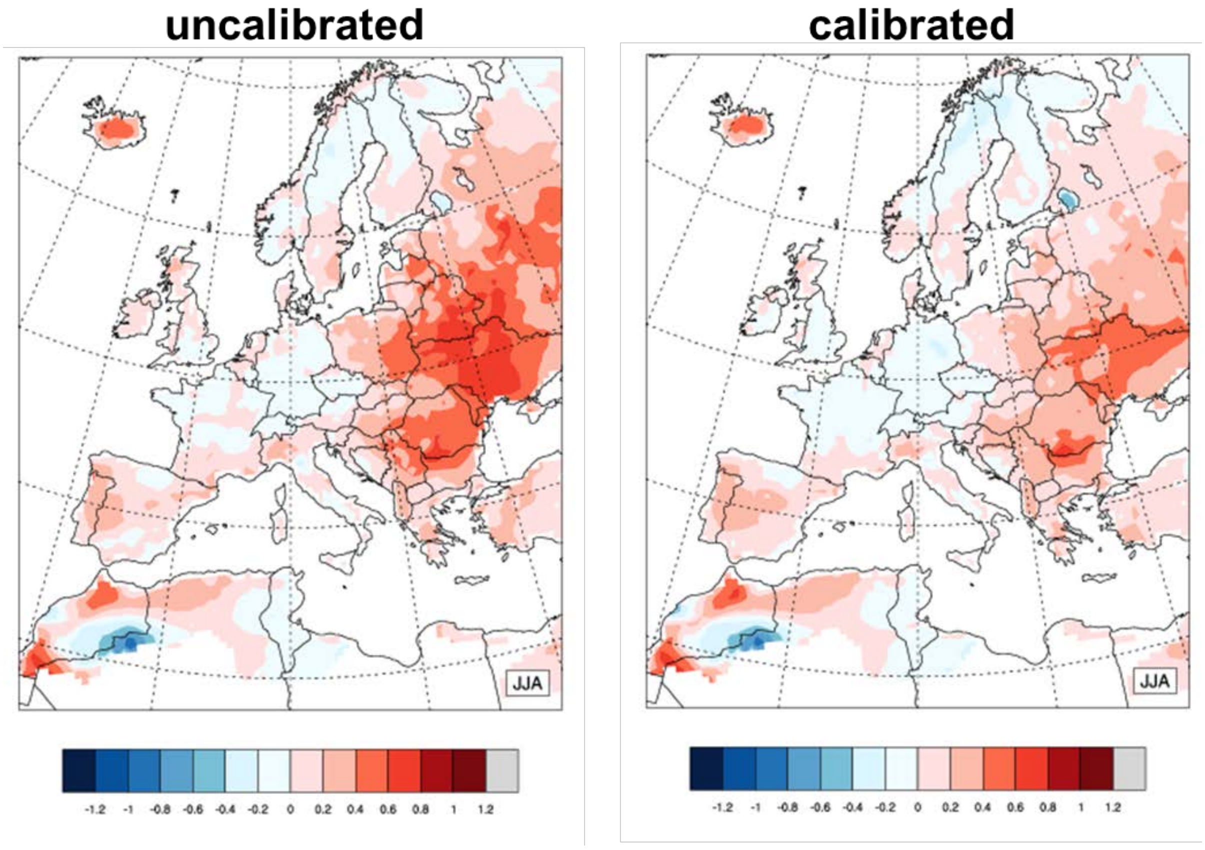

Figure 8: bias of interannual variability for summer season mean temperature with respect to E-OBS for the period 1980-2000. The left panel shows the bias for the reference simulation and the right panel for the simulation with the calibrated parameters.

Bottom line

The calibration process was able to improve the overall PS skill score of the cosmo5.0_clm6 based configuration for the EURO-CORDEX-044 domain by ~1.5%. This is not a dramatic improvement and leads to the conclusion that the uncalibrated model was already well tuned.

The main improvements can be seen in the 2-meter- temperature bias, which is most notably decreased during summer in the southern and eastern parts of Europe. However to the north of 60° latitude the negative temperature bias is increased.

Total precipitation and cloud cover remain almost unchanged by the calibration.

DTR and the interannual variability of mean summer temperature are also slightly improved even though they hadn't been evaluated explicitly in the calibration process.

Similar results are also found for simulations at 12km resolution.

References

Bellprat, O., S. Kotlarski, D. Lüthi, and C. Schär (2012), Objective calibration of regional climate models, J. Geophys. Res., 117, D23115, doi:10.1029/2012JD018262.

Bellprat, O., S. Kotlarski, D. Lüthi, R. De Elia, A. Frigon, R. Laprise, and C. Schär (2016), Objective Calibration of Regional Climate Models: Application over Europe and North America, J. of Climate, 39, 819-838, doi:10.1175/JCLI-D-15-0302.1

Acknowledgements

The reference configuration for this calibration exercise was developed within the COPAT project (a coordinated effort of the CLM-community). Contributors to this effort were I. Anders (ZAMG), B. Geyer (HZG), K. Keuler (BTU), H.-J. Panitz (KIT), M. Mertens (DLR), A. Ferrone (LIST), H. Wouters (KUL), M. Demuzere(KUL), S. Saeed (KUL), S. Brienen (DWD), B. Früh (DWD), J.P. Schulz (DWD), A. Roches (C2SM), D. Lüthi (ETHZ).

The computing and storage recourses needed for the calibration work was supported by a grant from the Swiss National Supercomputing Centre (CSCS) under project ID ch4.