Consortium

for

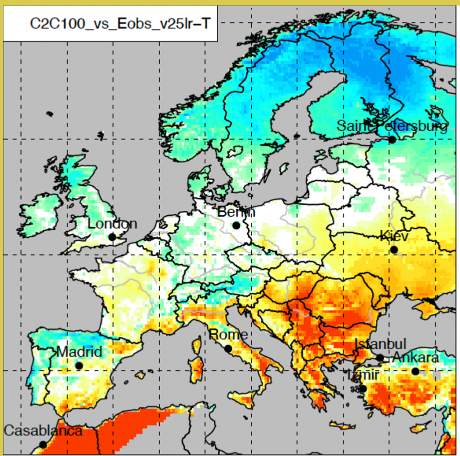
Small-Scale Modelling

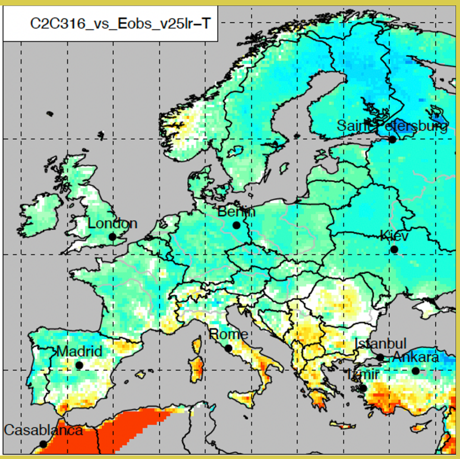
Technical Report No. 51

CLM Community WG EVAL, COordinated PArameter Testing project 2 (COPAT2): COSMO-CLM 6.0 clm1 recommended model configuration.

March 2024

DOI: 10.5676/DWD_pub/nwv/cosmo-tr_51





www.cosmo-model.org

Editor: Massimo Milelli, CIMA Foundation

CLM Community WG EVAL,
COordinated PArameter Testing project 2 (COPAT2):
COSMO-CLM 6.0 clm1 recommended model configuration

***E. Russo*[1], *B. Geyer*[2], *R. Petrik*[3],**

***K. Keuler*[4], *M. Adinolfi*[5] *H. Feldmann*[6]**

***K. Goergen*[7], *A. Kerkweg*[8] *P. Khain*[9]**

***P. Ludwig*[6], *M. Mertens*[10] *P. Pothapakula*[1]**

***M. Raffa*[5], *B. Rockel*[2] *J.-P. Schulz*[11,5]**

***M. Sulis*[12], *H. Thi Minh Ho-Hagemann*[2] *H. Truhetz*[13]**

***L. Uzan*[9], *U. Voggenberger*[14] *C. Steger*[11]**

[1]***Institute for Atmospheric and Climate Science (IAC) - ETH Zuerich, Switzerland***

[2]***Helmholtz-Zentrum Hereon, Germany***

[3]***Navy Command of the German Navy, Germany***

[4]***Chair of Atmospheric Processes - Brandenburg University of Technology (BTU) Cottbus-Senftenberg, Germany***

[5]***Euro-Mediterranean Center on Climate Change (CMCC), Italy***

[6]***Institute of Meteorology and Climate Research - Karlsruhe Institute of Technology (KIT), Germany***

[7]***Institute of Bio- and Geosciences (Agrosphere, IBG-3) - Research Centre Juelich (FZJ), Germany***

[8]*Institute of Energy and Climate Research (Troposphere, IEK-8) - Research Centre Juelich (FZJ), Germany*

[9]*Israel Meteorological Service, Israel*

[10]*Institute of Atmospheric Physics - German Aerospace Center (DLR), Germany*

[11]*Deutscher Wetterdienst (DWD), Offenbach am Main, Germany*

[12]*Environmental Research and Innovation Department - Luxembourg Institute of Science and Technology, Luxembourg*

[13]*Wegener Center for Climate and Global Change (WEGC) - University of Graz, Austria*

[14]*Department of Meteorology and Geophysics (IMGW) - University of Vienna, Austria*

1 Abstract

In this study we summarise the results of the extensive tests and analysis that have been conducted to assess the performance and optimize the configuration of the latest release of the COnsortium for Small-scale MOdelling model in CLimate Mode (COSMO-CLM 6.0). The presented work has been conducted as a joint effort by the members of the second phase of the COordinated Parameter Testing project (COPAT2) of the working group EVAL of the CLM Community. Here we provide all the technical details of the evaluation procedure and a final optimal configuration that is suggested to serve as a reference for simulations with the model over the European domain.

2 Introduction

The start of the development of the COSMO model dates back to the early 1990s when the Deutscher Wetterdienst (DWD) decided to develop a non-hydrostatic model for weather predictions at convection-permitting resolutions. The new model also seemed to be a good candidate for a regional climate model. Scientists from the Potsdam Institute for Climate Impact Research and later also from the Brandenburger Technische Universität Cottbus - Senftenberg and the Helmholtz-Zentrum Hereon developed a climate version of the COSMO model in the following years. They were able to release the first model version in the year 2002.

The CLM Community was founded in 2004 and formed the platform for collaboration and further development of the climate mode of the COSMO model (COSMO-CLM) in the years to come. After some years of parallel development, the first unified version of the model was released in 2007 (COSMO 4.0). This version included developments from the Numerical Weather Prediction (NWP) and climate communities.

The unification of the developments for NWP and climate was repeated in 2014 with the release of COSMO 5.0 and again in 2021 with the latest release COSMO 6.0. COSMO 6.0 was released on 15 December 2021. It is the last release of the COSMO model, since the DWD (as the main developer) and the other national meteorological services organised in the Consortium for Small-scale MOdelling (COSMO) have already or are planning to switch to the ICON modelling framework. The release of COSMO 6.0 marks the endpoint of the long and very successful history of the COSMO model, which was used for operational weather prediction and climate research in many meteorological services and research institutions around the world for more than 20 years.

The CLM Community always tried to provide well-tested model versions and optimized configurations to its members. To achieve this goal, the unified releases of the COSMO model were always extensively analysed and many tests were conducted in order to optimize the model set-up before a new recommended version of the model was suggested to the community members. This is an important achievement and sets the CLM Community apart from other modelling communities that leave this task to the users of the model. The tests and improvements of the model were always done in collaboration between different community members, ensuring a high quality of the simulation results.

This procedure is of course also applied to COSMO 6.0 and this report presents the results of the optimisation procedure for the model version for climate applications (COSMO-CLM 6.0) conducted over Europe. The presented work was conducted within a community internal project called COordinated PArameter Testing 2 (COPAT2, COPAT1 was for COSMO 5.0).

The goal of the first part of COPAT2 is to determine an optimal model configuration for the European CORDEX domain for COSMO-CLM 6.0. The second part of COPAT2 is dedicated to the setup of the climate limited-area mode (ICON-CLM) of the latest release of the ICON modelling framework. The results for ICON are not described in this report but will be published in a separate document as soon as the process will be completed.

The focus of COPAT2 for COSMO-CLM 6.0 is mainly on testing new features of the model that became available in the latest release, and not on optimizing the values of the tuning parameters. There are several reasons for this strategy. First, the usage of new configuration options or parameterisations is expected to result in the largest differences (improvements) with respect to the former standard setup, since the tuning parameters of the model have been optimized in the NWP and climate communities over more than two decades, leaving little room for improvements. Second, detailed testing of different values of many tuning parameters and their combinations is very time-consuming and resource intensive. Together with the expected little potential for improvement in the results, this did not seem to be a productive strategy.

The evaluation process starts from the recommended configuration of COSMO 5.0. The effects of single as well as combined changes in newly-available model configuration options and developments are evaluated through a series of coordinated simulations conducted by the project task-force, including members from different institutions across Europe.

In a first phase (COPAT2 Phase Ic, where c stands for the COSMO part of COPAT2), a set of simulations are integrated over a relatively short period of 7 years, with the preliminary goal of determining to which configuration changes the model is more sensitive. In a second phase (COPAT2 Phase IIc), a new set of simulations is performed, combining the most sensitive model configurations of the Phase Ic runs. Finally, in a third phase (COPAT2 Phase IIIc), the most promising experiments in terms of agreement with observations are extended over a total period of 12 years, from 1979 to 1990, with an additional narrower ensemble produced for a more recent period of time, covering the years from 2002 to 2008. A very final simulation with the best-performing configuration is further extended over the entire period from 1979 to 2020 (also corresponding to the reference period of the CORDEX-CMIP6 evaluation runs), in order to conduct a more robust comparison against the reference configuration of the current recommended model version COSMO-CLM 5.0.

Results show that configuration changes in model dynamics and in the representation of surface processes lead to significant improvements in model performance with respect to the recommended configuration of the previous model version COSMO-CLM 5.0.

This report is structured as follows: different details of the evaluation procedure are introduced in section 3, including information on the reference simulation configuration and model domain (section 3.1), as well as a description of all the conducted experiments, the selected observational data sets and the employed metrics (introduced in section 3.2, section 3.3, and section 3.4, respectively). The results are presented and discussed in section 4, divided into different subsections corresponding to the three different phases of the evaluation procedure. Concluding remarks and considerations on the recommended model version are finally presented in section 5.

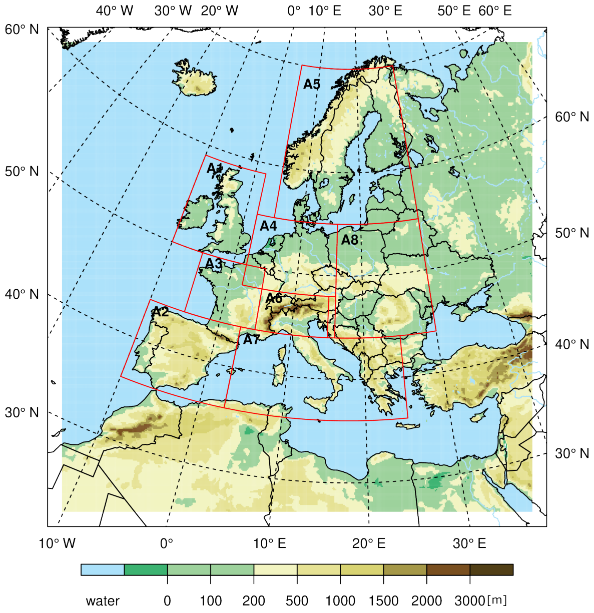


Figure 1: Simulation domain (EURO-CORDEX domain EUR-11). The transparent boundary area at the panel's borders denotes the simulations' sponge zone. The different considered evaluation regions (known as PRUDENCE regions [6]) are marked in red: A1 British Islands; A2 Iberian Peninsula; A3 France; A4 Mid-Europe; A5 Scandinavia; A6.Alps; A7 Mediterranean; A8 East-Europe.

3 Methods

3.1 Reference simulation and common choices in model configuration

The target domain of the conducted experiments is the CORDEX European domain (EUR-11) [10], covering the entire Europe and part of Northern Africa and the Middle-East. The grid used for the performed simulations presents a horizontal resolution of 0.11°(∼ 12.5 km), extending 450 grid boxes in longitudinal and 438 in latitudinal directions, including a sponge zone of 13 grid boxes on each side (see Fig.1). All the performed simulations use 40 terrain-following vertical levels with geometric height coordinate.

In this work, the reference configuration of COSMO-CLM 6.0, from which all the other configurations are derived, is the one based on the recommended model configuration proposed in the first COPAT initiative (COPAT1)[*] for COSMO5.0_clm9 (referred to as simulation C2C100), with additional changes leading to improved performances as tested in the NUKLEUS project[†]. The reference COPAT2 simulation is denoted experiment C2C201. The setups of experiments C2C100 and C2C201 are provided in the appendix A.1, via their namelist files C2C100_YUSPECIF and C2C201_YUSPECIF, respectively.

External parameters like orography, vegetation cover, surface roughness length, etc. are interpolated onto the target grid using the EXTernal PARarameter (EXTPAR) version 3.0 tool of the CLM Community. The following data sets are used as input: FAO DSMW[‡], GLOBE (orography,[11]), and Lake Database (GLDB [5]). For land cover inputs, all simula-

[*] https://hcdc.hereon.de/clm-community/uploads/media/material/02b03b83-3433-4215-bd5e-508c2fbe7128/Report_RecommendedVersion.docx

[†] https://www.regiklim.de/DE/Querschnittsprojekte/NUKLEUS/nukleus_node.html

[‡] FAO Digital Soil Map of the World

tions use the GLC2000[§]. Additionally, to confirm the validity of the obtained results when using a more up-to-date dataset, also the GlobCover2009[¶] database is employed in a final test simulation.

The reference simulation C2C201 covers the entire period from 1979 to 2020. The temporal integration is conducted with a two-time-level, third-order Runge-Kutta scheme. The horizontal advection of wind components is calculated with a third-order upwind scheme, and the vertical advection is performed with a third-order implicit advection scheme. The scalars are transported using a second-order Bott advection scheme with deformational correction. Fast processes are treated by the newest version of the fast waves solver by Baldauf [2],[8]. More details about the default configuration of the COSMO model dynamics can be found in [3]. Regarding the parameterisation of physical processes, the model configuration is close to the NWP configuration of COSMO 6.0 but adapted to follow the recommendations of the CLM Community. The albedo is chosen to depend on the soil moisture in order to appropriately capture the feedback between soil and near-surface temperatures (large albedo values for dry soil and reduced albedo for wet soil). The type of root distribution is not uniform but follows an exponential decay. The calculation of heat conductivity considers not only soil moisture but also soil ice. The parameterisation of bare soil evaporation is according to Noilhan [15].

The ERA5 reanalysis data [12] are used as boundary conditions (lateral, upper, and bottom boundaries) for all the performed simulations. In order to avoid a multi-year spin-up period, for an appropriate initialisation of the deeper soil layers, an averaged soil moisture field is used to initialize soil moisture in (nearly) all simulations, representing a climatological mean state around the starting date of the simulations. This field is generated from a previous long-term evaluation run with the same configuration as the reference run, by averaging the soil moisture values over a five-day window around the 1st of January between the years 1994 and 2008. This period is selected far enough from the initial date in order to ensure that the model is in equilibrium with respect to the given variables.

3.2 Tested model configurations

Starting from the configuration of the reference simulation, hereafter referred to as C2C201 (later on referred to as C2C301 in Phases II and III of the project[‖]), a series of experiments is conducted in Phase I of the project, modifying different model parameters and physical options belonging to three main categories: model physics, turbulence, and dynamics. These experiments, covering the years from 1979 to 1985, are listed in Tab. 1, ordered according to their experiment ID, going from C2C202 to C2C225. A schematic of the applied changes in the model setup of Phase I and the corresponding "area of influence" in the climate system are displayed in Fig. 2.

The changes in model dynamics include, among others, an option to choose the Bott 2nd order finite-volume scheme with deformational correction and local time stepping, an option to calculate the horizontal pressure gradient (i.e. in the u- and v- equations) by interpolating the tendency of pressure to a horizontal plane, a new scheme for the treatment of fast waves and an option for choosing an artificial divergence damping acting in a fully isotropic 3D manner.

[§] Global Land Cover 2000 database. European Commission, Joint Research Centre, 2003.

[¶] © ESA 2010 and UCLouvain http://due.esrin.esa.int/page_globcover.php

[‖] Configuration in namelist tool: `https://tools.clm-community.eu/NLT/tmp/upload_202204200911_9412fa6c76864bea2bd3.txt`

The tested changes in model physics include the use of a new calculation of skin temperature and an improved representation of bare soil evaporation (after [18]), as well as a new soil groundwater formulation allowing groundwater build-up (after [17]). For all these tests, we added to the external parameter inputs the skin conductivity derived from the GlobCover2009 data set, using the EXTPAR version 5.2.1.

Finally, the tested changes in the representation of turbulence include, among others, the use of the new turbulence scheme of the ICON modelling framework implemented in COSMO-CLM 6.0, a test calculating SSO-wake turbulence production for Turbulent Kinetic Energy (TKE), one considering horizontal shear production for TKE, and a simulation with clouds sub-grid scale condensation considering water clouds.

In the second phase of COPAT2 (Phase IIc), further tests are conducted for the same period 1979 to 1985, combining changes in the model setup of Phase Ic that showed the best performance in the comparison against observations. The only exceptions are the experiments C2C314 and C2C315, representing modified versions of experiment C2C214 for which no important improvements with respect to the reference model version were found during Phase Ic. The new experiment C2C314 includes some namelist changes with respect to C2C214, inherent to soil and surface properties, while experiment C2C315 is started from different initial conditions, with soil moisture and temperature derived from the simulation C2C314.

In the third Phase (Phase IIIc) of COPAT2, the best-performing simulations with combined configuration optimisations from Phase IIc are further extended until 1990 (covering overall the period 1979-1990). Additionally, the same experiments are conducted over a more recent period of time, covering the years from 2002 to 2008. The latter tests are all started from the restart file of the experiment C2C301 at the beginning of January 2002. For the analysis, only the years from 2003 to 2008 are considered. The main goal of the experiments of Phase IIIc is to test the robustness of the obtained results when considering different time periods. All the experiments of Phase IIc and IIIc are listed in Table 2.

The best-performing simulations based on these sets of experiments are further prolonged in a final step of Phase IIIc, covering the period from 1979 to 2020.

In a concluding step, a simulation using the determined optimal model setup is performed over the period 1979 to 1985 using the GLOBCOVER2009 database instead of the GLC2000, in order to test whether the obtained model improvements do not change significantly when using a different and more recent land cover dataset. This simulation is denoted as experiment C2C316er_bg in the following text.

The experiments of Phase I are conducted on Mistral, the previous High Performance Computing (HPC) system of the German Climate Computing Center (DKRZ), while the experiments of Phases IIc and IIIc are conducted on Levante, the new HPC System of DKRZ, that started its operations in March 2022. In order to take into account the possible effects of different hardware, compilers, and libraries of the different employed machines, the reference simulation is executed on both Mistral (experiment C2C201) and Levante (experiment C2C301). The results of the two simulations are almost identical (not shown).

3.3 Observational Datasets

For the evaluation of the model results, in Phase Ic a comparison against two datasets is conducted: the E-OBS observational data [7] and the reanalysis data ERA5 [12]. Both datasets are gridded products, retrieved on a regular grid with a spatial resolution of ~ 25 km. In Phase IIc and IIIc, additional satellite products are used for the 2003-2008 evaluation

Table 1: Description of changes in the model setup of the different experiments of the first evaluation Phase Ic with respect to the reference experiment C2C201.

ID	namelist parameter	value in C2C201	tested value
C2C202	scalar_advec	BOTT2	BOTTDC2
	itype_fastwaves	1	2
	l3D_div_damping	false	true
	ldyn_bbc	true	true
	itype_bbcw	1	114
C2C203	C2C202 + ldyn_bbc	true	false
C2C204	C2C203 + lhor_pgrad_Mahrer	false	true
C2C205	C2C203 + itype_outflow_qrsg	1	2
C2C210	itype_canopy	1	2
	cskinc	30	-1
	int2lm::lskinc	false	true
C2C212	itype_evsl	3	4
	c_soil	1	1.25
C2C213	cwimax_ml	0.000001	0.0005
C2C214	itype_hydmod	0	1
C2C220	ltkesso	false	true
C2C221	ltkeshs	false	true
C2C222	icldm_turb	2	2
	icldm_tran	0	2
C2C223	loldtur	true	false
	itype_vdif	-1	1
C2C224	lsuper_coolw	false	true
C2C225	itype_lbc_qrsg	1	1
	lana_qi	true	false
	lana_qrqs	true	false
	llb_qi	true	false
	llb_qr_qs	true	false

Table 2: Description of changes in the model setup of the experiments of the second and third evaluation phases (Phase II and Phase III, respectively) with respect to the reference experiment C2C201.

ID	namelist parameter	value in C2C301	tested value
C2C301	C2C201 (1980-1990)		
C2C301c	C2C201 (2002-2008)		
C2C302	C201 + C210 + C212		
C2C303	C2C302 +		
	y_scalar_advect	BOTT2	BOTTDC2
	itype_fast_waves	1	2
	l_3D_div_damping	false	true
	ldyn_bbc	true	false
	itype_bbc	1	114
C2C303c	C2C303 (2002-2008)		
C2C304	C2C303 + lhor_pgrad_Mahrer	false	true
C2C305	C2C303 + itype_outflow_qrsg	1	2
C2C306	C2C303 +		
	lhor_pgrad_Mahrer	false	true
	itype_outflow_qrsg	1	2
C2C314	C2C214 +		
	new S_ORO_MAX		
	itype_hydmod=0	0	1
C2C315	C2C314 + restart W_SO, T_SO from 1986010100		
C2C316	C2C303 +		
	itype_conv	0	2
	icapdcycl	0	2
	lconf_avg	true	false
C2C316c	C2C316 (2002-2008)		
C2C317	C2C303 +		
	lhor_pgrad_Mahrer	false	true
	itype_outflow_qrsg	1	2
	itype_conv	0	2
	icapdcycl	0	2
	lconf_avg	true	false
C2C317c	C2C317 2002-2008		

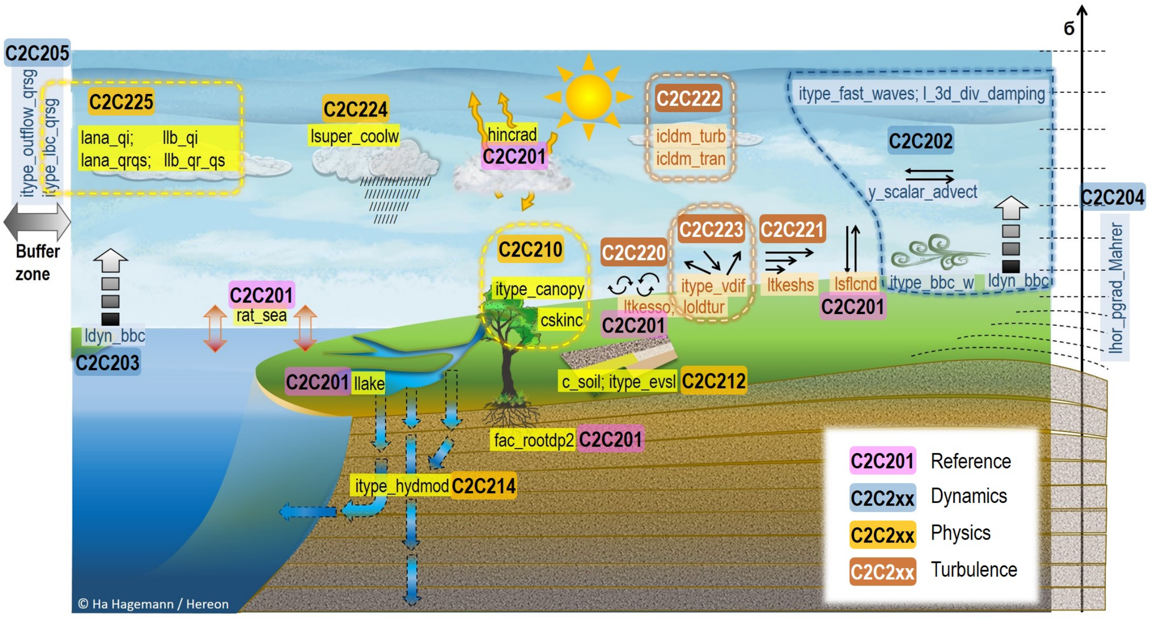


Figure 2: Scheme of conducted experiments of Phase I, with the tested configuration choices of each experiment reported in correspondence of the different components of the climate system they affect/represent. The experiments are highlighted in different colours, depending on whether their configuration is inherent to the model physics, dynamics, or turbulence representation.

period, namely CERES [14] and SARAH2 [16].

The variables considered in the evaluation procedure are listed in Table 3. Different weights are assigned to the different variables for aggregation of the evaluation metrics to a final score (see Section 3.4). When multiple observational data sets are used for the same output variable, the weights are split between the data sets, with the most reliable one (based on personal judgment) receiving a stronger weight. The different assigned weights are provided in the last column of Table 3.

The evaluation is performed point-by-point, considering 3-daily means of the given variables. Taking into account that the ERA5 data used as boundary conditions for the conducted experiments are "realistic" reanalysis data, we assume that, for a single grid-box, the model is able to properly represent the variability of a given variable at synoptic time-scales. This enables, on the one hand, to have a time-series long enough for testing the robustness of the employed metrics using Monte-Carlo approaches. On the other hand, it allows to make the variables more Gaussian through averaging: this then allows for the application of estimators such as the Root Mean Squared Error (RMSE), which are better suitable for normally distributed data [13]. Since the nature of precipitation data is expected to be more chaotic than the other considered variables, we have decided to conduct the analyses for total precipitation separately. Additionally, for the variable cloud cover of the CERES satellite product, monthly mean values are considered instead of 3-daily means, representing the original temporal resolution of the CERES data set.

Table 3: Description of the variables considered in the evaluation procedure and their corresponding weights used to calculate the metrics introduced in the following subsections. Given is the model output and the CMOR. (CMOR stands for "Climate Model Output Rewriter", that can be used to produce CF-compliant netCDF files that fulfil the requirements of many of the climate community standard model experiments).

Name	Description	Dataset	Weight
T2M/tas	near surface temperature	E-OBS	0.66
TMAX 2M/tasmax	maximum near surface temperature	E-OBS	0.66
TMIN 2M/tasmin	minimum near surface temperature	E-OBS	0.66
TOT PREC/pr_amount	total precipitation	E-OBS	-
PMSL/ps	mean sea level pressure	E-OBS and ERA5	0.2 and 0.8
CLCT/clc	total cloud cover	ERA5 and CERES	1
ASOD S/rsds	downward shortwave radiation at the surface	E-OBS/ERA5 or SARAH2	0.3/0.7 or 1
TQV/clwvi	precipitable water	ERA5	1

3.3.1 Radiosondes

Moreover, for the entire period 1979-2020, vertical profiles of different variables are compared against radiosonde observations (see Section A.5).

Data from radiosoundings from European stations out of the Integrated Global Radiosondes Archive (IGRA) dataset [9] and a Copernicus data set** are used for regular reporting times 00, 06, 12, and 18 h plus all intermediate reporting times falling into a time window of +/- 1 h centered around each regular reporting time. Regardless of the actual reporting time and the length of the ascending time, each radiosonde profile is assigned to its regular reporting time. Additionally, the measured variables temperature, relative and specific humidity, and wind speed are interpolated to the standard pressure levels 850, 700, 500, and 300 hPa prior to the analyses to reduce computational efforts (additional information can be found in the Appendix A.4).

The position of each radiosonde is reconstructed from wind data using the method explained in the Appendix A.5. Instead of using only the coordinates of the station where the balloon was launched, the full trajectory of the balloon is reconstructed. In this way we are able to consider drifts over several hundreds of km, and to conduct a more appropriate comparison of the observations against model results. The balloon trajectories are calculated for all profiles, where a sufficient amount of data on different observed levels is available. This is necessary to provide high quality trajectories and avoid misleading information.

Finally, vertical profiles at the regular reporting times and at the positions of each radiosonde from each COSMO-CLM simulation are extracted via interpolation and compared with the profiles of the radiosondes to retrieve the simulations' biases.

** Copernicus Projekt C3S 311C_Lot2, unpublished up to now

3.4 Evaluation Metrics

In a first instance, the evaluation is conducted point-by-point on the regular lon-lat grid of the selected observational data set. The simulation results are upscaled onto the observational grid by bi-linear interpolation prior to the analysis. The main metrics that we consider in our analyses are the following: mean error (mean BIAS), Root Mean Squared Error (RMSE), Linear Correlation in time (LCorr), and the Advanced (symmetric) Mean Squared Error SkillScore, AMSESS, defined after [21] according to the following equation:

$$\text{AMSESS} = \begin{cases} 1 - \overline{BIAS_{ts}^{2}} \cdot \overline{BIAS_{ref}^{-2}} & \text{if } \overline{BIAS_{ts}^{2}} \leq \overline{BIAS_{ref}^{2}} \\ \overline{BIAS_{ref}^{2}} \cdot \overline{BIAS_{ts}^{-2}} - 1 & \text{if } \overline{BIAS_{ts}^{2}} > \overline{BIAS_{ref}^{2}} \end{cases}, \tag{1}$$

where $\overline{BIAS_{ts}^{2}}$ and $\overline{BIAS_{ref}^{2}}$ represent the mean of the quared BIAS of the test and reference simulations against observations, respectively . The AMSESS varies between -1 and 1. Positive (negative) values of the AMSESS indicate improved (worsened) performance of the test simulation with respect to the reference one.

Given the stochastic nature of precipitation, for this variable we have decided to consider a metric that is less sensitive to extreme outliers: the Median Absolute Deviation (MAD) [13], defined as:

$$MAD = median(|X_i - M|) \tag{2}$$

where X_i represents each data point in the dataset (in our case a 3-day mean value), while M is the median of the entire time series.

Based on the aforementioned metrics, we compare the model error against observations of a given simulation and the one of the reference run for each variable and grid-box. The error of a given simulation against observations might be smaller or larger than the one of the reference run. However, whether these smaller or larger errors are significant must be tested. In the next section we present a method for making a statistically sound assessment on whether a given simulation is better or worse than the reference in the comparison against observations.

3.4.1 Significance tests

For a given simulation, for each considered metric, variable, season, and grid point, a distribution of the differences in the comparison against observations between the tested and the reference simulation is inferred through bootstrapping, by randomly sorting out data values in the time series (see the details below). The ratio of the inferred distribution of differences falling below or above zero is taken as the basis to decide whether the tested simulation can be considered "moderately" or "strongly" different than the reference run (see details below).

The procedure that we follow to conduct a significance test under the null hypothesis that the differences in a given metric are equal to zero involves the following steps: the uncertainty of a selected evaluation metric is first calculated for a given atmospheric variable, season and grid-box for the reference (A) as well as another model experiment (B). Uncertainty means that the metric is tested for its robustness in making small changes to the data series: starting from the original data series with, e. g., 3-daily mean BIAS values for the 2-meter temperature in the winter season of model A, we randomly generate 250 times new data

Table 4: Metric-dependent thresholds of the significance level α of the applied significance test. Thresholds for moderate and strong significance are defined by the authors and denoted with α_m and α_s, respectively.

Metric	α_m	α_s
BIAS	0.075	0.01
MAE	0.075	0.01
LCorr	0.075	0.01
RMSE/AMSESS	0.1	0.02

series of the BIAS via bootstrapping (random sampling with replacement). For each of these iterations, the mean BIAS is then calculated, finally retrieving 250 mean BIAS values. At each iteration, the differences in the mean BIAS against observations is calculated between the two considered simulations (A minus B). In this way we retain a distribution of the differences in the mean BIAS that, for a given significance level α, allows us to determine whether we can reject or accept the null hypothesis. In order to decide whether a given metric (e.g. mean BIAS) of model A and model B are significantly different, we simply consider the lower and upper bounds of a confidence interval determined by the significance level α. Then, we distinguish three possible cases:

- A) The entire confidence interval lies entirely to the right of 0: that indicates that the mean BIAS of model A and model B significantly differ, with model B improving over the performance of model A. We can basically state that the mean bias of B is significantly smaller than the one of model A.
- B) The entire confidence interval lies to the left of 0: that indicates that the mean bias of model A and model B significantly differ, but in this case model B presents worse performance compared to model A. Here we can affirm that the mean BIAS (or any other considered metric) of B is significantly larger than the one of model A.
- C) 0 lies within the confidence interval: we can affirm that the mean bias of model A and model B do not differ significantly.

For RMSE, MAE, AMSESS, and MAD the same procedure is followed.

In order to decide whether the bias between two distributions is strongly or moderately different, we consider for the different metrics different values of the significance level α, as reported in Table 4. In each case, two values of α are considered: a moderate (α_m) and a strong (α) one.

Bootstrapping has the advantage that it does not require an error distribution to follow a specific distribution type. However, it is time-consuming. For LCorr, due to time constraints, we conduct the analyses following the method of Zou et al. 2007 [22] instead of bootstrapping.

The significance test procedure for error metrics explained here is done separately for all atmospheric variables under consideration, for all four seasons and each grid-box of the whole analysis domain. The ratio of points with significant differences in the error distribution between reference simulation and model experiment is then calculated for each domain of interest, following the procedure that is described in the following subsection.

3.4.2 Score points of evidence - ScoPi

For a certain model region containing N_g grid points, we count in a first step the number of grid points presenting a strong improvement of model B with respect to the reference simulation A (N_{s+}). The same is done for the grid points showing a strong worsening of model B compared to the reference simulation A (N_{s-}). The grid points presenting no strong significant improvement or worsening at all are not considered. Afterwards, we repeat this procedure but counting the grid points with respect to a moderate significance: (N_{m+}) and (N_{m-}).

The proportion of the different grid points is then calculated using the two following formulas:

$$F_{ss} = \frac{N_{s+} - N_{s-}}{N_g} \ , \ \ F_{ms} = \frac{N_{m+} - N_{m-}}{N_g} \tag{3}$$

Depending on the shares of grid points with strong and moderate significance (F_{ss} and F_{ms}), a final score for a certain model region is calculated, defined here as the **Score Points of evidence, ScoPi**:

$$\text{ScoPi} = \begin{cases} 2 \cdot F_{ss} & \text{if } F_{ss} > 0.4 \\ 1 \cdot F_{ms} & \text{if } F_{ss} \leq 0.4 \,\text{and}\, F_{ms} \geq 0.4 \\ -1 \cdot F_{ms} & \text{if } F_{ss} \geq -0.4 \,\text{and}\, F_{ms} \leq -0.4 \\ -2 \cdot F_{ss} & \text{if } F_{ss} < -0.4 \\ 0.0 & \text{otherwise} \end{cases} \tag{4}$$

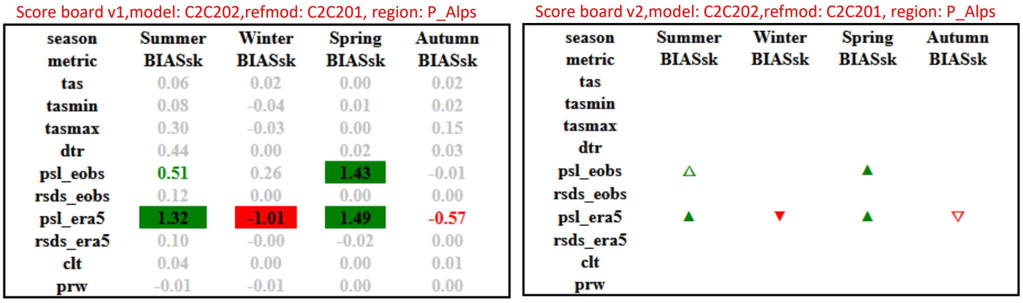
That means that the ScoPi score is equal to the share of grid points in a model region presenting a significant improvement or worsening with respect to the reference run. For a given experiment, if the share of significant grid points is too small (lower than 0.4), the ScoPi score is zero. In this case, it is assumed that there is no noticeable and large-scale change in the model results with respect to the reference. On the other hand, if the share of grid points in a model region with a strong (not only moderate) significance is larger than 0.4, the score is doubled.

The threshold of 0.4 applied in Eq. (4) ensures that the majority of the locations in a specific region and for a specific model configuration show a significant improvement / worsening compared to the reference experiment. For instance, if at least 40 % of the grid points show significantly improved performance for simulation B with respect to the reference A, the ScoPi score is 0.4, although 60 % of the grid points show no significant changes. The same is true if 60 % of the grid points show significant improvement and 20 % significant worsening.

The ScoPi score is computed for each PRUDENCE region (see Figure 1), for different metrics, seasons and various atmospheric variables, separately. This allows us to gain a comprehensive understanding of the reasons for possible improvement / worsening in the model performance for a given configuration. As an example, Figure 3 presents the ScoPi scores for the model experiment C2C202 calculated against the reference experiment C2C201 for the PRUDENCE region 'Alps'. The coloured scores indicate that the portion of grid points with 'moderate' significant improvement / worsening exceeds the threshold of 40% (see Eq. (4)), whereas the scores with coloured background indicate 'strong' significant improvement / worsening. The overview of the ScoPi scores calculated for single variables, seasons, regions and a selected metric is called 'Score board' in this report. As shown in Figure 3, the mean sea level pressure in spring and summer is improved with strong significance in C2C202 over the reference simulation, thanks to the new fast waves dynamics.

Table 5: ScoPi weights according to the region area (1) or to distance from the domain center (Germany) (2).

PRUDENCE Region	Alps	British Islands	East Europe	France	Iberian Peninsula	Mediterranean	Mid Europe	Scandinavia
weight1	0.0576	0.0589	0.2227	0.0639	0.1222	0.1168	0.1094	0.2484
weight2	0.2	0.0212	0.0802	0.0230	0.0440	0.0421	0.5000	0.0895



Score board v1,model: C2C202,refmod: C2C201, region: P_Alps

season	Summer	Winter	Spring	Autumn
metric	BIASsk	BIASsk	BIASsk	BIASsk
tas	0.06	0.02	0.00	0.02
tasmin	0.08	-0.04	0.01	0.02
tasmax	0.30	-0.03	0.00	0.15
dtr	0.44	0.00	0.02	0.03
psl_eobs	0.51	0.26	1.43	-0.01
rsds_eobs	0.12	0.00	0.00	0.00
psl_era5	1.32	-1.01	1.49	-0.57
rsds_era5	0.10	-0.00	-0.02	0.00
clt	0.04	0.00	0.00	0.01
prw	-0.01	-0.01	0.00	0.00

Score board v2,model: C2C202,refmod: C2C201, region: P_Alps

season	Summer	Winter	Spring	Autumn
metric	BIASsk	BIASsk	BIASsk	BIASsk
tas				
tasmin				
tasmax				
dtr				
psl_eobs	△		▲	
rsds_eobs				
psl_era5	▲	▼	▲	▽
rsds_era5				
clt				
prw				

Figure 3: Score boards showing an overview of ScoPi values obtained for the *Alps* region, comparing the mean BIAS against observations of simulation C2C202 with respect to the reference simulation C2C201. Left: ScoPi scores based on the differences in the mean BIAS against observations calculated, point-by-point, between simulation C2C202 and the reference C2C201. Positive (negative) numbers indicate better (worse) performance of simulation C2C202 with respect to the reference. Right: visual synthesis of the results of the left table using coloured triangles, where green stands for improvement and red for worsening, while the open or filled symbols indicate strong or weak differences, respectively.

The great advantage of the ScoPi score is its inherent standardisation for different types of variables. The values always range between -2 and 2. Therefore, the ScoPi score can be aggregated over different types of variables and metrics. The aggregation with respect to a specific PRUDENCE region is done by calculating the sum over the ScoPi values for all seasons, metrics, and variables. For integrating over different variables, different weights are considered each time, as given in Tab. 3. The resulting value is named ScoPi$_{region}$. Finally, the ScoPi$_{simulation}$ is defined as a weighted sum over all PRUDENCE regions, considering additional regional weights. A different weight is assigned for each PRUDENCE region, considering either its area or its distance to mid-Europe (domain A4 in Fig. 2). The weights for the PRUDENCE regions are given in Tab. 5.

As an example, the boxplot in Figure 4 shows the ScoPi$_{region}$ calculated for all variables and seasons, separately for all PRUDENCE regions (coloured points) and each simulation of the first phase of COPAT2 (rows). In addition, the ScoPi$_{simulation}$ is given on the y-axis labels. Using this so-called ScoPi-plot, it is possible to obtain a quick overview of the performance of the different COPAT2 experiments.

The ScoPi is calculated in a first place considering the mean BIAS against observations for all of the given variables beside precipitation. Then, it is also applied to other metrics, such as the RMSE (not shown here) the AMSESS, and the MAD only for precipitation. Additionally, the ScoPi is also calculated using as error metric the Pearson linear correlation and as significance test for the differences between two simulations the method proposed by Zou [22]. Finally, the same ScoPi analyses are further reiterated considering as input the original data interpolated on grids with different resolution, as well as different variable weights.

3.4.3 Additional Metrics

To support the results of the analyses based on the Score Points of evidence, we additionally perform an evaluation of the different simulations against ERA5, in a similar way as in [4]. Considering the same variables as described in section 3.3 (beside MSLP, currently not considered), in a first step we calculate for each variable and simulation spatial averages of 3-daily means for the different PRUDENCE regions. Afterwards, we calculate the RMSE of the obtained time-series of spatial means against ERA5. Finally, we compare the RMSE obtained for a given simulation *sim* against the one of the reference experiment ref on the base of the following Skill Score (SS):

$$SS_{var} = \frac{(RMSE_{var})_{ref} - (RMSE_{var})_{sim}}{(RMSE_{var})_{ref}} \times 100 \tag{5}$$

where *var* indicates the considered variable. When SS is positive (negative), experiment *sim* leads to an improvement (worsening) of the results for a given variable with respect to the reference run ref. For this analysis, no weights are assigned to the different regions.

The main advantage of this approach is, again, to make the different variables more Gaussian through averaging. At the same time, by considering spatial means over large areas of the study domain, we can additionally expect to reduce the error due to the model sensitivity to the initial conditions.

Finally, for each of the phases of the evaluation process, we complement the given analysis with a comparison of the maps of the mean BIAS against observations of the different simulations. These maps are shown here only for the final recommended simulation.

4 Results

4.1 Phase Ic

The analysis of the Phase Ic simulations in terms of ScoPi shows that a large number of different settings improve the results compared to the reference run (Fig. 4). This is particularly true for the new model dynamics (C2C202 to C2C205) and for changes in the model physics, such as the use of a new formulation of surface skin temperature and bare soil evaporation (C2C210 and C2C212). The experiment with a new formulation for ground water budget (C2C214) does not improve the results in all of the considered regions. The changes in the turbulence settings show nearly no improvement with respect to the reference run. The results of the experiments with the new dynamics show that changes to the model configuration including a modified dynamic bottom boundary condition (ldynbb=False) as well as the Mahrer discretization lead to improved model results. The latter particularly improves the model performance in regions with high mountains. It is important to notice here that the improvement produced by the changes in model dynamics is not obtained for all of the considered regions though: a remarkable worsening of the results is obtained for the Iberian Peninsula and Scandinavia in this case.

For precipitation (not shown), no significant improvement is obtained in terms of the mean BIAS for the performed experiments of phase I.

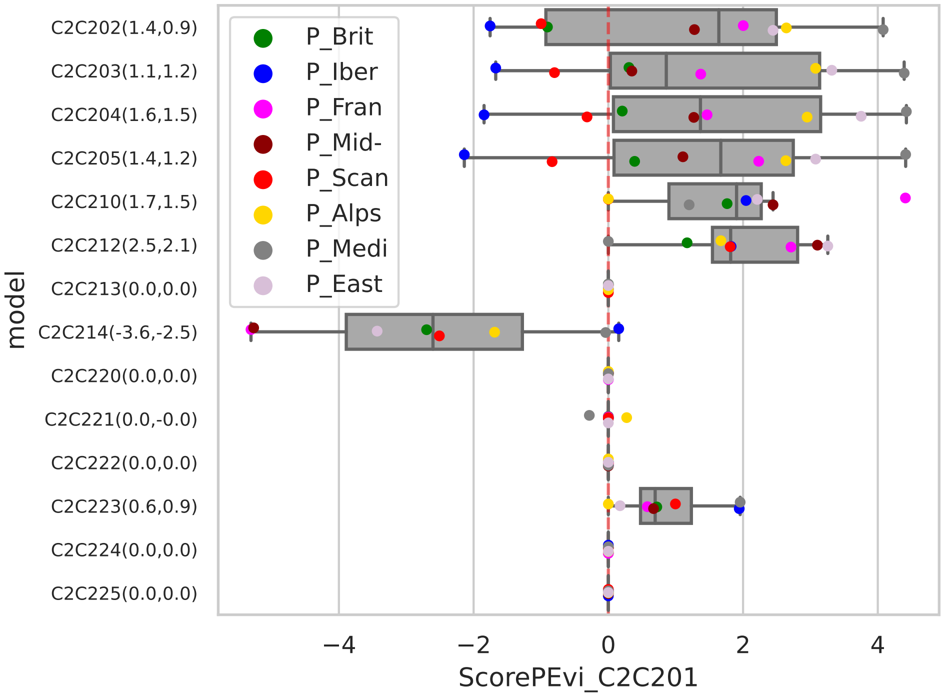


Figure 4: ScoPi$_{region}$ based on the differences in the mean BIAS between the observations and each simulation of Tab. 1, against the ones of the reference simulation C2C201. The colors indicate the different CORDEX regions. The numbers given on the y-axis labels are the ScoPi$_{simulation}$. The first value considers the area of a region, the second considers the distance to Mid-Europe (see Tab. 5).

4.2 Phase IIc

Figure 5 (upper panel) shows for the experiments of Phase IIc (see Tab. 2) that the combination of the most promising changes of Phase Ic in the model dynamics and the representation of surface processes have in general a positive impact on the quality of the simulations, in all regions. Additionally, the results of phase IIc confirm that the new formulation of soil moisture budget does not improve the results of our analysis against the reference. Nonetheless, the BIAS against observations for the experiments with the new formulation of soil moisture budget (C2C314 and C2C315) is now reduced with respect to the results of the same experiment conducted during phase Ic (C2C214). This is mainly attributable to the changes in the initial conditions and external forcings required for this configuration and that were not considered before: a new parameter for the description of the orography (S_ORO_max) and a longer spin-up time. The best configurations of this new set of experiments are C2C316 and C2C317 for which, beside the changes considered in the experiment C2C303, also a different convection scheme is considered. This in particular leads finally to significant changes in the representation of precipitation with respect to the reference run C2C301 (Fig. 5, lower panel), although only for some regions. The experiment C2C318 shows improvements with respect to the reference run comparable to the ones obtained for C2C316 and C2C317. However, after consulting the developers of the COSMO turbulence scheme from the Deutscher Wetterdienst (DWD), it was decided to discard experiment C2C318 as it does not align with their recommendations, even though this experiment is still considered in some of the successive analysis.

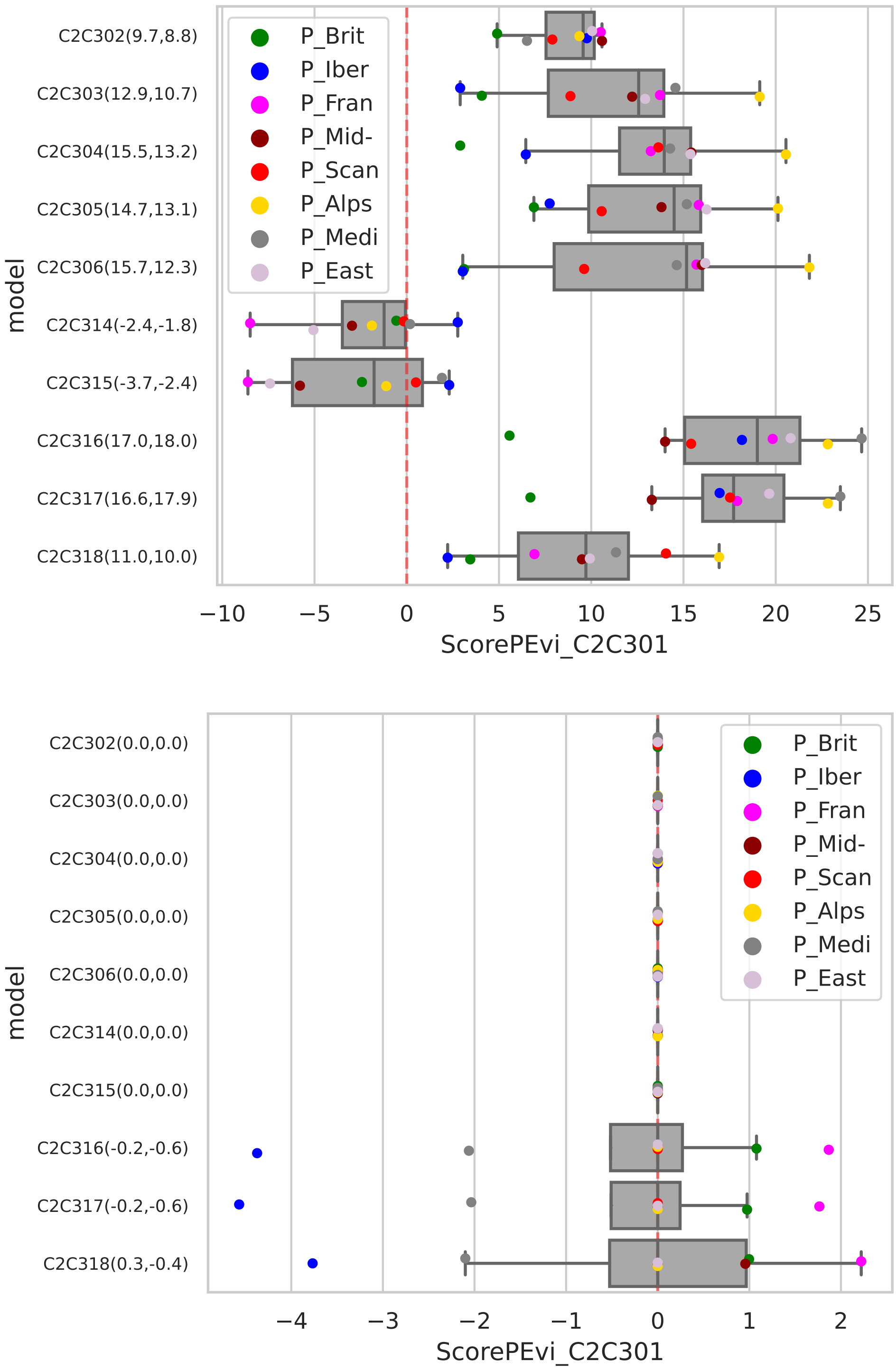


Figure 5: ScoPi-plot based on the differences in the mean BIAS between the observations and each simulation of Tab. 3 (excluding experiments C2C3xx), against the ones of the reference simulation C2C301. The colors indicate the different CORDEX regions. Top: All variables but precipitation. Bottom: Precipitation.

4.2.1 ScoPi based on different spatial aggregation, metrics and variables

The validity of the given results is additionally tested by repeating the same ScoPi analyses considering different spatial aggregations. Both the simulation results and the observations are upscaled from their original grid to several regular grids with spatial resolutions of 0.5°, 1°and 1.5°. Fig. 6 shows the ScoPi-plot calculated for the mean BIAS for all the simulations of Phase IIc, using as input the data on the grid with a spatial resolution of 1°. The results are very similar to the ones derived from Fig. 5 and the best performing simulations are again the ones with combined changes in the model dynamics and in the representation of surface processes (C2C316cg10 and C2C317cg10).

Additionally, we have further tested the robustness of the results of the ScoPi analysis considering, besides the mean BIAS, also the AMSESS and the LCorr. The aggregated results for the three metrics are shown in Fig. 7. The results are again very similar to the ones derived from Fig. 5. In particular, Fig. 7 shows that the best-performing simulations are again those with combined changes in the model dynamics and the representation of surface processes (C2C316 and C2C317). This is also true when considering the results of the ScoPi plot separately for each metric (see Fig. 16, in the Appendix).

Finally, we have also calculated the ScoPi based on the mean BIAS, but changing the weights of the considered variables for the best performing experiments of Phase IIc (namely C2C303, C2C316 and C2C317). The results are presented in Fig. 16, in the Appendix. Again, simulations C2C316 and C2C317 clearly improve in each case the performance of the reference run in the comparison against observations.

4.2.2 Ranking based on a different evaluation approach

In a final step, to further support the outcomes of the ScoPi analysis for the experiments of phase IIc, we also apply a different evaluation procedure based on the RMSE calculated between the different simulations and ERA5, as detailed in section 3.4. Fig. 8 shows the results of the Skill Score (SS) based on the RMSE (3.4.3) calculated for each variable separately, between each of the simulations of Phase IIc and the ERA5 reanalysis data, and considering as reference experiment C2C301. Fig. 8 shows again that the simulations with the most pronounced improvements with respect to the reference C2C301 are obtained for experiments C2C316 and C2C317. In this case, only for these two runs an improvement is evident for each of the considered variables. It is important to notice that for precipitation, both C2C316 and C2C317 lead to an improvement that is of the same magnitude as the one obtained for the other variables.

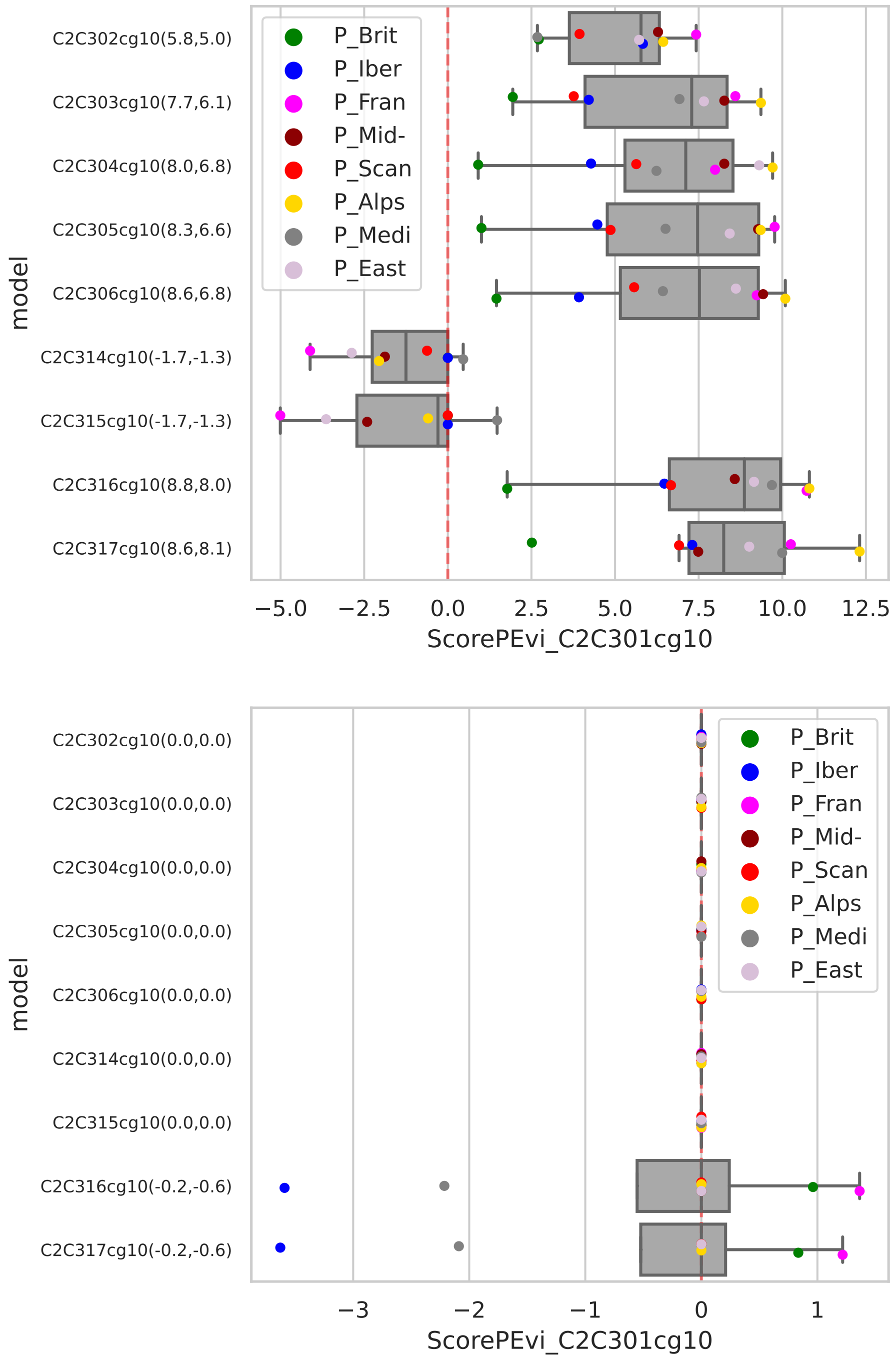


Figure 6: ScoPi-plot based on the differences in the mean BIAS between the observations and each simulation of Tab. 3 (only experiments C2C3xx), against the ones of the reference simulation C2C301, but considering the original data interpolated onto a grid with a spatial resolution of 1°lon. The colors indicate the different CORDEX regions. Top: All variables but precipitation. Bottom: Precipitation.

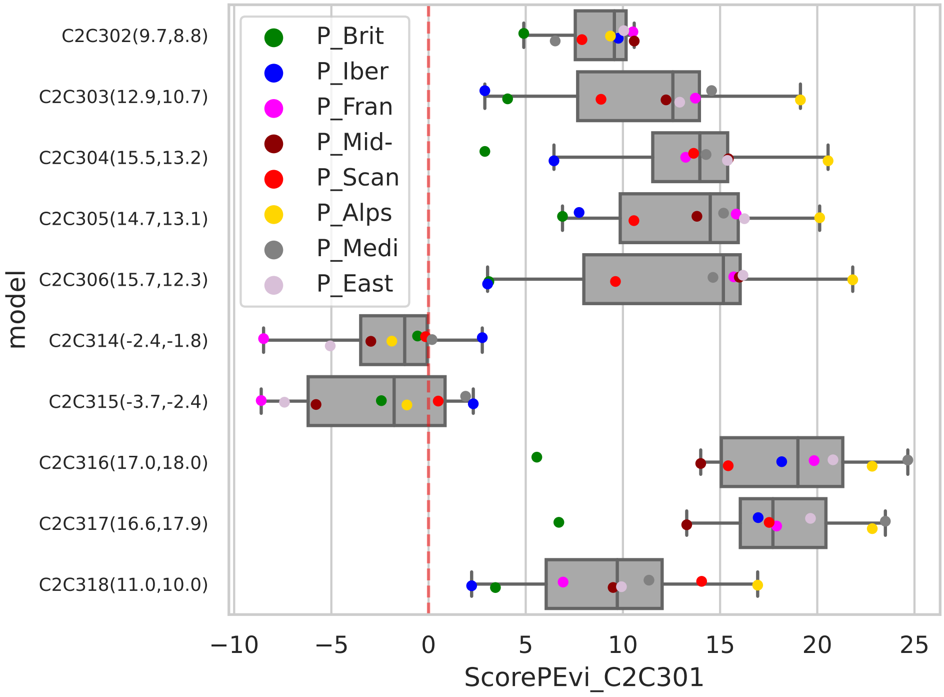


Figure 7: ScoPi-plot based on the differences in the mean BIAS, AMSESS and LCorr between the observations and each simulation of Tab. 3 (only phase IIc experiments C2C3xx), against the ones of the reference simulation C2C301. The colors indicate the different CORDEX regions. The plot shows the ScoPi calculated for all variables, except precipitation.

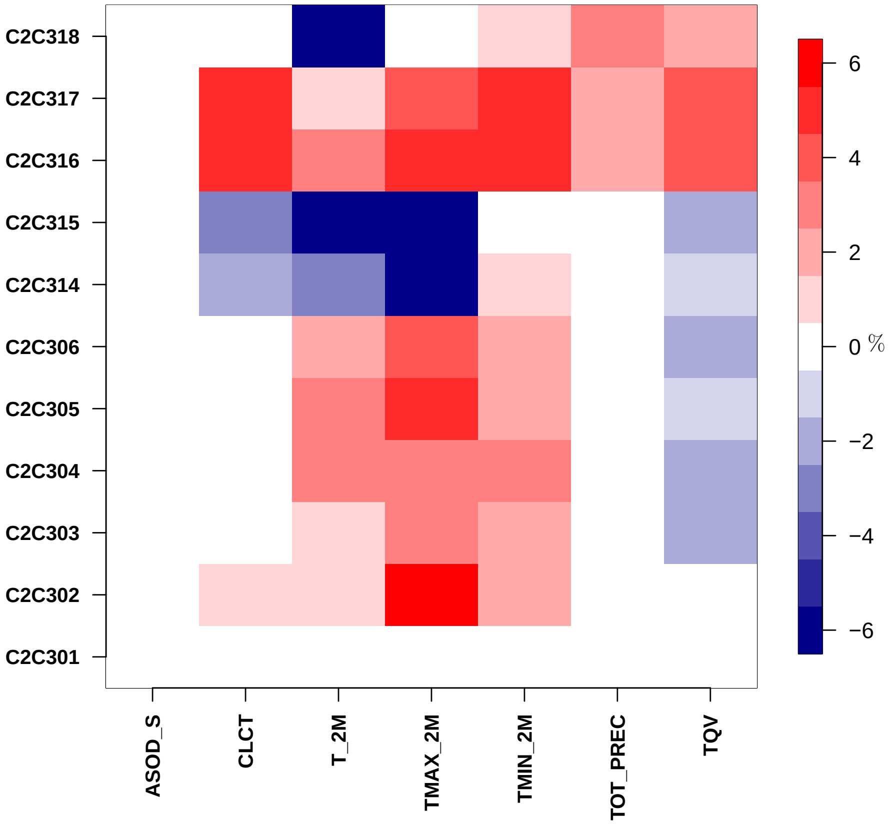


Figure 8: Skill Score based on the RMSE, as calculated according to equation 3.4.3 for each simulation of phase IIc, with respect to ERA5 and considering as reference the simulation C2C301, over the period 1981-1985.

4.2.3 Comparison against Radiosondes

We additionally conduct a seasonal analysis of the mean BIAS of COSMO-CLM results against radiosondes. Fig. 9 shows the vertical distribution of the mean BIAS of the simulation C2C316 against observations, calculated for (from left to right) temperature, relative and specific humidity, and wind speed for the PRUDENCE regions Mediterranean (upper panel) and Mid-Europe (bottom panel). The overview for the other regions is given in the Appendix A.6. For all regions except Scandinavia, we find an overestimation of the area- and seasonal mean air temperatures at 850 and 750 hPa. On the other hand, all regions except the British Islands present a positive bias for temperature in the upper levels, at 300 and 500 hPa. For humidity, the model underestimates the values derived from the radiosondes for East Europe, France, Mid-Europe, and Scandinavia. Conversely, the model presents a positive bias in humidity for the Alps, the Iberian Peninsula, the Mediterranean, and partly the British Islands. For all regions, the wind speed is overestimated in the model in the bottom layers, up to 400 hPa. Conversely, in the upper levels, simulation values better match the observed ones, even though the model presents in this case a negative bias. In the Appendix (Fig. 22 and Fig. 23) we also give an example of the seasonal spread of the 3-daily-mean bias calculated between simulation C2C316 and the radiosonde measurements for Mid-Europe. Figure 22 shows the results for the period 1981-1985 and Fig. 23 for the period 2002-2008.

In the next part, we calculate the ScoPi$_{region}$ metric using radiosondes data. Importantly, for the radiosondes we do not have data at each time-stamp and grid-box, as for the model. In this case, 3-daily means used as input for the analysis of ScoPi$_{region}$ are calculated for each single balloon launch, considering its effective location during its ascend. To assign a value of a radiosonde to a specific model time stamp we consider only the launches occurring in a window of +1 and -1 hour, at the regular reporting times of a given day. For a given balloon launch, information at each of these time stamps might not be available during a 3-day period for specific locations (especially high in the atmosphere). Consequently, for the comparison of model results against radiosondes, we calculate 3-daily means only considering model data at times when information from radiosondes is available. This ensures consistency and comparability between the two datasets. The ScoPi$_{region}$ metric is calculated according to Eq. 4, but with different weights considered in the aggregation of the results for the different variables, as indicated in Tab. 6. In order not to give too much weight to the variables for humidity, we assign to specific and relative humidity together the same weight as air temperature and wind speed. However, we decided to assign to relative humidity a weight of an order of 3 larger than for specific humidity, taking into account that the latter is extremely low at the upper levels.

Table 6: Weights for different variables for the calculation of the aggregated ScoPi$_{region}$ metric for upper air evaluation.

Variable	Weight
air temperature	1.0
relative humidity	0.75
specific humidity	0.25
wind speed	1.0

Although the upper air analysis is not taken into account during the selection of the most promising setup, here we show in Fig. 10 the synopsis of ScoPi$_{region}$ and ScoPi$_{simulation}$ for all simulations from Phase IIc, whereas the results of the Phase Ic are shown in the Appendix A.6, Fig. 20. To test whether the assessment of the model quality changes with a change

Table 7: ScoPi$_{region}$ for C2C316 and C2C316sp

	300 hPa	500 hPa	700 hPa	850 hPa
C2C316	-0.6 / 2.1	6.8 / 6.4	-6.6 / -6.3	1.4 / 1.4
C2C316sp	1.6 / 4.7	9.4 / 7.7	-2.3 /-0.4	0.8 / 0.5

to more recent times, when more satellite information is assimilated into the forcing data, modern radiosonde instruments are employed and aerosol distribution in the atmosphere has changed, we repeat the same analysis for the period 2002 to 2008, for the simulations C2C303, C2C316 and C2C317. The final results are not affected: the 700 hPa level remains problematic, and the other values remain in the same range (see Appendix A.6, Fig. 21).

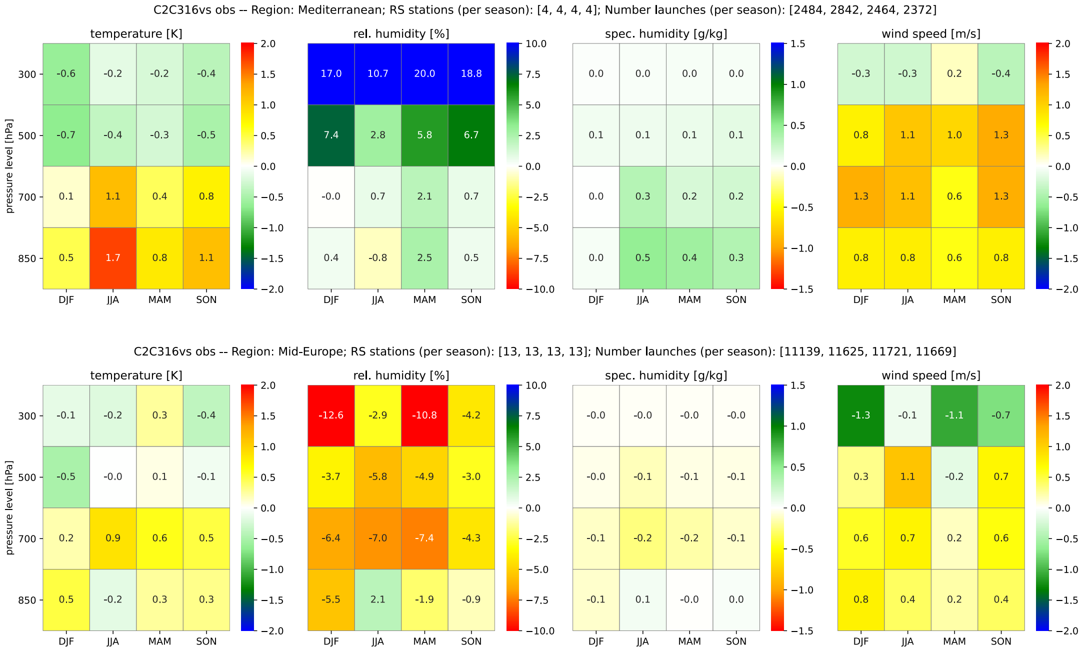


Figure 9: Comparison of the simulation C2C316 to radiosonde observations: mean seasonal BIAS calculated over the period 1981-1985 for the variables (from left to right) temperature, relative humidity, specific humidity, and wind speed. In the upper panels, the results for Mid-Europe are presented, while the ones for the Mediterranean region are shown in the bottom panels.

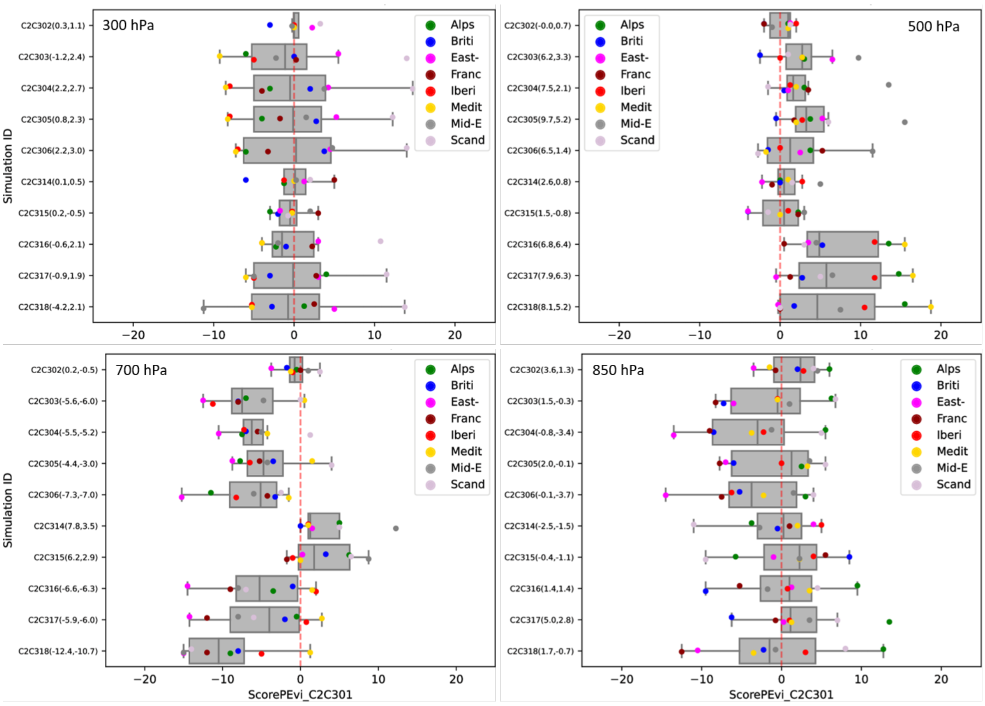


Figure 10: ScoPi-plot based on radiosonde data, calculated for Phase IIc simulations by comparing them to the reference run C2C301. The data is plotted for altitudes of 300, 500, 700, and 850 hPa, covering the time span from 1981 to 1985.

4.3 Phase IIIc

4.3.1 Different integration periods

One important point to consider for proving the robustness of the results obtained in the previous two phases is to test whether the most promising simulations show the same performances in the comparison against the reference run, even when integrated over different time periods. Therefore, the best performing simulations of the previous phases, experiment C2C316, C2C317, and C2C303, as well as the reference C2C301, are further extended until the year 1990. In addition, experiments with these configurations for the period 2001-2008 are also performed (only the years 2003-2008 are considered for the analysis in this case). The results of a simulation with the recommended configuration of COSMO-CLM version COSMO5.0_clm9 (C2C100) are also included in the analysis this time.

The ScoPi score obtained for this new set of simulations is presented in Fig. 11 (upper panel). The results confirm that for the comparison of the mean BIAS obtained for all of the considered variables besides precipitation, the new configurations lead to an improvement of the results of the reference run against observations. This holds true for all of the considered regions and also for the integration period 2003-2008. Importantly, the three newly tested configurations (i.e. C2C303c, C2C316c, C2C317c) have remarkably better performances than the former reference C2C100. In this case, though, only experiments C2C316c and C2C317c show a clear improvement of the results for the considered metric.

For precipitation, significant improvements in the MAD are obtained for France for simulations C2C316c and C2C317c. However, these simulations show worse performance with respect to the reference run for the Iberian peninsula and the Mediterranean region. For the latter region, the best results are obtained using the reference configuration of the former model version C2C100.

4.3.2 Longer integration times

Given the results of the previous analyses, and taking into account the fact that a simulation with the setup of C2C317 appeared to be unstable when integrated over longer time periods (violation of the Courant-Friedrichs-Lewy condition), configuration C2C316 is selected as the best-performing configuration. Hence, a new simulation for the entire evaluation period 1979 - 2020 is performed with the setup of experiment C2C316. For testing whether this configuration leads to significant improvements compared to the recommended configuration of COSMO 5.0_clm9 (C2C100), the score points of evidence are then calculated for the entire period 1980 to 2020, choosing experiment C2C301 as the reference in this case.

The results shown in Fig. 12 demonstrate that the new model version, with the setup of experiment C2C316, leads to significant improvements with respect to the reference configuration of the former model version. This is true for all of the considered variables and regions. The only exceptions are the Iberian peninsula and the Mediterranean regions in the case of the ScoPi based only on precipitation.

As examples for the improvements obtained for experiment C2C316 with respect to the recommended configuration of COSMO5.0_clm9 (C2C100[††]) , we show two highlights as 2D maps: the mean BIAS of summer 2-meter temperature and monthly precipitation sum calculated in the two cases against E-Obs (see Fig. 13 and 14). 2D maps for all seasons are shown in the Appendix Fig. 17 and 18).

The comparison of 2 m temperature shows a remarkable improvement of the results of experiment C2C316 compared to C2C100 (Fig. 13). There is a large reduction of the cold model BIAS in Scandinavia and a very large reduction of the warm BIAS in South-Eastern Europe. Importantly, the warm BIAS over the latter region was a long-standing problem that characterised many previous versions of the COSMO model. Unfortunately, these improvements go along with an overall shift to cooler temperatures in the central part of the domain. However, this doesn't compromise the general improvements obtained with the configuration C2C316.

The improvements of experiment C2C316 over the one with the latest recommended configuration of COSMO-CLM 5.0 (C2C100) are less pronounced for precipitation than in the case of temperature, when considering the period 1980-2020 (Fig. 14). There is a reduction of the wet model BIAS in France, mid and eastern Europe. In particular, a reduction of a very pronounced negative precipitation BIAS that characterised the results of experiment C2C100 over the Balkan region is now largely reduced. However, the results of C2C316 appear to generate generally drier conditions over the entire domain. But, as demonstrated by the score points of evidence (Fig. 12), there is an overall improvement also for precipitation in all regions except the Iberian Peninsula and the Mediterranean.

4.3.3 Influence of the land cover datasets GLC2000 and GLOBCOVER2009

In a very concluding step we have performed an additional simulation with the same setup as experiment C2C316 for the period 1979-1985, but using as land cover input the GLOBCOVER2009 dataset instead of the GLC2000. In general, we have found that the new model

[††] Configuration in namelist tool: `https://tools.clm-community.eu/NLT/tmp/upload_202204200911_9412fa6c76864bea2bd3.txt`

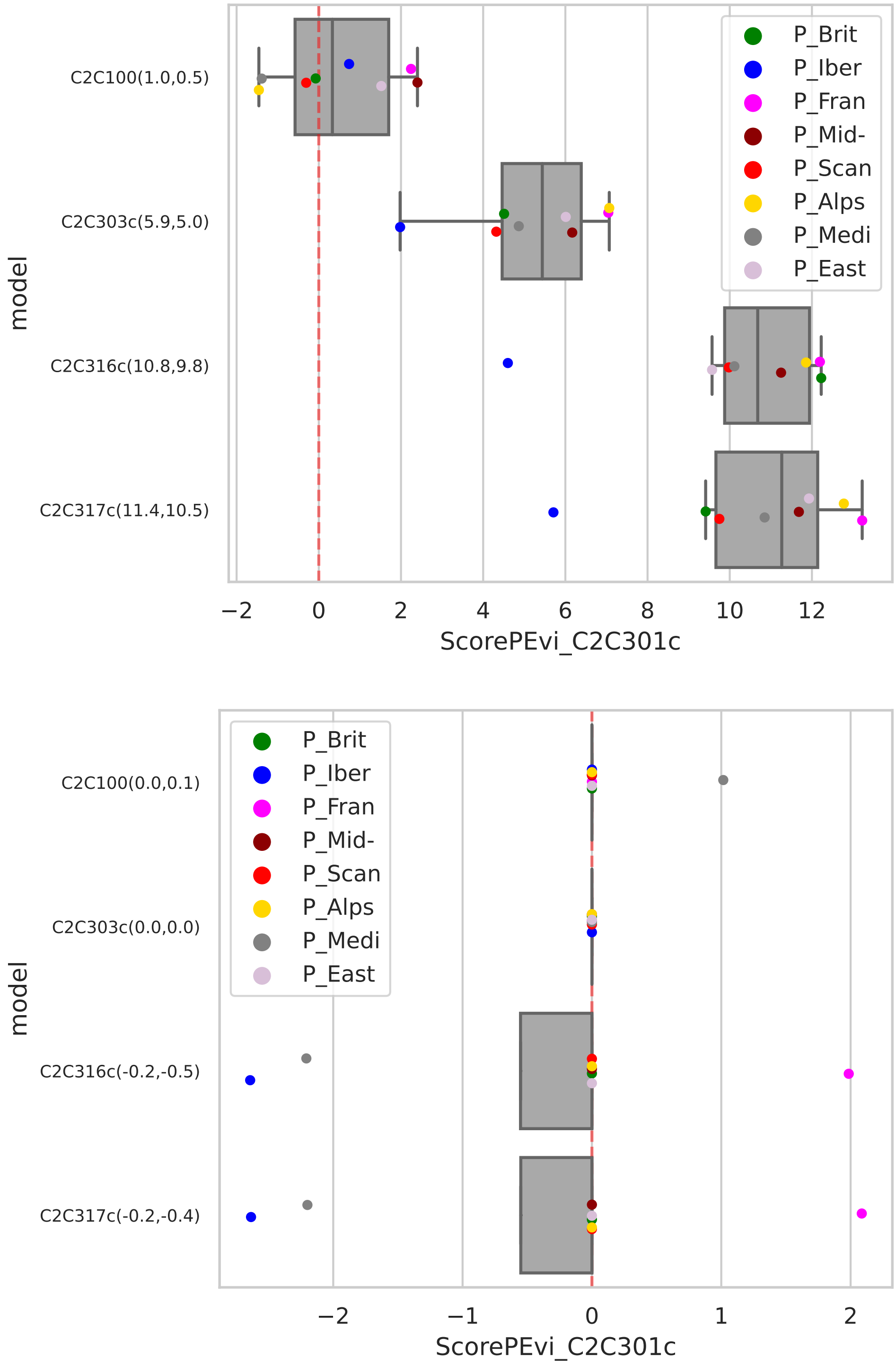


Figure 11: ScoPi-plot based on the differences in the mean BIAS (upper panel) and the MAD for precipitaion (bottom panel) between the observations and each realisation of Tab. 3 (only experiments C2C3xxc), against the ones of the reference simulation C2C301 for the period 2003-2008. In the first row of the plots the results for simulation C2C100, the former reference of COSMO-CLM 5.0, are added. The colors indicate the different CORDEX regions. Top: All variables but precipitation. Bottom: Precipitation.

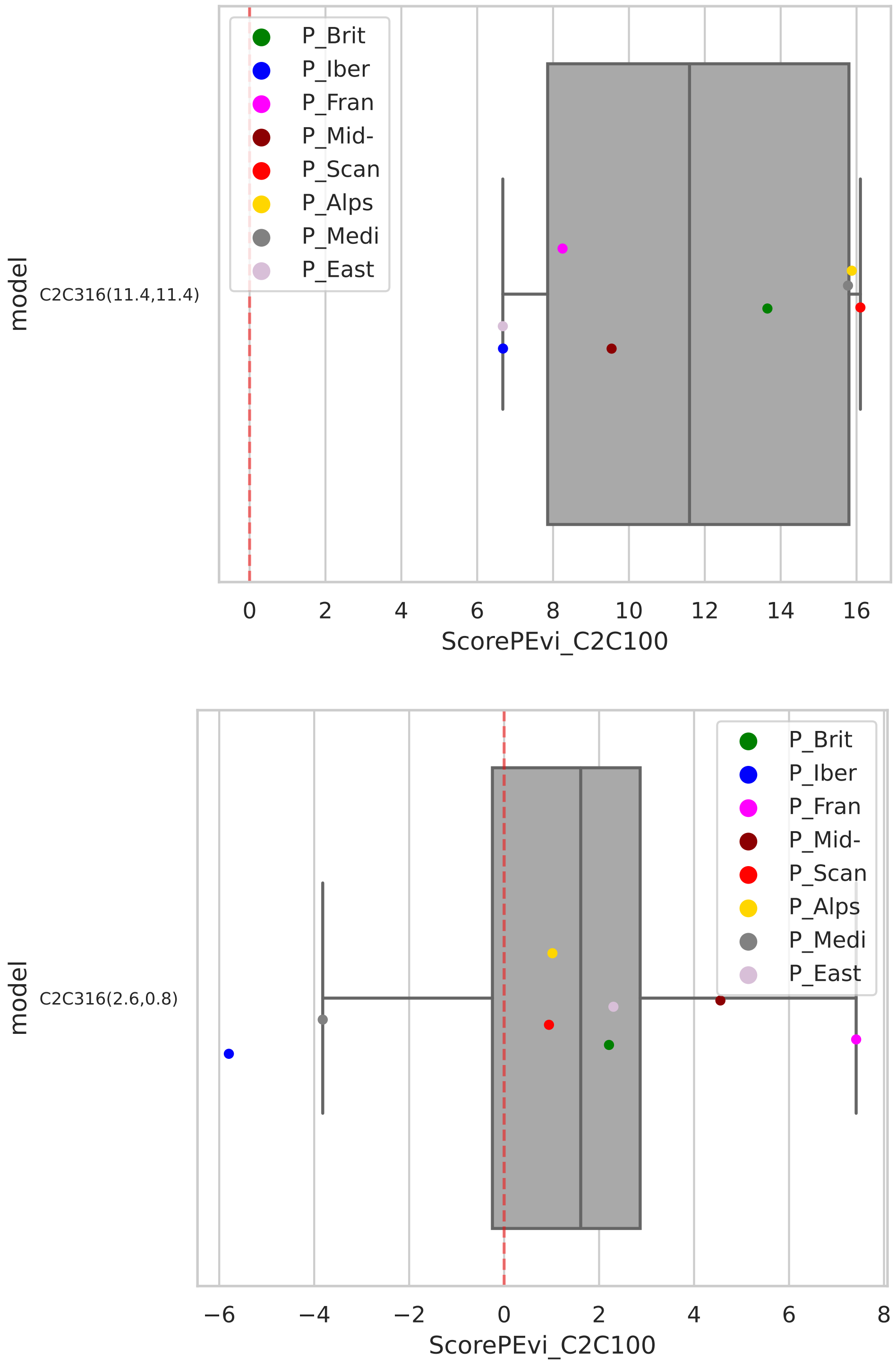


Figure 12: ScoPi-plot based on the differences in the mean BIAS (upper panel) and the MAD for precipitation (bottom panel) between the observations and simulation C2C316, against the ones of the simulation C2C100, calculated over the period 1980-2010. The colored dots indicate the different CORDEX regions. The numbers provided in brackets on the y-axis (beside the different experiment names) are weighted means of all regions. The first weight considers the area of a region. The second weight considers the distance to Mid-Europe. Top: All variables but precipitation. Bottom: Precipitation.

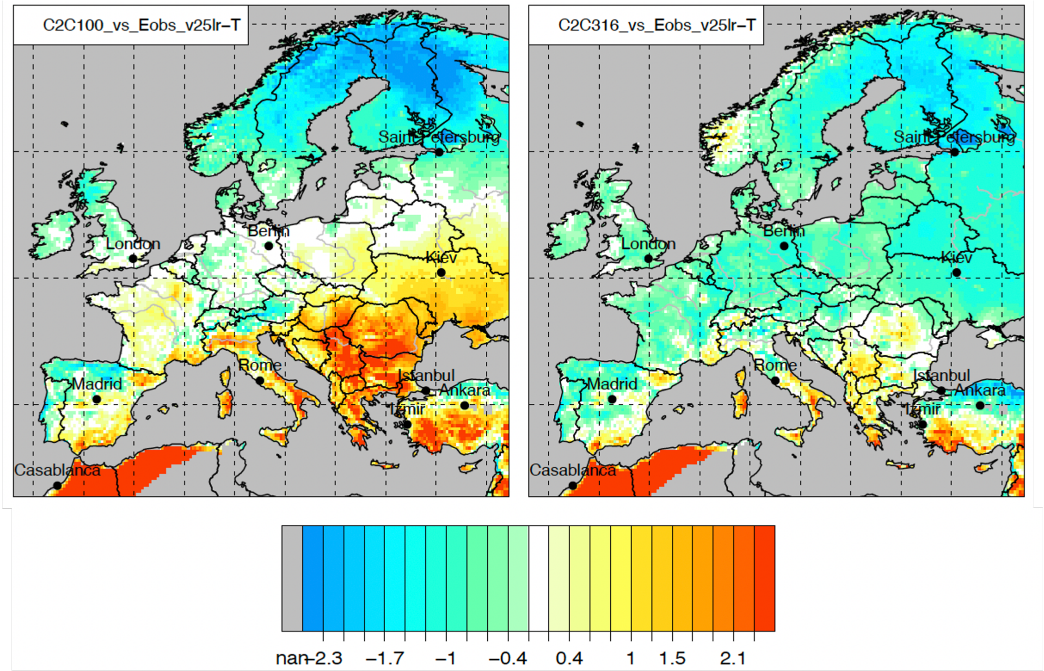


Figure 13: Comparison of the recommended version of 2018 (C2C100,*left*) and experiment C2C316 (*right*) in terms of mean summer bias of 2 m temperatures [K] calculated against Eobs for the period 1980-2020.

results are consistent with the ones of experiment C2C316, with some differences mainly for 2-meter temperature. When analysing the mean monthly BIAS of 2-meter temperature for single regions, in many cases experiment C2C316er_bg actually leads to a reduction of the model BIAS against observations with respect to C2C316, such as for Switzerland and the Balkan region (Fig. 15). Other times though a slightly larger BIAS is observed for simulation C2C316er_bg compared to C2C316. However, the differences between the two simulations are always not very pronounced, such as in the case of Spain where the two experiments

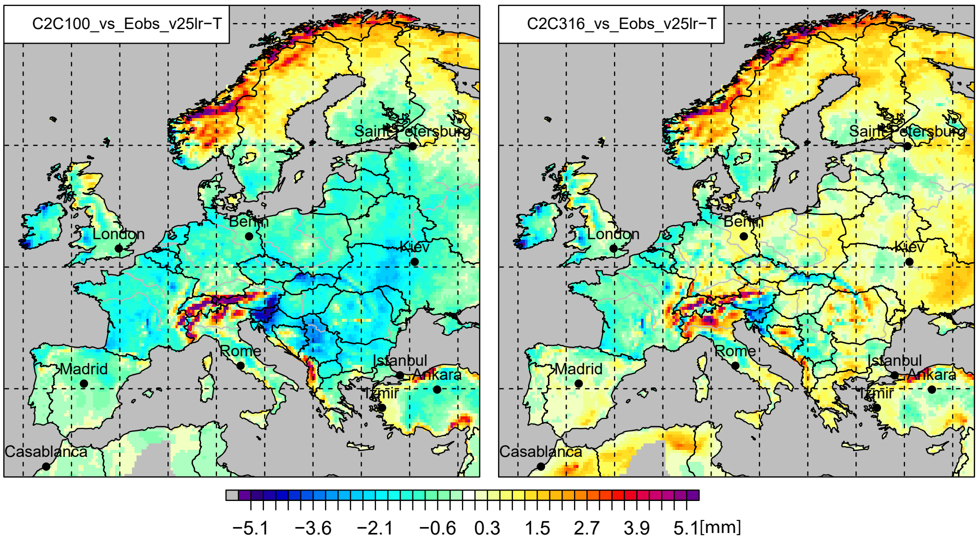


Figure 14: Comparison of the recommended version of 2018 *left*) and experiment C2C316 (*right*) in terms of mean summer BIAS of monthly precipitation sums [mm/mon] compared to Eobs for the period 1980-2020.

differ by no more than 0.2 °C. In general, the use of a more recent land cover dataset such GLOBCOVER2009 should possibly be preferred to less recent ones such as GLC2000. However, we have shown here that the provided optimal setup of experiment C2C316 leads to improved model performances for both the considered sets of data, confirming that our main conclusions hold true even when considering different surface boundaries.

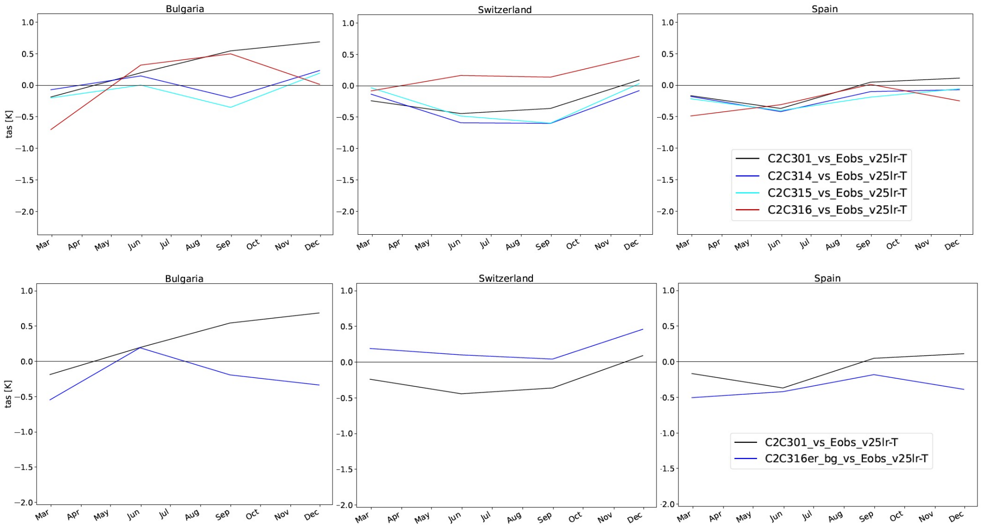


Figure 15: Regional mean BIAS of 2-meter temperature calculated between simulations C2C301, C2C314, C2C315, C2C316 (upper row) and C2C316er_bg (bottom row) and the EOBS observational data. From left to right, the mean BIAS calculated in each case for Bulgaria, Switzerland and Spain is shown.

5 Conclusions

This report summarizes the results of the extensive tests and analysis that have been conducted to assess the performance and optimize the configuration of COSMO-CLM 6.0. The work has been done in the COPAT2 project of the CLM Community working group Evaluation (EVAL). The followed evaluation procedure mainly focused on testing new options and parameterisations of COSMO-CLM 6.0, but no tuning parameters. The reasons for this decision are illustrated in the introduction section.

For the evaluation of the model results we introduced a new score, the score points of evidence (ScoPi). This score combines the results for different variables, matrices and regions in a single number, hence simplifying the comparison and assessment of different simulations. It supports the general strategy that a preferably optimized configuration for the whole European domain should be determined. This can include a worsening of the results for some variables in some regions of the domain, as long as the overall results across all variables and most parts of the domain are improved.

According to the results presented in the previous sections, COSMO-CLM 6.0 with the configuration C2C316 has shown the best results, also with respect to the recommended configuration of the former model version COSMO-CLM 5.0. Consequently, COSMO-CLM 6.0 with the configuration of simulation C2C316 is proposed as the new recommended version of the COSMO-CLM model framework for the European domain by the working groups EVAL and SUPTECH. According to the regulations of the CLM Community agreement, the decision about the new recommended model version and the corresponding configuration will be taken by the community members at the CLM Community Assembly 2023 in Leuven, Belgium.

For the reasons mentioned in the introduction, the optimisation of the values for tuning parameters was not the focus of this initiative. Together with the aim of optimizing the overall model performance for a large domain, this leaves certainly some room for further improvements. Especially when looking at specific processes, output variables or domain sub-regions. Depending on the goals of a given simulation and the target region within the European domain, users of the model are encouraged to do more detailed tests to further improve model performance.

Users that want to apply the model in other regions than Europe are advised to do intensive tests to optimize the model set-up before starting with the production of simulations. The recommendation for the configuration given in this report is only valid for the European domain. For other domains, this configuration will very likely not provide optimal results. The procedure described in this report can be used to optimize also the performance of the model in other regions of the World. Set-ups of COSMO 5.0 that have been used for other domains can be used as starting points for the tests. Results of tests and improved configurations for other domains should be reported back and shared with the community via the working group EVAL or the coordination office.

References

[1] P. Alexander and A. De La Torre. Uncertainties in the measurement of the atmospheric velocity due to balloon-gondola pendulum-like motions. *Adv. Space Res.*, 47(4):736–739, 2011.

[2] M. (2013) Baldauf. A new fast-waves solver for the runge-kutta dynamical core. techreport 21, Deutscher Wetterdienst.

[3] Michael Baldauf, Axel Seifert, Jochen Förstner, Detlev Majewski, Matthias Raschendorfer, and Thorsten Reinhardt. Operational convective-scale numerical weather prediction with the cosmo model: Description and sensitivities. *Mon. Wea. Rev.*, 139(12):3887–3905, Apr 2011.

[4] O. Bellprat, S. Kotlarski, D. Lüthi, and C. Schär. Exploring perturbed physics ensembles in a regional climate model. *Journal of Climate*, 25(13):4582–4599, 2012.

[5] Margarita Choulga, Ekaterina Kourzeneva, Elena Zakharova, and Arkady Doganovsky. Estimation of the mean depth of boreal lakes for use in numerical weather prediction and climate modelling. *Tellus A: Dynamic Meteorology and Oceanography*, 66(1):21295, 2014.

[6] Jens Hesselbjerg Christensen and Ole Bøssing Christensen. A summary of the PRUDENCE model projections of changes in european climate by the end of this century. *Climatic Change*, 81(S1):7–30, mar 2007.

[7] Richard C Cornes, Gerard van der Schrier, Else JM van den Besselaar, and Philip D Jones. An ensemble version of the e-obs temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17):9391–9409, 2018.

[8] G. Doms and M. Baldauf. A description of the nonhydrostatic regional cosmo-model part ii: Dynamics and numerics. Technical report, Deutscher Wetterdienst.

[9] Imke Durre, Yin Xungang, Russell S. Vose, Scott Applequist, and Jeff Arnfield. Integrated global radiosonde archive (igra).

[10] Filippo Giorgi, Colin Jones, and G. Asrar. Addressing climate information needs at the regional level: The cordex framework. *WMO Bull*, 53:175–183, 11 2009.

[11] GLOBE-Task-Team. The global land one-kilometer base elevation(globe) digital elevation model, version 1.0. Technical report, NationalOceanic and Atmospheric Administration, National Geophysical Data Center.

[12] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

[13] T.O. Hodson. Root mean square error (rmse) or mean absolute error (mae): when to use them or not. *Geoscientific Model Development Discussions*, pages 1–10, 2022.

[14] NASA/LARC/SD/ASDC. Ceres energy balanced and filled (ebaf) toa monthly means data in netcdf edition4.1, 6 2019.

[15] J. Noilhan and S. Planton. A simple parameterization of land surface processes for meteorological models. *Monthly Weather Review*, 117(3):536 – 549, 1989.

[16] Uwe Pfeifroth, Steffen Kothe, Jörg Trentmann, Rainer Hollmann, Petra Fuchs, Johannes Kaiser, and Martin Werscheck. Surface radiation data set - heliosat (sarah) - edition 2.1, 2019.

[17] Linda Schlemmer, Christoph Schär, Daniel Lüthi, and Lukas Strebel. A groundwater and runoff formulation for weather and climate models. *Journal of Advances in Modeling Earth Systems*, 10(8):1809–1832, 2018.

[18] Jan-Peter Schulz and Gerd Vogel. Improving the processes in the land surface scheme TERRA: Bare soil evaporation and skin temperature. *Atmosphere*, 11(5):513, 2020.

[19] Sun B. Pettey M. Seidel, D. J. and A. Reale. Global radiosonde balloon drift statistics. *Geophys. Res.*, 116(D07102), 2011.

[20] U. Voggenberger. Reconstruction of missing balloon drift data. *Possible Journals: ESSD, JAMES, GMD*, to be released in 2023.

[21] Joerge Winterfeldt, Beate Geyer, and Ralf Weisse. Using quikscat in the added value assessment of dynamically downscaled wind speed. *International Journal of Climatology*, 31(7):1028–1039, 2011.

[22] G. Y. Zou. Toward using confidence intervals to compare correlations. *Psychological Methods*, 12:399–413, 2007.

A Appendices

A.1 Configurations

A.1.1 Namelist parameters description

Table 8: Description of changes in the model setup of the conducted experiments.

Namelist Parameter	Description
scalar_advec	Type of advection of scalar fields
itype_fastwaves	Type of fast waves solver for Runge-Kutta dynamics
l3D_div_damping	The artificial divergence damping either acts only on the u- and v-equation (.FALSE.) or in a fully isotropic 3D manner (.TRUE.)
ldyn_bbc	To choose a dynamical bottom boundary condition
itype_bbcw	Option for choosing bottom boundary condition for vertical wind
lhor_pgrad_Mahrer	The horizontal pressure gradient (i.e. in the u- and v- equations) is either calculated in the terrain-following system (.FALSE.) or by interpolation of p' to a horizontal plane
itype_outflow_qrsg	To choose the type of relaxation treatment for qr, qs, qg
itype_canopy	Type of vegetation-canopy parameterisation. In case of itype_canopy = 2, use skin layer formulation
cskinc	Skin conductivity [10.0, 1000.0]. cskinc < 0: Usage of an external parameter field SKC from EXTPAR
itype_evsl	Parameter to select the type of parameterisation for evaporation from bare soil
c_soil	Surface area density of the (evaporative) soil surface
cwimax_ml	Factor for calculation of maximum interception water
itype_hydmod	Type of soil water transport and ground water runoff
ltkesso	Switch to calculate SSO-wake turbulence production for TKE
ltkeshs	Switch to consider horizontal shear production for TKE
icldm_turb	Mode of cloud representation to take into account sub-grid scale condensation within the turbulence parameterization in case of itype_turb = 3
icldm_tran	Mode of cloud representation to take into account sub-grid scale condensation within the new surface layer parameterization itype_tran = 2
loldtur	To choose ICON or COSMO turbulence scheme
itype_vdif	Parameter to choose the type of vertical diffusion calculation
lsuper_coolw	Switch to activate effects of supercooled liquid water in the microphysics (only activated for graupel and cloudice scheme)
itype_lbc_qrsg	To choose the type of lateral boundary treatment for qr, qs, qg, i.e., which values are used at the boundary zone
lana_qi	Switch to use the cloud-ice field contained in the initial data file as initial condition for cloud-ice
lana_qr_qs	Switch to use the values for rain and snow contained in the initial data file
llb_qi	Switch to take cloud-ice values contained in the lateral boundary data as boundary condition for cloud ice
llb_qr_qs	Switch to take rain and snow values contained in the lateral boundary data as boundary condition
itype_conv	Type of convection scheme
icapdcycl	CAPE: diurnal cycle correction
lconf_avg	Switch to apply a horizontal smoothing of the convective forcings (moisture convergence, surface moisture flux and vertical velocity) prior to calling the convection scheme

A.1.2 C2C100_YUSPECIF & C2C316_YUSPECIF

The files can be find in the community portal: C2C100_YUSPECIF C2C316_YUSPECIF

A.1.3 Differences in setups of C2C100, C2C201/301, and C2C316

Table 9: Namelist settings where differences exist between C2C100, C2C301, and C2C316. Missing entries in C2C100 are labeled with '-' and indicate that the corresponding namelist parameter did not exist in version COSMO5.0_clm9/16. It is possible that the setting was fixed in the code.

Namelist	Parameter	C2C100	C2C301	C2C316
DIACTL	ldursun_mch	–	False	False
	lhailcast	–	False	False
	ninchail	–	30	30
DYNCTL	crltau_inv	1 (crltau)	1	1
	divdamp_slope	20	1	1
	itype_bbc_w	1	1	114
	itype_fast_waves	1	1	2
	l_3d_div_damping	–	False	True
	l_diff_cold_pools	–	True	True
	l_diff_cold_pools_uv	–	False	False
	l_euler_dynamics	–	True	True
	l_satad_dyn_iter	–	True	True
	lcpp_dycore	–	False	False
	ldyn_bbc	True	True	False
	leulag	–	False	False
	lhor_pgrad_mahrer	–	False	False
	thresh_cold_pool	–	10	10
	y_scalar_advect	BOTT2	BOTT2	BOTTDC2
LMGRID	delta_t	–	75	75
	dt0lp	–	42	42
	h_scal	–	10000	10000
	irefatm	–	2	2
	p0sl	–	100000	100000
	t0sl	–	288.15	288.15
PHYCTL	cskinc	–	30	-1
	cskinc_urb	–	1000	1000
	curb_ahf	–	-1	-1
	curb_alb_so	–	0.1	0.1
	curb_alb_th	–	0.14	0.14
	curb_fr_bld	–	0.67	0.67
	curb_h2w	–	1.5	1.5
	curb_h_bld	–	15	15
	curb_hcap	–	1250000	1250000
	curb_hcon	–	0.777	0.777
	cwimax_ml	–	1.00E-06	1.00E-06
	hincrad	1	0.25	0.25

	icapdcycl	–	0	2
	icpl_aero_conv	–	0	0
	idiag_snowfrac	–	1	1
	itype_ahf	–	1	1
	itype_canopy	–	1	2
	itype_conv	0	0	2
	itype_eisa	–	2	2
	itype_evsl	3	3	4
	itype_hydmod	–	0	0
	itype_kbmo_uf	–	1	1
	itype_mire	–	0	0
	itype_sher	1	0	0
	itype_snow	–	1	1
	itype_snow_start	–	2	2
	itype_vdif	–	-1	-1
	l3dturb_metr	True	False	False
	lconf_avg	True	True	False
	ldetrain_conv_prec	–	True	True
	lgsp_first	–	False	False
	llake	False	True	True
	loldtur	– (true)	True	True
	lsflcnd	–	True	True
	lshallowconv_only	–	False	False
	lsoil_init_fill	–	False	False
	lsuper_coolw	–	False	False
	lterra_urb	– (false)	False	False
	ltkeshs	–	False	False
	lurbfab	–	True	True
	nincrad	40	10	10
	y_conv_closure	–	standard	standard
RUNCT	asynio_block_size	–	10	10
	asynio_host_mem	–	1	1
	asynio_prefetch_mem	–	8	8
	itype_iau	–	0	0
	itype_pert	–	0	0
	l_t_check	–	False	False
	ldebug_sso	–	False	False
	lsppt	–	False	False
	ltraj	–	False	False
	luse_radarfwo	–	False	False
	nblock	–	-1	-1
	ncomm_type	1	3	3
	nprocio_radar	–	0	0
	nprocy	30	32	32
	nproma	–	16	16
	peri_iau	–	3600	3600
	rperturb	–	0	0
TUNING	c_soil	1	1	1.25
	cmfctop	–	0.33	0.33
	cprcon_dc	–	0.0002	0.0002
	fac_rootdp2	0.9	1	1

l_g	–	2.5	2.5
radqc_fact	–	0.5	0.5
radqg_fact	–	0.5	0.5
radqi_fact	–	0.5	0.5
radqs_fact	–	0.5	0.5
rat_sea	20	9	9
thick_dc	–	20000	20000
tmpmin_dc	–	250.16	250.16
tune_capdcfac_et	–	0.5	0.5
tune_capdcfac_tr	–	0.5	0.5
tune_entrorg	–	0.0019	0.0019
tune_minsnowfrac	–	0.125	0.125
tune_qexc	–	0.0125	0.0125
tune_rcucov	–	0.05	0.05
tune_rcucov_trop	–	0.03	0.03
tune_rdepths	–	20000	20000
tune_rhebc_land	–	0.75	0.75
tune_rhebc_land_trop	–	0.7	0.7
tune_rhebc_ocean	–	0.85	0.85
tune_rhebc_ocean_trop	–	0.76	0.76
tune_rprcon	–	0.0014	0.0014
tune_texc	–	0.125	0.125

A.2 Sensitivity Analysis of ScoPi$_{simulation}$

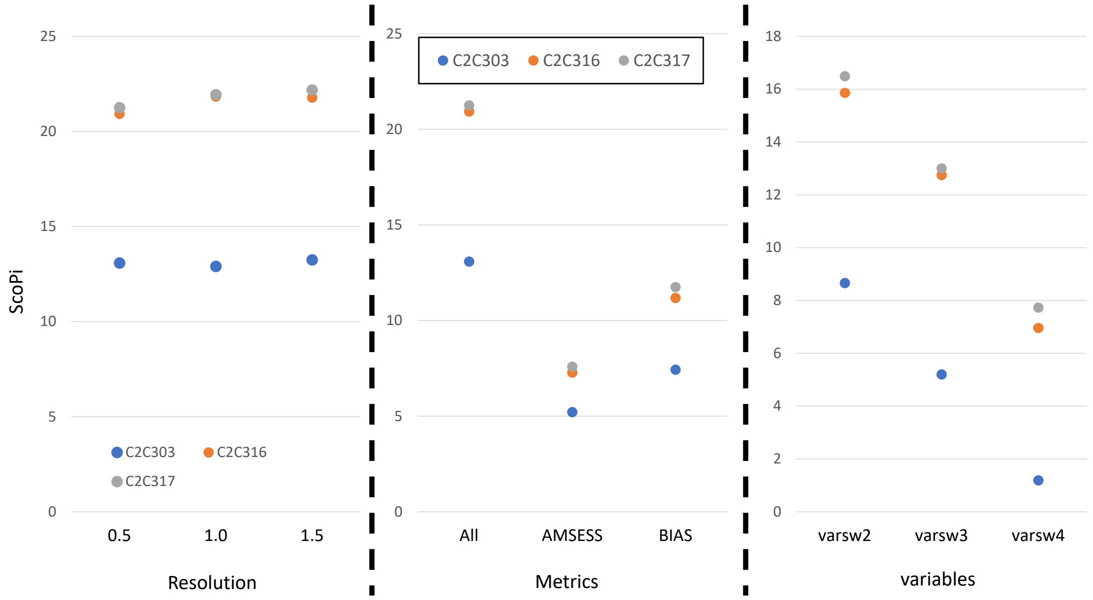


Figure 16: Sensitivity of the ScoPi$_{simulation}$ with respect to the resolution (left), the metrics (center), and the atmospheric parameters (right). The analysis is done for three model configurations from phase II (coloured points) tested against the reference simulation C2C301 (Tab. 2). 'varsw2', 'varsw3' and 'varsw4' indicate modified weights for the different variables when calculating the ScoPi$_{region}$. The ScoPi$_{simulation}$ is calculated by using both weights1 and weights2 from Tab. 5, and then averaging the resulting two scores.

A.3 Seasonal 2D plots for 2m temperature and precipitation

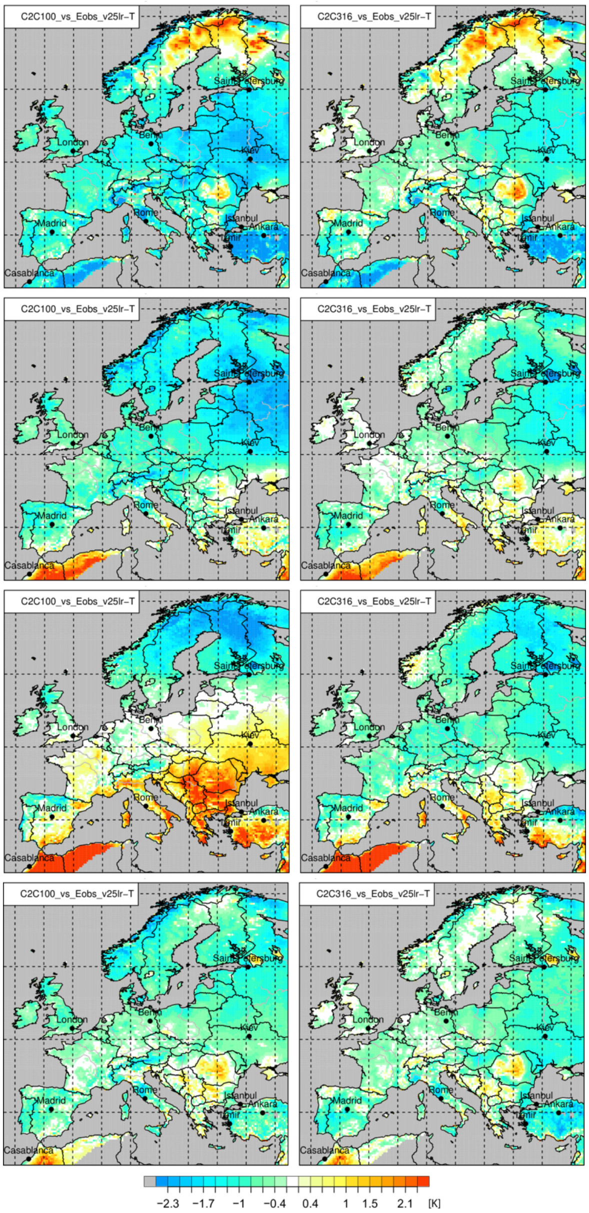


Figure 17: Comparison of the recommended version of 2018 (C2C100, *left*) and experiment C2C316 (*right*): seasonal mean bias of 2-meter temperatures compared to Eobs [K] (from top to bottom winter, spring, summer, and autumn).

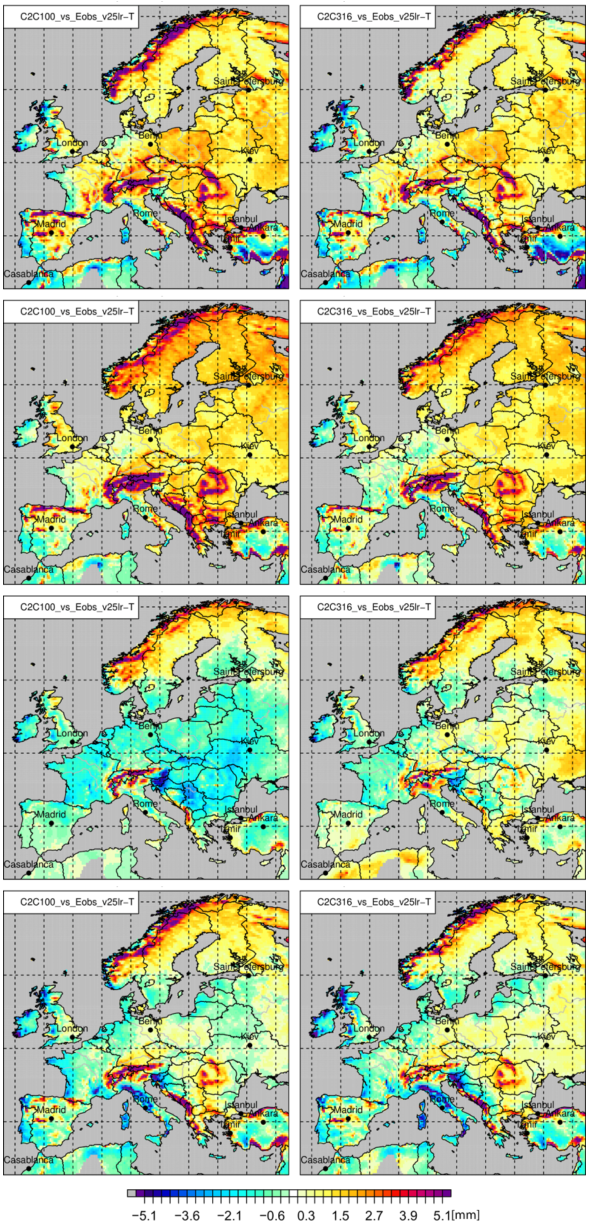


Figure 18: Comparison of the recommended version of 2018 (C2C100, *left*) and C2C316 (*right*): seasonal BIAS of precipitation compared to Eobs [mm/mon](from top to bottom winter, spring, summer, and autumn).

A.4 Preparatory work for the comparison of vertical model data to radiosondes observations

The simulation output for the variables temperature, relative and specific humidity, geopotential height, and wind speed on model levels were interpolated to the standard pressure levels 950, 850, 700, 500, 300, and 200 hPa. To reduce the amount of data after the first analysis steps, the number of levels was reduced and only the levels at 850, 700, 500, and 300 hPa were taken into account. As for the analysis of the single level variables, here the analysis is conducted on 3-daily means of the given variables.

A.5 Radiosonde Trajectory Reconstruction

Depending on the given conditions, as well as the time of the year, radiosondes can drift horizontally from the starting point [19]. This drift (hereafter displacement) can be as large as a few hundred kilometers. In the Northern Hemisphere mid-latitudes, where many radiosonde sites are located, large systematic displacements can be observed, especially in the winter months, or when zonal circulation regimes are more prevalent.

The approach to estimating the actual radiosonde balloon position is to calculate how it was transported with the wind. It is reminded that the 'measured wind' from balloon-borne instruments is not the result of an Eulerian measurement (like a LIDAR wind would provide), but the result of a quasi-Lagrangian measurement: the horizontal wind is estimated from the balloon trajectory. Balloon movement relative to the ambient air occurs only in the vertical direction. We solve the inverse problem and reconstruct the balloon trajectory using the wind information available.

Both estimates, from the trajectory to the wind, and vice-versa, assume that inertia can be neglected, i.e. the balloon has a very small mass, and yet a large surface area, so it is assumed to be advected with the wind, at least in the horizontal direction. Oscillations of the balloon around its center of gravity during ascent are supposed not to affect its horizontal speed, at least at the gridscale of a larger observed area, like in this case. However, high-frequency oscillations can interfere with high-frequency wind estimates in some cases [1].

If the balloon position relative to the launch position was not reported at the time of its ascent, it must be calculated from the available wind data. The calculated position is always given as latitude displacement and longitude displacement (decimal degrees). For each vertical level, these two values can be added to the station coordinates to obtain the new (latitude, longitude) position at the given level.

For the position calculation, we apply the same simple physical laws that have been used to derive the reported winds. Only a few initial parameters are necessary for this:

- Station coordinates or starting point of the sonde (latitude and longitude)
- Wind (u and v), measured by the sonde at different pressure levels
- Measurement time (t), at different pressure levels

These variables enable calculation of how long the sonde was exposed to which wind, and therefore can be used to estimate the displacement of the sonde. Especially older datasets often only contain the starting time of the ascent, but temporal information is not available for all the reported pressure levels. To estimate the time elapsed since the release of the balloon, two variables are needed:

- Reported pressure levels (radiosondes) or heights (PILOT balloons)
- Sonde ascent speed

The only information available regarding the radiosonde vertical position is often the pressure. However, pressure can be transformed to a height profile using the temperature profile, assuming a piecewise polytropic atmosphere.

$$\frac{\Delta T}{\Delta p} \Longrightarrow \frac{\Delta T}{\Delta z} \tag{6}$$

T	temperature	$[K]$
z	height	$[m]$
p	pressure	$[Pa]$

Subsequently, this information is used to determine the height of all pressure levels.

The determination of the sonde's ascent speed is more uncertain. It depends on some unknown variables such as the vertical wind speed, or the weight to buoyancy ratio of the probe and the balloon. Deviations in the filling level of the balloon, the air resistance of the balloon skin, as well as the ambient temperature or even that of the balloon gas, can lead to further small deviations.

The study of data with known altitude time series has shown that the rate of ascent varies substantially. For most ascents, the vertical speed varies around $5.0\frac{m}{s}$. This observation is also consistent with other sources [19].

As a first step in the process it is necessary to calculate the height profile from temperature and pressure information by applying the dry polytropic height formula. The height profile is then used to calculate the time interval spent by the sonde between the noted levels. These time intervals are then used to determine the transport of the balloon according to the mean wind inside the layer between the levels $i \rightarrow i+1$.

$$s_{i+1-longitude} = u_{i\rightarrow i+1} * \frac{z_{i\rightarrow i+1}}{w_{balloon}} \tag{7}$$

$$s_{i+1-latitude} = v_{i\rightarrow i+1} * \frac{z_{i\rightarrow i+1}}{w_{balloon}} \tag{8}$$

s_{i+1}	displacement distance at level $i+1$ (longitude and latitude)	$[m]$
u	mean eastward wind component of layer $i \rightarrow i+1$	$[\frac{m}{s}]$
v	mean northward wind component of layer $i \rightarrow i+1$	$[\frac{m}{s}]$
z	height of layer $i \rightarrow i+1$	$[m]$
$w_{balloon}$	assumed vertical speed of balloon	$[\frac{m}{s}]$

Afterwards, this distance is converted into latitude and longitude using the inverse Haversine method on an assumed sphere. Those are the final displacements, which then can be used to determine the more precise position of the observation. This process will be described in detail in the 2023 paper Voggenberber et al. [20]

A.6 Additional results of upper air data analysis

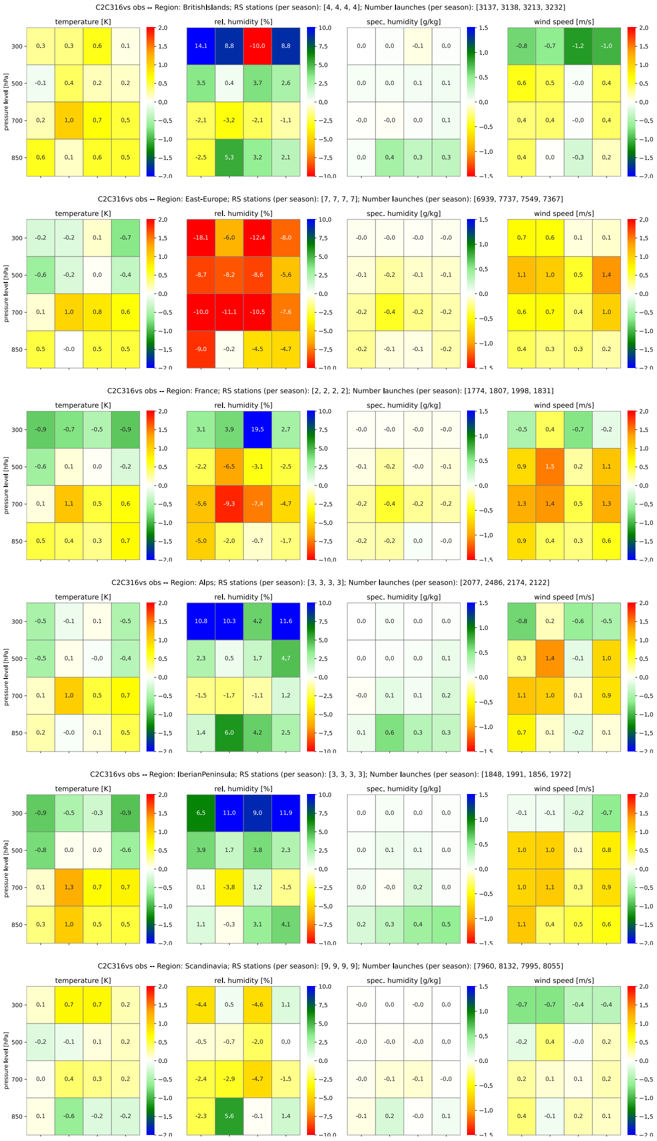


Figure 19: as Fig. 9: Comparison of the simulation C2C316 to radiosonde observations: mean seasonal BIAS calculated over the period 1981-1985 for the variables, from left to right, temperature, relative humidity, specific humidity, and wind speed, for all considered PRUDENCE region beside Mediterranean and mid-Europe.

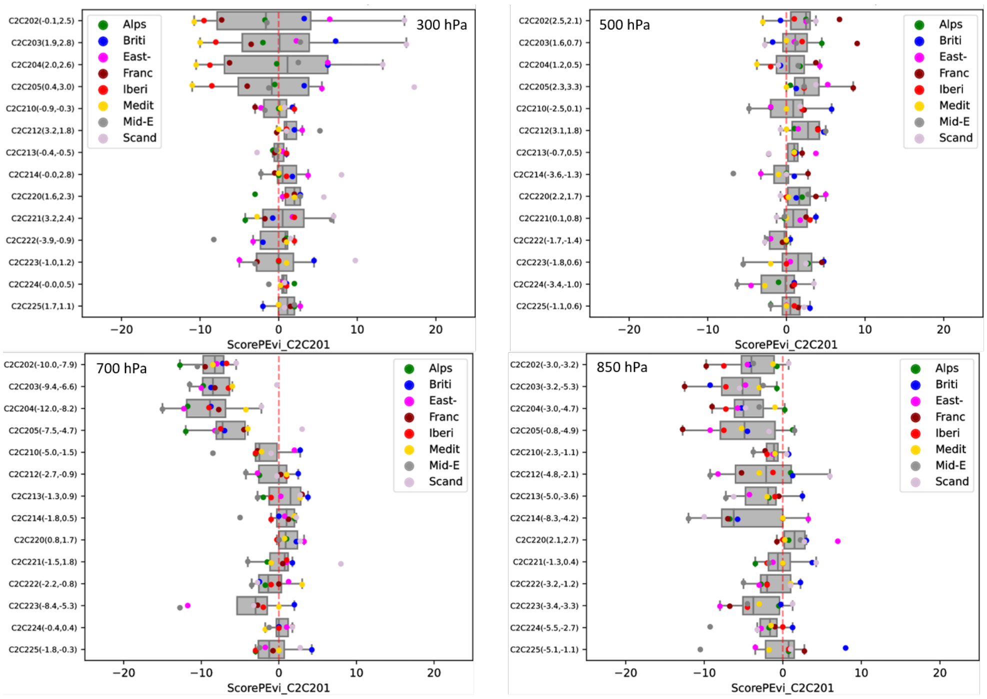


Figure 20: ScoPi-plot for the simulations of Phase Ic with respect to C2C201 for the levels 300, 500, 700, and 850 hPa for the period of 1981-1985.

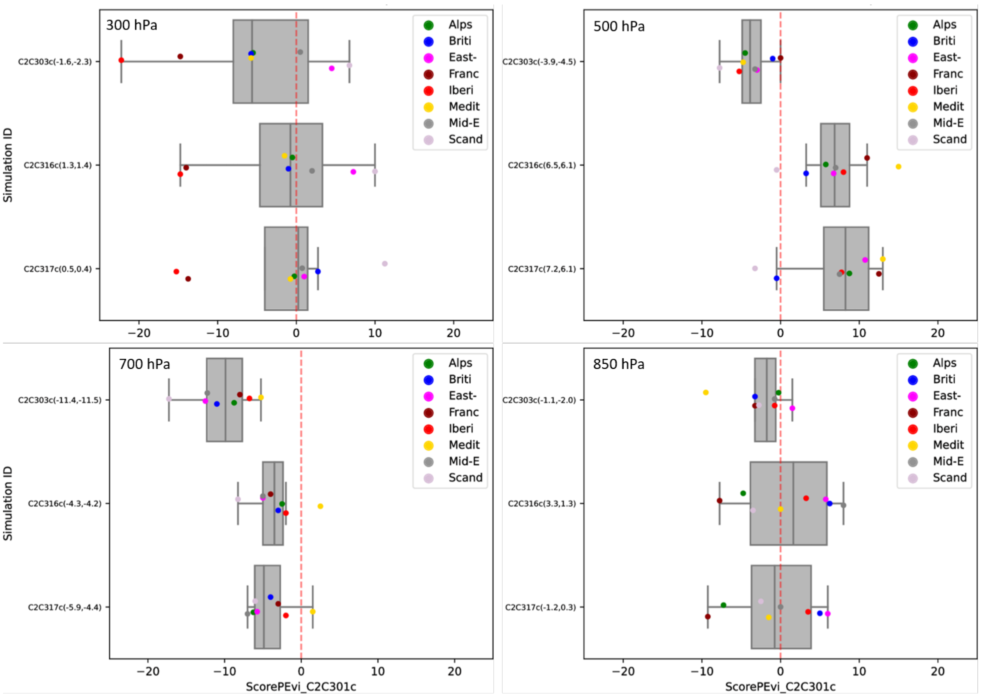


Figure 21: ScoPi-plot for the simulations of Phase IIIc with respect to C2C301 for the levels 300, 500, 700, and 850 hPa for the period of 2002-2008.

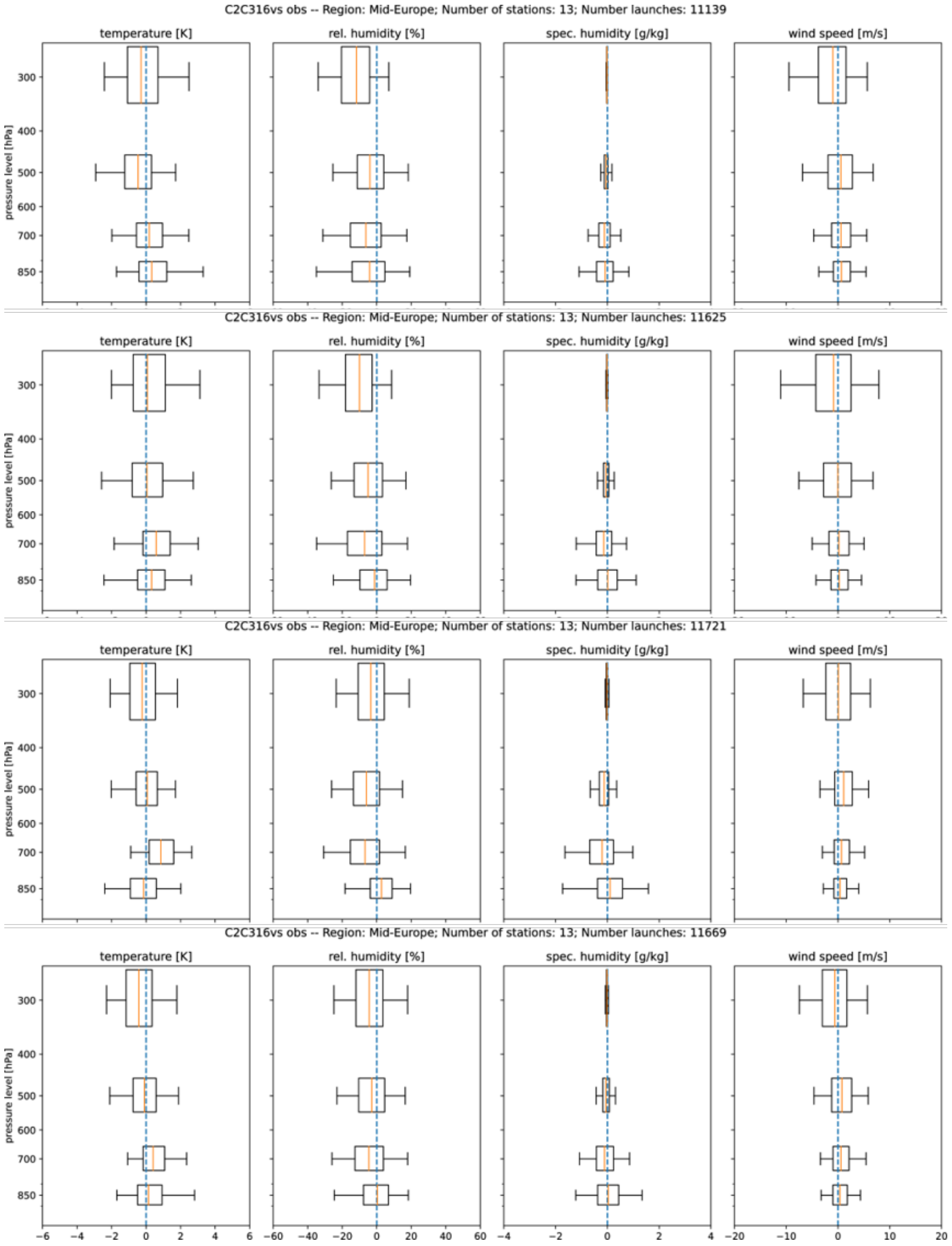


Figure 22: Boxplot for deviations of the simulation C2C316 from radiosondes data for the levels 300, 500, 700, and 850 hPa for the period of 1981-1985 for Mid-Europe. The number of observations per season is given in the headers of the plots.

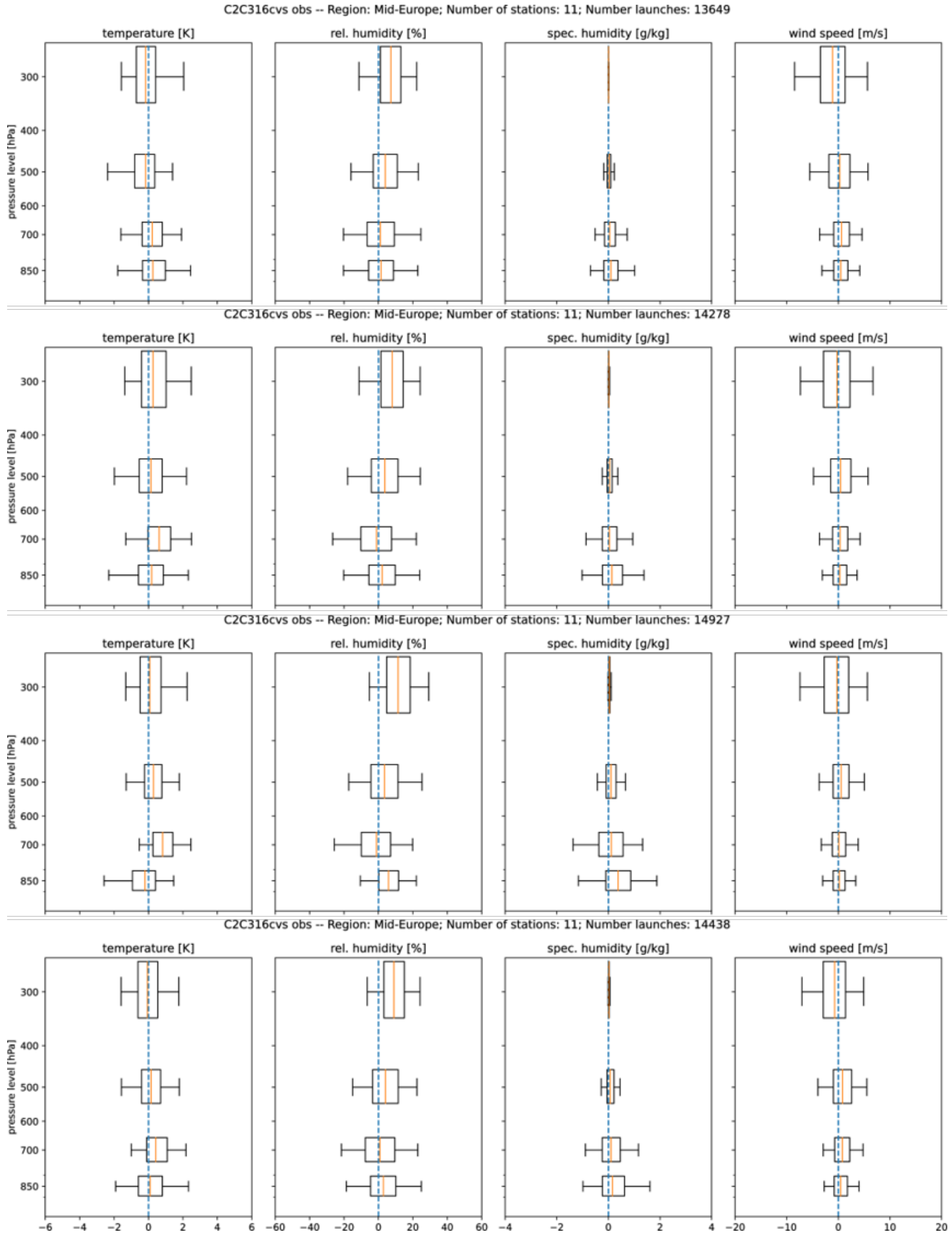


Figure 23: as Fig. 22 but for the period of 2002-2008 for Mid-Europe.

B Acknowledgments

We acknowledge the DKRZ for the use of resources in terms of granted computing time and storage capacity (project bg1155). Additionally, we used data from the DKRZ /pool/data section provided by the CLM Community. At the same time we express our gratitude towards Daniel Rieger from the Deutscher Wetterdienst who has reviewed the document prior to its publication.

List of COSMO Newsletters and Technical Reports

(available for download from the COSMO Website: www.cosmo-model.org)

COSMO Newsletters

No. 1: February 2001.

No. 2: February 2002.

No. 3: February 2003.

No. 4: February 2004.

No. 5: April 2005.

No. 6: July 2006.

No. 7: April 2008; Proceedings from the 8th COSMO General Meeting in Bucharest, 2006.

No. 8: September 2008; Proceedings from the 9th COSMO General Meeting in Athens, 2007.

No. 9: December 2008.

No. 10: March 2010.

No. 11: April 2011.

No. 12: April 2012.

No. 13: April 2013.

No. 15: July 2015.

No. 16: July 2016.

No. 17: July 2017.

No. 18: November 2018.

No. 19: October 2019.

No. 20: December 2020.

No. 21: May 2022.

No. 22: May 2023.

COSMO Technical Reports

No. 1: Dmitrii Mironov and Matthias Raschendorfer (2001):
Evaluation of Empirical Parameters of the New LM Surface-Layer Parameterization Scheme. Results from Numerical Experiments Including the Soil Moisture Analysis.

No. 2: Reinhold Schrodin and Erdmann Heise (2001):
The Multi-Layer Version of the DWD Soil Model TERRA_LM.

No. 3: Günther Doms (2001):
A Scheme for Monotonic Numerical Diffusion in the LM.

No. 4: Hans-Joachim Herzog, Ursula Schubert, Gerd Vogel, Adelheid Fiedler and Roswitha Kirchner (2002):
LLM — the High-Resolving Nonhydrostatic Simulation Model in the DWD-Project LITFASS.
Part I: Modelling Technique and Simulation Method.

No. 5: Jean-Marie Bettems (2002):
EUCOS Impact Study Using the Limited-Area Non-Hydrostatic NWP Model in Operational Use at MeteoSwiss.

No. 6: Heinz-Werner Bitzer and Jürgen Steppeler (2004):
Documentation of the Z-Coordinate Dynamical Core of LM.

No. 7: Hans-Joachim Herzog, Almut Gassmann (2005):
Lorenz- and Charney-Phillips vertical grid experimentation using a compressible non-hydrostatic toy-model relevant to the fast-mode part of the 'Lokal-Modell'.

No. 8: Chiara Marsigli, Andrea Montani, Tiziana Paccagnella, Davide Sacchetti, André Walser, Marco Arpagaus, Thomas Schumann (2005):
Evaluation of the Performance of the COSMO-LEPS System.

No. 9: Erdmann Heise, Bodo Ritter, Reinhold Schrodin (2006):
Operational Implementation of the Multilayer Soil Model.

No. 10: M.D. Tsyrulnikov (2007):
Is the particle filtering approach appropriate for meso-scale data assimilation ?

No. 11: Dmitrii V. Mironov (2008):
Parameterization of Lakes in Numerical Weather Prediction. Description of a Lake Model.

No. 12: Adriano Raspanti (2009):
COSMO Priority Project "VERification System Unified Survey" (VERSUS): Final Report.

No. 13: Chiara Marsigli (2009):
COSMO Priority Project "Short Range Ensemble Prediction System" (SREPS): Final Report.

No. 14: Michael Baldauf (2009):
COSMO Priority Project "Further Developments of the Runge-Kutta Time Integration Scheme" (RK): Final Report.

No. 15: Silke Dierer (2009):
COSMO Priority Project "Tackle deficiencies in quantitative precipitation forecast" (QPF): Final Report.

No. 16: Pierre Eckert (2009):
COSMO Priority Project "INTERP": Final Report.

No. 17: D. Leuenberger, M. Stoll and A. Roches (2010):
Description of some convective indices implemented in the COSMO model.

No. 18: Daniel Leuenberger (2010):
Statistical analysis of high-resolution COSMO Ensemble forecasts in view of Data Assimilation.

No. 19: A. Montani, D. Cesari, C. Marsigli, T. Paccagnella (2010):
Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO–LEPS system: main achievements and open challenges.

No. 20: A. Roches, O. Fuhrer (2012):
Tracer module in the COSMO model.

No. 21: Michael Baldauf (2013):
A new fast-waves solver for the Runge-Kutta dynamical core.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_21

No. 22: C. Marsigli, T. Diomede, A. Montani, T. Paccagnella, P. Louka, F. Gofa, A. Corigliano (2013):
The CONSENS Priority Project.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_22

No. 23: M. Baldauf, O. Fuhrer, M. J. Kurowski, G. de Morsier, M. Müllner, Z. P. Piotrowski, B. Rosa, P. L. Vitagliano, D. Wójcik, M. Ziemiański (2013):
The COSMO Priority Project 'Conservative Dynamical Core' Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_23

No. 24: A. K. Miltenberger, A. Roches, S. Pfahl, H. Wernli (2014):
Online Trajectory Module in COSMO: a short user guide.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_24

No. 25: P. Khain, I. Carmona, A. Voudouri, E. Avgoustoglou, J.-M. Bettems, F. Grazzini (2015):
The Proof of the Parameters Calibration Method: CALMO Progress Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_25

No. 26: D. Mironov, E. Machulskaya, B. Szintai, M. Raschendorfer, V. Perov, M. Chumakov, E. Avgoustoglou (2015):
The COSMO Priority Project 'UTCS' Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_26

No. 27: J-M. Bettems (2015):
The COSMO Priority Project 'COLOBOC': Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_27

No. 28: Ulrich Blahak (2016):
RADAR_MIE_LM and RADAR_MIELIB - Calculation of Radar Reflectivity from Model Output.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_28

No. 29: M. Tsyrulnikov and D. Gayfulin (2016):
A Stochastic Pattern Generator for ensemble applications.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_29

No. 30: D. Mironov and E. Machulskaya (2017):
A Turbulence Kinetic Energy – Scalar Variance Turbulence Parameterization Scheme.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_30

No. 31: P. Khain, I. Carmona, A. Voudouri, E. Avgoustoglou, J.-M. Bettems, F. Grazzini, P. Kaufmann (2017):
CALMO - Progress Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_31

No. 32: A. Voudouri, P. Khain, I. Carmona, E. Avgoustoglou, J.M. Bettems, F. Grazzini, O. Bellprat, P. Kaufmann and E. Bucchignani (2017):
Calibration of COSMO Model, Priority Project CALMO Final report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_32

No. 33: N. Vela (2017):
VAST 2.0 - User Manual.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_33

No. 34: C. Marsigli, D. Alferov, M. Arpagaus, E. Astakhova, R. Bonanno, G. Duniec, C. Gebhardt, W. Interewicz, N. Loglisci, A. Mazur, V. Maurer, A. Montani, A. Walser (2018):
COsmo Towards Ensembles at the Km-scale IN Our countries (COTEKINO), Priority Project final report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_34

No. 35: G. Rivin, I. Rozinkina, E. Astakhova, A. Montani, D. Alferov, M. Arpagaus, D. Blinov, A. Bundel, M. Chumakov, P. Eckert, A. Euripides, J. Förstner, J. Helmert, E. Kazakova, A. Kirsanov, V. Kopeikin, E. Kukanova, D. Majewski, C. Marsigli, G. de Morsier, A. Muravev, T. Paccagnella, U. Schättler, C. Schraff, M. Shatunova, A. Shcherbakov, P. Steiner, M. Zaichenko (2018):
The COSMO Priority Project CORSO Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_35

No. 36: A. Raspanti, A. Celozzi, A. Troisi, A. Vocino, R. Bove, F. Batignani (2018):
The COSMO Priority Project VERSUS2 Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_36

No. 37: A. Bundel, F. Gofa, D. Alferov, E. Astakhova, P. Baumann, D. Boucouvala, U. Damrath, P. Eckert, A. Kirsanov, X. Lapillonne, J. Linkowska, C. Marsigli, A. Montani, A. Muraviev, E. Oberto, M.S. Tesini, N. Vela, A. Wyszogrodzki, M. Zaichenko, A. Walser (2019):
The COSMO Priority Project INSPECT Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_37

No. 38: G. Rivin, I. Rozinkina, E. Astakhova, A. Montani, J-M. Bettems, D. Alferov, D. Blinov, P. Eckert, A. Euripides, J. Helmert, M.Shatunova (2019):
The COSMO Priority Project CORSO-A Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_38

No. 39: C. Marsigli, D. Alferov, E. Astakhova, G. Duniec, D. Gayfulin, C. Gebhardt, W. Interewicz, N. Loglisci, F. Marcucci, A. Mazur, A. Montani, M. Tsyrulnikov, A. Walser (2019):
Studying perturbations for the representation of modeling uncertainties in Ensemble development (SPRED Priority Project): Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_39

No. 40: E. Bucchignani, P. Mercogliano, V. Garbero, M. Milelli, M. Varentsov, I. Rozinkina, G. Rivin, D. Blinov, A. Kirsanov, H. Wouters, J.-P. Schulz, U. Schättler (2019):

Analysis and Evaluation of TERRA_URB Scheme: PT AEVUS Final Report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_40

No. 41: X. Lapillonne, O. Fuhrer (2020):
Performance On Massively Parallel Architectures (POMPA): Final report.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_41

No. 42: E. Avgoustoglou, A. Voudouri, I Carmona, E. Bucchignani, Y. Levy, J. -M. Bettems (2020):
A methodology towards the hierarchy of COSMO parameter calibration tests via the domain sensitivity over the Mediterranean area.
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_42

No. 43: H. Muskatel, U. Blahak, P. Khain, A. Shtivelman, M. Raschendorfer, M. Kohler, D. Rieger, O. Fuhrer, X. Lapillonne, G. Rivin, N. Chubarova, M. Shatunova, A. Poliukhov, A. Kirsanov, T. Andreadis, S. Gruber (2021):
The COSMO Priority Project $T^2(RC)^2$: Testing and Tuning of Revised Cloud Radiation Coupling, Final Report
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_43

No. 44: M. Baldauf, D. Wojcik, F. Prill, D. Reinert, R. Dumitrache, A. Iriza, G. deMorsier, M. Shatunova, G. Zaengl, U. Schaettler (2021):
The COSMO Priority Project CDIC: Comparison of the dynamical cores of ICON and COSMO, Final Report
DOI: 10.5676/DWD_pub/nwv/cosmo-tr_44

No. 45 Marsigli C., Astakhova E. Duniec G., Fuezer L., Gayfulin D., Gebhardt C., Golino R., Heppelmann T., Interewicz W., Marcucci F., Mazur A., Sprengel M., Tsyrulnikov M., Walser A. (2022):
The COSMO Priority Project APSU: Final Report.

No. 46 A. Iriza-Burca, F. Gofa, D. Boucouvala, T. Andreadis, J. Linkowska, P. Khain, A. Shtivelman, F. Batignani, A. Pauling, A. Kirsanov, T. Gastaldo, B. Maco, M. Bogdan, F. Fundel (2022):
The COSMO Priority Project CARMA: Common Area with Rfdbk/MEC Application Final Report.

No. 47 A. Voudouri, E. Avgoustoglou, Y. Levy, I. Carmona, E. Bucchignani, J. M. Bettems (2022):
Calibration of COSMO Model, Priority Project CALMO-MAX: Final Report.

No. 48 D. Rieger et al. (2022):
The Priority Project C2I, Transition of COSMO to ICON - Final Report.

No. 49 E. Churiulin, M. Toelle, V. Kopeikin, M. Uebel, J. Helmert and J.-M. Bettems (2022):
The COSMO Priority Task VAINT: Vegetation Atmosphere INTeractions Report.

No. 50 F. Gofa, A. Bundel, M. S. Tesini, C. Marsigli, M. Hoff, D. Boucouvala, A. Mazur, J. Linkowska, G. Duniec, D. Cattani, B. Pasquier, A. Muraviev, E. Tatarinovich, Y. Khlestova, D. Zakharchenko (2023):
The COSMO Priority Project AWARE: Appraisal of "Challenging WeAther" FoREcasts Final Report.

COSMO Technical Reports

Issues of the COSMO Technical Reports series are published by the *COnsortium for Small-scale MOdelling* at non-regular intervals. COSMO is a European group for numerical weather prediction with participating meteorological services from Germany (DWD, AWGeophys), Greece (HNMS), Italy (USAM, ARPA-SIMC, ARPA Piemonte), Switzerland (MeteoSwiss), Poland (IMGW), Romania (NMA) and Russia (RHM). The general goal is to develop, improve and maintain a non-hydrostatic limited area modelling system to be used for both operational and research applications by the members of COSMO. This system is initially based on the COSMO-Model (previously known as LM) of DWD with its corresponding data assimilation system.

The Technical Reports are intended

- for scientific contributions and a documentation of research activities,
- to present and discuss results obtained from the model system,
- to present and discuss verification results and interpretation methods,
- for a documentation of technical changes to the model system,
- to give an overview of new components of the model system.

The purpose of these reports is to communicate results, changes and progress related to the LM model system relatively fast within the COSMO consortium, and also to inform other NWP groups on our current research activities. In this way the discussion on a specific topic can be stimulated at an early stage. In order to publish a report very soon after the completion of the manuscript, we have decided to omit a thorough reviewing procedure and only a rough check is done by the editors and a third reviewer. We apologize for typographical and other errors or inconsistencies which may still be present.

At present, the Technical Reports are available for download from the COSMO web site (www.cosmo-model.org). If required, the member meteorological centres can produce hard-copies by their own for distribution within their service. All members of the consortium will be informed about new issues by email.

For any comments and questions, please contact the editor:

Massimo Milelli

massimo.milelli@cimafoundation.org