The Fifth Generation Regional Climate Modeling System, RegCM5: the first CP European wide simulation and validation over the CORDEX-CORE domains.

Erika Coppola\textsuperscript{1}, Filippo Giorgi\textsuperscript{2}, Graziano Giuliani\textsuperscript{3}, Emanuela Pichelli\textsuperscript{4}, James M Ciarlo\textsuperscript{5}, Francesca Raffaele\textsuperscript{6}, Rita Nogherotto\textsuperscript{5}, Michelle Simões Reboita\textsuperscript{7}, Chen Lu\textsuperscript{5}, Natalia Zazulie\textsuperscript{8}, Luiza Vargas-Heinz\textsuperscript{5}, Andressa Andrade Cardoso\textsuperscript{9}, and Johannes de Leeuw\textsuperscript{10}

\textsuperscript{1}Abdus Salam ICTP
\textsuperscript{2}The Abdus Salam International Centre for Theoretical Physics
\textsuperscript{3}ICTP
\textsuperscript{4}Abdus Salam International Center for Theoretical Physics
\textsuperscript{5}Abdus Salam International Centre for Theoretical Physics
\textsuperscript{6}The Abdus Salam International Centre for Theoretical Physics - ICTP
\textsuperscript{7}Universidade Federal de Itajubá
\textsuperscript{8}Universidad de Buenos Aires
\textsuperscript{9}Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo
\textsuperscript{10}National Institute of Oceanography and Applied Geophysics

January 16, 2024

Abstract

The Regional Climate Modeling system (RegCM) has undergone a significant evolution over the years, leading for example to the widely used versions RegCM4 and RegCM4-NH. In response to the demand for higher resolution, a new version of the system has been developed, RegCM5, incorporating the non-hydrostatic dynamical core of the MOLOCH weather prediction model. In this paper we assess the RegCM5’s performance for 5 CORDEX-CORE domains, including a pan-European domain at convection-permitting resolution.

We find temperature biases generally in the range of -2 to 2 degrees Celsius, higher in the northernmost regions of North America and Asia during winter, linked to cloud water overestimation. Central Asia and the Tibetan Plateau show cold biases, possibly due to sparse station coverage. The model exhibits a prevailing cold bias in maximum temperature and warm bias in minimum temperature, associated with a systematic overestimation of lower-level cloud fraction, especially in winter.

Taylor diagrams indicate a high spatial temperature pattern correlation with ERA5 and CRU data, except in South America and the Caribbean region. The precipitation evaluation shows an overestimation in South America, East Asia, and Africa. RegCM5 improves the daily precipitation distribution compared to RegCM4, particularly at high intensities. The analysis of wind fields confirms the model’s ability to simulate monsoon circulations. The assessment of tropical cyclone tracks highlights a strong sensitivity to the tracking algorithms, thus necessitating a careful model interpretation.

Over the European region, the convection permitting simulations especially improve the diurnal cycle of precipitation and the hourly precipitation intensities.

Hosted file

983024_0_art_file_11721499_s643zt.docx available at https://authorea.com/users/713227/articles/697591-the-fifth-generation-regional-climate-modeling-system-regcm5-the-first-
The Fifth Generation Regional Climate Modeling System, RegCM5: the first CP European wide simulation and validation over the CORDEX-CORE domains.

Coppola Erika¹, Filippo Giorgi², Graziano Giuliani¹, Emanuela Pichelli¹, James M. Ciarlo¹, Francesca Raffaele¹, Rita Nogherotto¹, Michelle Simões Reboita¹,², Chen Lu¹, Natalia Zazulie¹,³, Luiza Vargas-Heinz¹,⁴, Andressa Andrade Cardoso¹,⁶, Johannes de Leeuw¹,⁷

¹The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy
²Federal University of Itajubá, Itajubá-MG, Brazil
³National Council of Scientific and Technical Research (CONICET), Buenos Aires, Argentina
⁴Università degli Studi di Trieste, Trieste, Italy
⁵University of Malta, Malta
⁶Departamento de Ciências Atmosféricas, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão 1226, Cidade Universitária, São Paulo, SP, Brazil.
⁷The National Institute of Oceanography and Applied Geophysics, Trieste, Italy.

Abstract

The Regional Climate Modeling system (RegCM) has undergone a significant evolution over the years, leading for example to the widely used versions RegCM4 and RegCM4-NH. In response to the demand for higher resolution, a new version of the system has been developed, RegCM5, incorporating the non-hydrostatic dynamical core of the MOLOCH weather prediction model. In this paper we assess the RegCM5’s performance for CORDEX-CORE domains, including a pan-European domain at convection-permitting resolution.

We find temperature biases generally in the range of -2 to 2 degrees Celsius, higher in the northernmost regions of North America and Asia during winter, linked to cloud water overestimation. Central Asia and the Tibetan Plateau show cold biases, possibly due to sparse station coverage. The model exhibits a prevailing cold bias in maximum temperature and warm bias in minimum temperature, associated with a systematic overestimation of lower-level cloud fraction, especially in winter.

Taylor diagrams indicate a high spatial temperature pattern correlation with ERA5 and CRU data, except in South America and the Caribbean region. The precipitation evaluation shows an overestimation in South America, East Asia, and Africa. RegCM5 improves the daily precipitation distribution compared to RegCM4, particularly at high intensities. The analysis of wind fields confirms the model’s ability to simulate monsoon circulations. The assessment of tropical cyclone tracks highlights a strong sensitivity to the tracking algorithms, thus necessitating a careful model interpretation.

Over the European region, the convection permitting simulations especially improve the diurnal cycle of precipitation and the hourly precipitation intensities.
Introduction

Since the initial work of Dickinson et al. (1989) and Giorgi and Bates (1989) introducing the first version of the Regional Climate Modeling system (RegCM1), the dynamical downscaling technique using limited-area Regional Climate Models (RCMs) has become a well-known method used worldwide (Giorgi 2009). The RCM community has witnessed the evolution of various RCM systems, including subsequent model versions of the RegCM framework: RegCM2, RegCM2.5, RegCM3, and the latest RegCM4 (Giorgi et al., 1993a, b; Giorgi and Mearns, 1999; Pal et al., 2007; Giorgi et al., 2012). These model developments largely stemmed from the incorporation of new and more advanced physics packages, with the exception of the RegCM1 to RegCM2 transition, which brought an update to the model's dynamical core, adopting the MM5's hydrostatic dynamical representation (Grell et al., 1994).

RegCM4, in particular, has emerged as a cornerstone in the field, finding extensive use in a diverse range of projects and applications, from process studies to paleo and future climate projections. This includes participation in the Coordinated Regional Downscaling Experiment (CORDEX, Giorgi et al., 2009; Gutowski et al., 2016). RegCM4 is designed to be coupled with ocean, land, chemistry, and aerosol modules in a fully interactive way, adding to its versatility (Sitz et al., 2017).

However, as the demand for higher resolutions escalates, with the RCM community increasingly reaching "convection-permitting" resolutions of a few kilometers, RegCM4's hydrostatic dynamical core has been recognized as a limiting factor for such applications. As a result, the RegCM4 dynamical core underwent a significant upgrade, including the MM5 non-hydrostatic dynamics and leading to the development of RegCM4-NH (Coppola et al., 2021). RegCM4-NH has already extensively been used for climate simulations at convection-permitting scales, e.g. within the European Climate Prediction System (EUCP) project and the CORDEX Flagship Pilot Study dedicated to convection (CORDEX-FPSCONV) (Coppola et al. 2020). Its potential has been demonstrated through multi-model experiments, including those carried out over the greater Alpine region by Ban et al. (2021) and Pichelli et al. (2021), over the South America region of La Plata basin (Betolli et al., 2021; da Rocha et al., 2023) and the region of Lake Victoria in Africa (Lipzig et al., 2023; Glazer et al. 2023).

One of the major drawbacks of the RegCM4-NH is the computational cost to run the model, since the MM5 dynamical core is still based on a split explicit scheme requiring short time steps for stability constraints. In addition, the MM5 scheme includes a relatively high diffusion term, also to increase stability. For this reason, a new version of the RegCM modeling system, RegCM5 was developed by incorporating the dynamical core of the non-hydrostatic weather prediction model MOLOCH (Buzzì et al., 2014; Malguzzi et al., 2006; Trini Castelli et al., 2020) as part of a collaborative effort between the ICTP RegCM modeling team and the Institute of Atmospheric Sciences and Climate (ISAC) of the National Research Council (CNR) of Italy. The first version RegCM5 was introduced by Giorgi et al. (2023), who tested it at convection parametrized and convection permitting resolutions over the Euro-CORDEX domain and the CORDEX FPS convection Alpine domain. In these experiments, not only the model was 4-5 times more computationally efficient than the old RegCM4 and RegCM4-NH counterparts, but also improved different aspects of model performance, and in particular the occurrence of extreme precipitation events and some systematic temperature biases (Giorgi et al. 2023).
RegCM5 thus represents an important step forwards for model users, in particular when using
the model at very high resolutions. It is important to acknowledge that the success of the
RegCM system is not only the work of the core development teams, but also a result of
contributions from the broader user community, who play a vital role in testing the model,
identifying errors, customizing model configurations, and implementing new components. As
RegCM5 has become available for public use, ongoing feedback and optimization efforts
from prospective users will continue to refine the model’s performance and applicability. This
is especially important in view of the fact that the RegCM system includes multiple
representations of different physics processes, which can be quite sensitive to the region of
application.

For this reason, it is very helpful to provide model users with some basic information of the
performance of a standard version of the model optimized over a variety of climatic settings,
which can then provide the basis of more detailed customizations for different applications.
Therefore, in this paper we extend the analysis of Giorgi et al. (2023) by presenting a version
of the model optimized and tested over nine domains used in the CORDEX-CORE effort
(Giorgi et al. 2021; Teichman et al. 2020; Coppola et al., 2020), along with a convection-
permitting experiment covering for the first time the entire European region. A number of
different aspects of model performance are assessed using a variety of observation datasets
for model validation, and for all experiments the model is driven at the lateral boundaries by
reanalyses of observations.

We first present in section 2 a brief summary of the main model features, the methodology
and setting for the simulations reported in section 3, results are discussed in section 4 and
summary and future outlooks are provided in section 5.

**RegCM5 model description**

RegCM5 includes both hydrostatic and non-hydrostatic dynamical cores, as well as a wide
range of physics options. It can be employed as a limited area model for any region globally
or using a tropical band configuration (Coppola et al., 2012). The significant enhancement in
RegCM5 compared to the previous version RegCM4 is the integration of the non-hydrostatic
dynamical core from the MOLOCH weather prediction model, along with some upgrades to
the model physics.

The MOLOCH dynamical core used in RegCM5 is described by Giorgi et al. (2023) and
references therein. It uses a hybrid terrain-following uniform vertical coordinate and an
Arakawa and Lamb C horizontal grid with uniform spacing and staggered wind components.

The model equations are expressed in terms of the variables \((T, P, \Pi, \Theta, u, v, w, q, T_v)\), where

- \(T\) is the temperature
- \(P\) is the pressure
- \(q_v, q_c, q_i\) are the mass mixing ratio of water vapor, liquid water and ice water
- \(\Pi = \left(\frac{\rho}{\rho_0}\right)^{\frac{R_d}{R}}\) is the Exner function
\[ \Theta_v = \frac{T_v}{\Pi} \] is the virtual potential temperature and

\[ T_v \approx T (1 + 0.61 q_v - q_c - q_i) \] is the virtual temperature

The prognostic equations for \( \Pi \) and \( \Theta_v \) are a good approximation of the exact thermodynamic and continuity equation of moist air. The horizontal and vertical derivatives are computed using a second order, centered finite difference scheme, while the time integration follows a three-step explicit scheme: vertical sound wave propagation with an implicit Euler-backward scheme with time step \( dt_s \), advection terms with a second-order total variation method with time step \( dt_a \), and physical parameterization terms added with a user-configured large time step \( dt_p \). The \( dt_a \) and \( dt_s \) time steps are integer fractions of \( dt_p \), i.e.

\[ dt_a = \frac{dt_p}{n_{adv}}, \quad dt_s = \frac{dt_a}{n_{sound}} \]

with \( n_{sound} \) and \( n_{adv} \) being user configurable parameters. The generalized vertical velocity is zero at the surface and at the model top. No explicit diffusion is required and numerical stability is attained by applying a second order spatial filter on the divergence of the horizontal wind with a user configurable coefficient.

For further technical details we refer to Giorgi et al. (2023) and Malguzzi et al. (2006) who provide comprehensive information on the model equations and solution procedures. A summary of the additional features available in the new RegCM5 model version optimized over the CORDEX-CORE domains is reported in Table 1.

| Table 1: Dynamics, Physics and Coupled Component Options Available in RegCM5. |
| Note. Bold letters highlight the options newly available since the RegCM5 version described by Giorgi et al. (2023). |

<table>
<thead>
<tr>
<th>Model aspects</th>
<th>Available options</th>
</tr>
</thead>
</table>
| Dynamics      | ● Hydrostatic, vertical pressure coordinate (Giorgi et al, 1993a)  
                ● Non-hydrostatic, vertical pressure coordinate (Coppola et al, 2012)  
                ● Non-hydrostatic, height based coordinate (MOLOCH, Malguzzi et al, 2006, Davolio et al. 2020) |
| Radiative transfer | ● Modified CCM3 (Kiehl et al, 1996)  
                        ● RRTM (Mlawer et al, 1997a,b) |
| Planetary Boundary Layer | ● Modified Holtslag (Holtslag et al. 1990)  
                               ● UW-PBL (Bretherton et al. 2004) |
| Cumulus convection | ● Simplified Kuo (Anthes et al. 1987, not available for MOLOCH dynamics)  
                                ● Grell (Grell 1993)  
                                ● MIT (Emanuel & Zivkovic-Rothman 1999)  
                                ● Tiedtke (Tiedtke 1989)  
                                ● Kain-Fritsch (Kain 2004) |
| Resolved scale precipitation | ● SUBEX (Pal et al, 2000)  
  ● WRF-single-moment-microphysics classes 5 (Hong, Dudhia and Chen, 2004)  
  ● Nogherotto-Tomkins (Nogherotto et al, 2016) |
|-------------------------------|---------------------------------------------------------------|
| Cloud fraction               | ● Sundqvist (Sundqvist, 1988)  
  ● Xu-Randall (1996)  
  ● Both modified according to Liang et al. (2005) |
| Land Surface                 | ● BATS (Dickinson et al. 1993)  
  ● CLM3.5 (Steiner et al. 2009)  
  ● CLM4.5 (Oleson et al, 2013)  
  ● Sub-grid BATS (Giorgi et al. 2003) and CLM4.5 |
| Land Use                     | ● Dynamical land use forcing from LUCAS LUC V1.1, based on LUH2 (Hoffmann et al. 2023) for the European Domain |
| Ocean fluxes                 | ● BATS (Dickinson et al. 1993)  
  ● Zeng (Zeng et al. 1998)  
  ● COARE (Fairall et al., 2003)  
  ● Diurnal sea surface temperature (Zeng & Beljaars 2005) |
| Interactive aerosols         | ● Organic and black carbon, SO4 (Solmon et al. 2006)  
  ● Dust (Zakey et al. 2006)  
  ● Sea salt (Zakey et al. 2008)  
  ● Gas-phase (Shalaby et al, 2012)  
  ● Pollen (Liu et al, 2016)  
  ● Implementation of Global Aerosol OPP Profile Reanalysis from MERRA-2 (Gelaro et al. 2017, last version available at: DOI: 10.34730/bc801a23b8bf48e98a50e23e909bf19c ), but only with one optical band (visible) |
| Interactive lake             | ● 1D diffusion/convection (Hostetler et al. 1993) |
| Interactive vegetation       | ● CLM4.5 CNDV (Shi et al, 2018) |
| Tropical band                | ● (Coppola et al 2012) |
| Coupling                     | ● RegCM-ES (Sitz et al. 2017)  
  ○ ROMS Ocean (Ratnam et al, 2009)  
  ○ MIT GCM Ocean (Artale et al. 2010)  
  ○ ChyM hydrology (Di Sante et al, 2019)  
  ○ BFM biogeochemical (Reale et al, 2020) |
| Sea ice                      | ● BATS (Dickinson et al. 1993) |
The RegCM5 model has been tested over the entire set of CORDEX-CORE domains, which were previously simulated with the RegCM4.7 version (Coppola et al., 2020; Giorgi et al., 2021). Additionally, the model was tested for the first time at a convection-permitting resolution over a pan-European domain. For each domain, multiple observations and reanalysis data have been utilized for model assessment, as reported in Table 2.

Table 2: Observational Datasets.

<table>
<thead>
<tr>
<th>Observed Datasets</th>
<th>Domain</th>
<th>Variables</th>
<th>Data type</th>
<th>Spatial Resolution</th>
<th>Temporal Resolution</th>
<th>Period</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPC_Global</td>
<td>Global Land</td>
<td>PRECIP</td>
<td>Gridded, Station based</td>
<td>0.50 degrees</td>
<td>DAILY</td>
<td>1979-2021</td>
<td>Chen et al. (2008)</td>
</tr>
<tr>
<td>TRMM</td>
<td>Tropics</td>
<td>PRECIP</td>
<td>Satellite observation based</td>
<td>0.25 degrees</td>
<td>3-HOURLY</td>
<td>1998-2017</td>
<td>Kummerow et al. (1998)</td>
</tr>
<tr>
<td>MSWEP</td>
<td>Global</td>
<td>PRECIP</td>
<td>Derived by optimally merging a range of gauge, satellite, and reanalysis estimates</td>
<td>0.10 degrees</td>
<td>DAILY</td>
<td>1979-2020</td>
<td>Beck et al. (2019)</td>
</tr>
<tr>
<td>GPCP</td>
<td>Global</td>
<td>PRECIP</td>
<td>Satellite observation based</td>
<td>0.25 degrees</td>
<td>DAILY</td>
<td>1979-2009</td>
<td>Adler et al. (2003)</td>
</tr>
<tr>
<td>CRU</td>
<td>Global Land</td>
<td>PRECIP</td>
<td>Station based</td>
<td>0.50 degrees</td>
<td>MONTHLY</td>
<td>1901-2015</td>
<td>Harris et al. (2020)</td>
</tr>
<tr>
<td>APHRO</td>
<td>India and East Asia</td>
<td>PRECIP</td>
<td>Grid</td>
<td>0.25 degrees</td>
<td>DAILY</td>
<td>1951-2007</td>
<td>Yatagai et al. (2009)</td>
</tr>
<tr>
<td>E_OBS</td>
<td>Europe Land</td>
<td>PRECIP</td>
<td>Grid</td>
<td>0.25 degrees</td>
<td>DAILY</td>
<td>1950-2015</td>
<td>Cornes et al. (2018)</td>
</tr>
<tr>
<td>Dataset</td>
<td>Country</td>
<td>Variable</td>
<td>Source Type</td>
<td>Resolution</td>
<td>Frequency</td>
<td>Time Period</td>
<td>References</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>-------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>CN05.1</td>
<td>China</td>
<td>PRECIP, TMEAN</td>
<td>Station based</td>
<td>0.25 degrees</td>
<td>DAILY</td>
<td>1961-2012</td>
<td>Wu &amp; Gao (2013)</td>
</tr>
<tr>
<td>ERA5</td>
<td>Global</td>
<td>WIND, PRECIP, CLOUD FRACTION, CLOUD WATER, CLOUD ICE, MEAN SEA LEVEL PRESSURE, TMEAN</td>
<td>Reanalysis</td>
<td>0.25 degrees</td>
<td>HOURLY</td>
<td>1940-Present</td>
<td>Hersbach et al. (2020)</td>
</tr>
<tr>
<td>IBTrACS</td>
<td>Global</td>
<td>TROPICAL CYCLONES TRACK</td>
<td>Merging datasets from different agencies</td>
<td>-</td>
<td>DAILY</td>
<td>1842-Present</td>
<td>Knapp et al. (2010, 2018)</td>
</tr>
<tr>
<td>REGNIE</td>
<td>Germany</td>
<td>PRECIP</td>
<td>Station based</td>
<td>1 km</td>
<td>DAILY</td>
<td>1961-2014</td>
<td>Rauthe et al., 2013</td>
</tr>
<tr>
<td>RADKLIM</td>
<td>Germany</td>
<td>PRECIP</td>
<td>Radar based (rain gauges calibration)</td>
<td>1 km</td>
<td>HOURLY</td>
<td>2001-2009</td>
<td>Kreklo et al. (2020)</td>
</tr>
<tr>
<td>SPAIN02</td>
<td>Spain</td>
<td>PRECIP</td>
<td>Station based</td>
<td>0.11 degrees</td>
<td>DAILY</td>
<td>1971-2010</td>
<td>Herrera et al., 2010</td>
</tr>
<tr>
<td>CARPATCLIM</td>
<td>Carpathian States</td>
<td>PRECIP</td>
<td>Station based</td>
<td>0.1 degrees</td>
<td>DAILY</td>
<td>1961-2010</td>
<td>Szalai et al. (2013)</td>
</tr>
<tr>
<td>ENG_REGR</td>
<td>Great Britain</td>
<td>PRECIP</td>
<td>Station based</td>
<td>5 km</td>
<td>DAILY</td>
<td>1990-2010</td>
<td><a href="http://www.precisrcm.com/Erasmo/ncic.uk.11.tgz">http://www.precisrcm.com/Erasmo/ncic.uk.11.tgz</a></td>
</tr>
<tr>
<td>COMEPHO RE</td>
<td>France</td>
<td>PRECIP</td>
<td>Reanalysis based on radar and rain gauges</td>
<td>1 km</td>
<td>HOURLY</td>
<td>1997-2017</td>
<td>Tabary et al. (2012)</td>
</tr>
<tr>
<td>GRIPHO</td>
<td>Italy</td>
<td>PRECIP</td>
<td>Station based gridded dataset</td>
<td>3 km</td>
<td>HOURLY</td>
<td>2001-2016</td>
<td>Fantini (2019)</td>
</tr>
<tr>
<td>EURO4M</td>
<td>Alps</td>
<td>PRECIP</td>
<td>Station based gridded dataset</td>
<td>5 km</td>
<td>DAILY</td>
<td>1971-2008</td>
<td>Isotta et al. (2014)</td>
</tr>
<tr>
<td>METNO</td>
<td>Norway</td>
<td>PRECIP</td>
<td>Station based gridded dataset</td>
<td>1 km</td>
<td>DAILY</td>
<td>1980-2008</td>
<td>Mohr et al. (2009)</td>
</tr>
</tbody>
</table>
All simulations use ERA5 reanalysis fields (Hersbach et al., 2020) as initial and lateral boundary conditions. Specific model configurations for each domain, including spatial resolution and the simulation period, are provided in Table 3.

**Table 3:** Model configuration for each domain.

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>Period</th>
<th>Horizontal Resolution</th>
<th>Vertical Resolution</th>
<th>Boundary Layer Scheme (ib ltyp)</th>
<th>Cumulus convection scheme (icup_lnd/ocl)</th>
<th>Moisture scheme (ipptis)</th>
<th>Cloud fraction algorithm (icldfrac)</th>
<th>Dynamical Land Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australasia</td>
<td>2000-2009</td>
<td>25 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture</td>
<td>SUBEX</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nogherotto/Tompkins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Asia</td>
<td>2000-2009</td>
<td>25 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture</td>
<td>Xu-Randall empirical</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nogherotto/Tompkins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South East Asia</td>
<td>2000-2009</td>
<td>25 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture</td>
<td>SUBEX</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nogherotto/Tompkins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td>2000-2009</td>
<td>25 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture</td>
<td>SUBEX</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nogherotto/Tompkins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central America</td>
<td>2000-2009</td>
<td>25 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture</td>
<td>SUBEX</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nogherotto/Tompkins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>2000-2004</td>
<td>3 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture</td>
<td>Xu-Randall empirical</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>1980-2010</td>
<td>12 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture</td>
<td>Xu-Randall empirical</td>
<td>YES</td>
</tr>
<tr>
<td>Region</td>
<td>Period</td>
<td>Resolution</td>
<td>Levels</td>
<td>PBL Model</td>
<td>Moisture Model</td>
<td>Validation</td>
<td>Percentage</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>--------</td>
<td>-----------</td>
<td>----------------</td>
<td>------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>South Asia</td>
<td>2000-2009</td>
<td>25 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture Nogherotto/To mpkins</td>
<td>Xu-Randall empirical</td>
<td>NO</td>
</tr>
<tr>
<td>North America</td>
<td>2000-2009</td>
<td>25 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture Nogherotto/To mpkins</td>
<td>Xu-Randall empirical</td>
<td>NO</td>
</tr>
<tr>
<td>Africa</td>
<td>2000-2009</td>
<td>25 km</td>
<td>30 levels</td>
<td>Holtslag PBL</td>
<td>Tiedtke/Tiedtke</td>
<td>Explicit moisture Nogherotto/To mpkins</td>
<td>SUBEX</td>
<td>NO</td>
</tr>
</tbody>
</table>

The model validation was conducted over the set of sub-regions identified in the AR6 WGI IPCC report covered by the RegCM5 domains. The regions are described by Iturbide et al. (2020). Various metrics were computed to validate the model, encompassing both mean climate and extreme climate distribution, as shown in Table 4.

**Table 4:** Metrics used for model validation.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Definition</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{\text{mean}}$</td>
<td>Daily mean 2-m temperature</td>
<td>°C</td>
</tr>
<tr>
<td>$T_{\text{max}}$</td>
<td>Daily maximum 2-m temperature</td>
<td>°C</td>
</tr>
<tr>
<td>$T_{\text{min}}$</td>
<td>Daily minimum 2-m temperature</td>
<td>°C</td>
</tr>
<tr>
<td>pr</td>
<td>Daily/hourly total precipitation</td>
<td>mm/day, mm/hr</td>
</tr>
<tr>
<td>pr-frq</td>
<td>Total number of wet days/hours (i.e., days with total precipitation greater than 1 mm)</td>
<td>day/year</td>
</tr>
<tr>
<td>pr-int</td>
<td>Average amount of wet-day precipitation</td>
<td>mm/day, mm/hr</td>
</tr>
<tr>
<td>p99</td>
<td>The 99th percentile of the precipitation distribution over the time period considered</td>
<td>mm/day, mm/hr</td>
</tr>
<tr>
<td>p99.9</td>
<td>The 99.9th percentile of the precipitation distribution over the time period considered</td>
<td>mm/day, mm/hr</td>
</tr>
<tr>
<td>cl</td>
<td>Cloud Fraction</td>
<td>%</td>
</tr>
</tbody>
</table>
The mean seasonal bias for 2 meter, mean, maximum and minimum temperature ($T_{\text{mean}}$, $T_{\text{max}}$, and $T_{\text{min}}$ respectively), mean precipitation ($pr$), precipitation intensity and frequency ($pr_{-\text{int}}$ and $pr_{-\text{frq}}$), as well as the annual total precipitation above the 99th percentile ($p_{99}$), were used for the validation of the model mean climatology (definition of the metrics can be found in Table 4). For temperature, the model results are compared with observations from the Climate Research Unit (CRU) dataset. For mean precipitation, the reference dataset is the Global Precipitation Climatology Centre (GPCC), and for precipitation intensity/frequency and $p_{99}$, is the Climate Prediction Center (CPC) one. The seasonal means are first calculated over the baseline period (1980 to 2010 for Europe and 2000 to 2009 for all other domains) at the original resolutions and are subsequently interpolated (distance-weighted average for temperature, and nearest neighbour for precipitation and related metrics) to the resolution of the observations. The area-weighted averages of all variables are then computed over the AR6 WGI IPCC regions contained within each domain, and the biases are then derived by taking the difference between the simulated and observed values. The global bias is obtained in the same way, except that the area-weighted average is calculated over all grids of all domains.

### Precipitation distribution

Boxplots were computed for daily precipitation in all regions considered, from RegCM4, RegCM5 and observations. We use the station-based data from CPC except for Europe, for which the observation dataset is E-OBS. Due to the steepness of the distribution, the box plots include the 5th and 95th and 99th percentiles.

Note that over some regions, and particularly the Mediterranean, RegCM4 exhibited a notable overestimation of extreme events due to the occurrence of numerical point storms, a problem that is considerably improved in RegCM5. Therefore, in the box plots, events with excessively large amounts in RegCM4 were excluded by adjusting the plot to align with the distribution from observations and RegCM5.

Hourly precipitation distributions for the period 2000-2004 were calculated for the RegCM5 CP and 12 km simulation over Europe and compared with high-resolution hourly observations over Italy, Switzerland, Germany, France and Great Britain (see Table 2). Furthermore, results were compared with the ERA5 reanalysis estimates. Distributions are calculated by taking all available time steps and grid points within each dataset considered. Some of the observational datasets did not have observations at the start of the RegCM5 simulations (e.g. Switzerland observational dataset starts in 2003). Therefore, in order to consider a consistent time period for the observations and model simulations, we used the first five available years for each of the observational datasets (e.g. Switzerland 2003-2007).
Daily precipitation distributions are calculated for 2000-2004 for the Europe RegCM5 model simulations, ERA5 and all available observations in the simulated region. In addition to the observational datasets mentioned above, daily precipitation estimates from Sweden, Norway, Spain and the Carpathians are also available (see table 2). All the biases were computed interpolating each observational dataset on the model grid.

Precipitation sub daily analysis

Seasonal daily precipitation cycles were computed for Europe, analysing both the 12 km and the 3 km simulations. The comparison was carried out against ERA5 data as well as different sub-regional hourly observation datasets: GRIPHO (Italy), Rdisaggh (Switzerland), RADKLIM (Germany), COMEPHORE (France) and CEH-GEAR (Great Britain). Each high-resolution dataset was interpolated on the coarser model grid and the daily cycle was computed spatially averaging only in the region covered by observations. Precipitation intensity and frequency for the hourly observation and RegCM5 datasets were calculated using hourly minimum precipitation thresholds of 0.1 mm/hr and 0.5 mm/hr in order to investigate the uncertainties in the data at very low intensities, which can strongly influence the biases. Note that the choice of threshold does not influence the p99.9 estimates as the whole distribution (including dry hours) is used to calculate this variable.

Taylor diagram

Taylor diagrams were used to validate the mean seasonal precipitation and temperature against several reference datasets. For precipitation, the model results are compared with ERA5, CRU, MSWEP, CPC, and GPCC. For temperature, ERA5 and CRU are used, except for additional observation datasets for Europe and East Asia. Specifically, for Europe, precipitation and daily mean temperature are compared against E-OBS, while for several subregions of East Asia, they are compared against APHRO and CN05.1. For each subregion, the gridded seasonal averages of the observed and simulated data are used to calculate the area-weighted centered pattern correlation and the ratio between the simulated and observed standard deviations, which are then used to generate the diagrams.

Cloud distributions

Vertical profiles were computed over each region for the mean seasonal cloud fraction, cloud liquid water and cloud ice in June-July-August (JJA) and December-January-February (DJF) using twelve pressure levels: 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150 and 100 hPa. The calculations were done for both RegCM5 and the ERA5 reanalysis data and covered the period 2000-2009 for all domains, except for Europe, for which 1980-2010 was used.

Upper level circulations

Composites of zonal and meridional wind were computed for 3 different pressure levels, i.e., 850, 500, and 250 hPa. RegCM5 includes a function to perform this task, called sigma2p. This function is first executed to interpolate the wind components from the sigma coordinates to pressure levels. Wind at the selected levels is then extracted, and its seasonal means are
calculated over the baseline period. Results of different domains are subsequently interpolated onto the grids of the reference dataset, i.e., ERA5, using distance-weighted average mapping. The composite of global wind is then obtained by directly combining the wind of all domains. In cases where there is an overlap between multiple domains, the average is calculated. For ERA5, wind at the three pressure levels averaged over 2000-2009 is used for all domains except for Europe, where the 1980-2010 average is employed. Wind of the reference dataset is then masked with respect to the RegCM composite to facilitate an intuitive comparison between the two.

**Tropical and extratropical cyclones**

Tropical and extratropical cyclones were tracked in each domain, but a graphical representation was created by combining all domains into a single map. Three different algorithms for identifying and tracking tropical cyclones (Reboita et al., 2010; Fuentes-Franco et al., 2014, 2017; Hodges, 1994, 1995, 1999) were employed, while one algorithm was used for extratropical systems (Reboita et al., 2010).

**a) Reboita et al. (2010)’s algorithm**

This algorithm identifies and tracks tropical and extratropical cyclones using cyclonic relative vorticity every 6 hours (0000, 0600, 1200, and 1800 UTC). Before applying the algorithm, the horizontal wind components at 925 hPa (zonal and meridional) are interpolated to a grid with a resolution of 1.5° x 1.5° in latitude and longitude. Once the data are provided to the algorithm, relative vorticity is computed and smoothed to reduce noisy features using the Cresmann (1959) method. The algorithm consists of three main steps: (1) initially, in a specific time step of the dataset, it searches for the minimum relative vorticity by comparing each grid point value with those of 24 surrounding points (nearest-neighbour method). A grid point is a cyclone center candidate when a minimum of relative vorticity is found and is smaller or equal to a threshold (defined as -1 x 10^{-5} s^{-1}); (2) the coordinates of the grid point identified in (1) are located in the follow time step of the dataset to limit the search area to the 24 neighboring grid points; and (3) once two positions are known, the algorithm calculates the displacement velocity of the cyclone center and uses it as an initial estimate (first guess) to locate the cyclone center in the following time step. This procedure continues until the dissipation (cyclolysis) of the cyclone. When cyclolysis occurs, the algorithm returns to the specific time step of the initial identification and searches for other grid points that could be a cyclone center, and all three steps are repeated. After the cyclone position is identified at a given time step, the algorithm performs an interpolation on a high resolution grid to refine the cyclone center searching a new position in a 250 km radius. Only cyclones with lifetime equal to or higher than 24 hours and equal to or lower than 10 days are included in the statistics. It is important to highlight that we will present all synoptic cyclonic systems detected by the algorithm and not only extratropical and tropical cyclones. For selection of a specific cyclone type, the tracking output would need to be used as input to the Cyclone Phase Space (CPS; Hart, 2003), which analyses the vertical structure of the systems.

**b) Fuentes-Franco et al. (2014, 2017)’ algorithm**

This algorithm, named Kyklop (Fuentes-Franco et al., 2017), is configured to work with three variables (near-surface wind speed at 10 m, mean sea level pressure -MSLP-, and sea surface temperature - SST) and with the time frequency and horizontal resolution (see [https://github.com/kyklop-climate/kyklop/blob/master/kyklop/kyklop.py](https://github.com/kyklop-climate/kyklop/blob/master/kyklop/kyklop.py)) of the NetCDF file.
provided as input. In this study, 3-hourly data (0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 UTC) are used. The Kyklop algorithm has two main procedures: (1) it searches for grid points that are candidates to be a tropical cyclone in all time steps and (2), subsequently performs the matching of grid points to determine the cyclone’s trajectory. This logical sequence differs from Reboita et al. (2010) in that, in their approach, once a grid point is identified as a cyclone center candidate, it is tracked until cyclolysis. In (1), for each time step, Kyklop searches grid points that satisfy the following criteria: wind speed >20 m s\(^{-1}\), MSLP <1005 hPa, and SST >25 °C. As these conditions may be satisfied by some neighbouring grid points, the centroid of the area encompassed by these grid points is considered as the center of the tropical cyclone. In (2), for each detected cyclone grid point in a specific time step, its tracking (following positions) is carried out by checking on next time steps if there are grid points that meet the conditions presented in (1) within a radius of 6° × 6° longitude–latitude. These conditions need to persist for at least 24 hours.


Hodges (1994, 1995, 1999) named his algorithm TRACK, which searches for various types of cyclones based on relative vorticity. However, this algorithm can also be configured for identifying only tropical disturbances. In this case, the TRACK uses the zonal and meridional wind components at different vertical levels (10 m, 850, 700, 600, 500, 400, 300 and 200 hPa), and at 6-hour intervals (0000, 0006, 1200 and 1800 UTC). The identification of tropical disturbances involves three main steps: (1) pre-processing filtering, (2) tracking performed following Hodges’s references, and the (3) post-tracking filtering - an additional procedure integrated within TRACK (Hodge et al., 2017). The data used in this study were first interpolated to a regular grid of 0.25° x 0.25° before being processed by TRACK. In step (1), the algorithm calculates the vertically averaged relative vorticity between 850-600 hPa. Subsequently, a spectral filter (triangular truncation) is applied, retaining wavenumbers between 6 and 63, in order to remove the noise associated with the smallest spatial scales and the large-scale background. In step (2), the nearest-neighbor method is applied to the processed data from step (1) to identify all tropical disturbances (tropical cyclones will be separated from all systems in step 3). Unlike Reboita et al. (2010), TRACK standardizes the relative vorticity field to positive values in both hemispheres, so it identifies the cyclonic disturbances by maxima of relative vorticity, and, in addition, it applies a threshold: candidates for tropical disturbance need to have relative vorticity > 5 \times 10^{-6} \text{ s}^{-1} (in the Southern Hemisphere the field is scaled by -1). The tropical disturbance location is then refined using a B-spline interpolation. Additionally, the algorithm refines the tracks by minimizing a cost function for track smoothness. The final step (3) is post-tracking filtering, selecting only the tropical cyclones from all tracked tropical disturbances. Tropical cyclones are identified based on three parameters describing their structure: presence of coherent vertical symmetry (presence of a maximum of relative vorticity at each vertical level), warm core, and high near-surface wind speeds. These three parameters must be satisfied for at least 2 days, with a minimum of 24 hours over the ocean. To identify the symmetry, the scheme searches the maximum relative vorticity at the vertical levels (850, 700, 600, 500, 400, 300 and 200 hPa). The algorithm uses the location of tropical disturbance computed at the 850-600 hPa level as the starting point, and then a circle with a radius of 5° (geodesic) is delimited. The maximum relative vorticity is then searched inside this area, and the location of this maximum is used as reference for the level above and this procedure is repeated until the uppermost level. The warm core is calculated as the difference between the relative vorticity fields at 850 and 200 hPa (at T63 resolution) and must be greater than 6 \times 10^{-5} \text{ s}^{-1} (indicating stronger winds near the surface than at upper levels). Additionally, the 10-m wind
speed must be greater than 17.5 m s\(^{-1}\) and is searched within a 6° radius from the cyclone center identified using the vorticity average between 850-600 hPa.

All algorithms provide as output the latitude and longitude at each time step of the cyclone’s lifecycle and other features such as MSLP, relative vorticity etc., depending on the algorithm. With the tracking information, it is possible to compute the track density, which is the number of cyclones passing by an area of 1° x 1° divided by the area of this box. We compared the RegCM5 performance in reproducing the cyclonic systems against the ERA5 reanalysis when working with the Reboita et al. (2010) algorithm and against the International Best Track Archive for Climate Stewardship (IBTrACS, version v04; Knapp et al. 2010, 2018) for the other algorithms. IBTrACS collects observed tropical cyclone data from 11 agencies around the world covering all major ocean basins and provides 6-hour data of tropical cyclones locations.

**Results**

**CORDEX-CORE domains**

The mean regional biases for mean, maximum, and minimum temperature, mean precipitation, precipitation frequency and intensity, and precipitation above the 99th percentile are presented in Figure 1 for all four seasons (DJF, MAM, JJA, and SON) and for each region, as well as for the global average. Mean temperature biases are generally constrained between -2 and 2 degrees, except for the two northernmost regions of the North American continent (NWN and NEN) and the northernmost eastern region of Asia (RFE) in DJF, where a stronger warm bias is evident.
Figure 1. Mean seasonal bias of each region for Tmean, Tmin, Tmax, pr, pr-freq, pr-int and the annual value of p99. The period covered is 2000-2009, except for the European domain (MED, WCE and NEU regions): 1980-2010. A global mean season bias table with its respective values has been added to the figure.

This is likely due to the overestimation of cloud water for low and middle clouds which increases downward infrared radiation (Figure 2), derived from an excessively stable boundary not well reproduced by the Holtslag PBL scheme (see Table 3), as previously noted in Güttler et al., 2014, or Bae et al., 2023; Gao and Giorgi (2017).

Figure 2. Cloud liquid water vertical profiles for DJF (a) and JJA (b). The period covered is 2000-2009, except for the European domain (MED, WCE and NEU regions): 1980-2010.

Other outlier regions include central Asia, where the Tibetan Plateau is located, showing a cold bias between 3 and 4 degrees in DJF. This is possibly at least partially due to the well-known sparse nature of available stations at high elevations, especially considering that gauge stations are often placed in valleys and only few or none on mountain tops (Xu et al., 2009). Overall, the model has a tendency for a cold bias in maximum (T\text{max}) temperature and a warm bias in minimum (T\text{min}) temperature across almost all seasons and regions. This tendency is associated with a systematic overestimation of the lower-level cloud fraction (see Figure S1), more pronounced in winter than in summer in both hemispheres but consistently present due to an overestimation of cloud liquid water (Figure 2). In this case, biases are generally within a 2-degree range, except for the warm T\text{min} bias in the Caribbean, western South Africa, and
Australasia regions, where the overestimation of the cloud profile is pronounced, and the Tibetan Plateau, showing a cold bias mainly in winter and spring. Cloud ice vertical profiles for DJF and JJA are shown in Figure S2.

In Figure 3a, Taylor diagrams are presented to validate the spatial temperature patterns in each domain and region, considering only land points. The results show for all seasons a strong correlation (0.9 or higher) between the model and the ERA5 and CRU datasets, except for NSA in South America and the Caribbean region (with respect to CRU), where the correlation drops to 0.7. Similar correlations are observed in Central Africa for all seasons except SON and Western Southern Africa for DJF and MAM. Spatial temperature variability is well captured in all regions, with a tendency to overestimate it in South and Central America (mostly in all regions and seasons) and East Asia, where variability is slightly underestimated for the northernmost regions and overestimated for the southern ones. Similar behaviour is observed for maximum and minimum temperature in Figure S3-S4.

Figure 3. Taylor diagrams for the mean temperature (panel a) and precipitation (panel b) for selected domains. Symbols represent seasons and colors are the subregions of a specific domain.

Taylor diagrams for precipitation are presented in Figure 3b for selected domains and in Figure S5 and S6 for the remaining domains and regional observational datasets. Five different datasets are used for comparison, varying in spatial resolution and origin. Correlation and spatial variability for all domains are in better agreement with the MSWEP and GPCC observational products, which have the highest resolution. Spatial correlation of precipitation ranges between 0.5 and 0.8 in most seasons and regions (Figure 3b and S3). The model tends to overestimate spatial variability, especially in South America, East Asia, and Africa.

Figure 4 illustrates the comparison of the precipitation intensity distribution in each region between the RegCM5 and RegCM4 models and the observations through box plots. RegCM5 shows a good representation of the precipitation distribution compared to observations and is more realistic than the previous model version, especially for the long tails and most extreme events, where the model strongly ameliorated the problem of numerical point storms found in RegCM4.
In Figure 5, the 850 hPa wind field is analyzed to validate monsoon circulation in different continents. The model well represents the South Asia monsoon system in terms of intensity and direction of the wind jet. It slightly overestimates the West African monsoon with more inland penetration and a west-east direction compared to observations. The Central America and North America monsoons are well located with correct intensity, while the East Asia monsoon circulation intensity is slightly underestimated. The South America Low-Level Jet (SALLJ) is well reproduced in intensity and direction in the austral summer (DJF), while during JJA the jet intensity over south Bolivia and Paraguay is weaker in the model compared to ERA5. The Caribbean Low-Level Jet is well positioned in both seasons with the right intensity and direction. The wind fields at 500 and 200 hPa are also reported in Figure S7 for completeness.
The model's ability to reproduce tropical and extra tropical cyclone tracks was tested using the different tracking algorithms discussed in Methods. Figure 6 (a,b) shows cyclone track densities in the RegCM5 simulations and the ERA5 reanalysis calculated with the tracking algorithm of Reboita et al. (2010). The model has a good performance in locating the core of trajectories in all regions but in some cases with differences in density from ERA5. While there is overestimation over the western Indian Ocean (coastal region of the Arabian Peninsula) and in the extratropical northern European areas, an underestimation occurs in western North America, southern Indian ocean and in the eastern coast of South America. The other two tropical cyclone tracking schemes (Figure 6d and 6e) also reproduce the areas of maximum track density but exhibit different behaviors in the western tropical Atlantic Ocean, southern Indian Ocean region, and eastern Asia tropical Pacific Ocean. The cyclone track density identified using the Reboita et al. (2010) and Fuentes-Franco et al. (2014, 2017) algorithms is underestimated in RegCM5 in the western tropical Atlantic compared to the Hodges (1994, 1995, 1999). However, the Hodges et al. (1994, 1995, 1999)’s algorithm overestimates track density in the eastern Asia tropical Pacific Ocean compared to the other two schemes. Differences are also found in the northern Australia coasts and southern Indian Ocean. These results highlight the importance of the choice of the tracking algorithm and the associated uncertainty in model results.

Figure 5. Wind intensity (m/s) and direction (arrows) at 850 hPa.
Figure 6. Total track density of all synoptic cyclones identified in ERA5 (panel a) and RegCM5 (panel b), from 2000 to 2009, using Reboita et al. (2010)’s algorithm. The unit is the number of cyclones with the center inside a 1º x 1º grid-box; total track density of tropical cyclones identified in the IBTrACS (panel c) and RegCM5 (panel d), from 2000 to 2009, using Fuentes-Franco et al. (2014, 2017)’s algorithm. The unit is the number of cyclones with the center inside a 1º x 1º grid-box; panel e is the same as panel d but using Hodges (1994, 1995, 1999)’ algorithm.
As mentioned, by being much more computationally efficient than previous versions of the model, RegCM5 allows simulations for a pan-European domain at convection-permitting resolution. Figure 7 illustrates a time sequence of summer convective events in the southern regions of Italy and Greece within the 3km CP domain, which is highlighted in the grey square, while ERA5 precipitation is shown outside of this region. The sequence starts on the night of June 8, 2000. A storm enters the CP domain from the western boundary, crossing Ireland throughout the day. Convection initiates in Sicily, Calabria, and northern Greece in the early afternoon, reaching its peak at 18:00 UTC and diminishing later in the evening. The time lapse demonstrates the consistency between the ERA5 boundary conditions and the CP model simulation in the evolution of the storm event.

Figure 7. Precipitation estimates [mm hr⁻¹] from ERA5 and RegCM5 CP for 6 different time steps on the 8th and 9th June 2000. The precipitation estimates inside the gray box are from the RegCM5 CP simulation, while the rest of the domain outside the gray box shows the ERA5 precipitation estimates. The insert figure in each panel shows the REGCM5 CP precipitation estimates over a smaller section of the full domain to highlight the presence of the diurnal cycle in convective activity.
Figure 8 shows seasonal precipitation and temperature biases, precipitation frequency and intensity, and p99 biases for the convection-parametrized 12 km resolution run and the explicit convection 3km resolution run. Table 2 presents the observed datasets used for model validation, which are station-based or radar-based national datasets for various European countries. Both resolutions exhibit similar mean temperature and precipitation biases, mean daily bias frequency of events, while improvements in daily precipitation intensity and P99 biases at the 3km resolution are found, in particular reducing the dry bias in central northern Europe.

Figure 8. Mean seasonal bias for Europe CP and Europe 12 km simulation are shown as calculated with respect to the high resolution observation datasets. Mean seasonal daily precipitation and mean seasonal temperature are shown in panel a, the seasonal daily precipitation intensity and the precipitation frequency (> 1mm/day) in panel b and the annual P99 bias in panel c. For each variable the left column shows the CP simulation, while the right column represents the results for the Europe 12 km simulations.

Figure 9 compares precipitation probability density function distributions at a daily timescale for each observed dataset. The CP precipitation distribution aligns closely with the high-resolution datasets, outperforming the 12 km resolution model and ERA5 precipitation distribution in most regions. However, in Norway, the CP model distribution underestimates the observed one, and in the Carpathians and Spain regions, the model overestimates the precipitation distribution, possibly due to the lower resolution of station-based observations.
Figure 9. Probability density function distributions of the daily precipitation [mm day\(^{-1}\)] for the 10 regions investigated in the European domain. Each panel shows the distribution estimated from combining all available data in each domain for the years 2000-2004 for RegCM5 CP (orange), RegCM5 12km (red), ERA5 (blue) and observations (black). Details about the observational datasets for each region can be found in Table 2.

Supplementary Figure S8 illustrates daily temperature PDFs for the same regions, showing reasonable temperature distributions with a slight underestimation of maximum temperature values in areas of complex topography, such as the Alps and Swiss regions, likely attributed to a precipitation overestimation.
Finally, Figure 10, 11, and 12 present precipitation statistics at hourly timescale. Frequency, intensity, and very extreme hourly precipitation (p99.9) are computed for events above the threshold of 0.5 mm/h, revealing an orographically driven positive bias. Despite some regional discrepancies, the explicit representation of convection in the 3km resolution run improves systematic biases compared to the 12km simulation across all statistics and seasons. Supplementary Figures S9a-S9b show results with a more commonly used threshold of 0.1mm/h, indicating a noticeable negative and positive bias for intensity and frequency, respectively, in the 3km resolution, primarily attributed to very light events occurring between 0.1 and 0.5 mm/h. This is also evident in Figure 11, where the hourly precipitation distributions are reported for five regions. The high resolution model precipitation matches well the observed distribution with the only mismatch occurring in the range 0.1-0.5 mm/h for all the model resolutions and the ERA5 precipitation distributions.

![Figure 10](image-url)

**Figure 10:** Precipitation intensity, wet frequency and P99.9 seasonal bias for hourly REGCM5 CP (panel a) and REGCM5 12km (panel b) versus high resolution observations. In each panel, the first column shows the seasonal biases for precipitation intensity, second column the precipitation frequency bias and the third column the P99.9 bias. The threshold used as the minimum precipitation for the REGCM5 simulations is 0.5 mm/hr. Figure S9 (panels a and b) shows the same seasonal biases but using the minimum threshold of 0.1 mm/hr.
Figure 11. PDFs of hourly precipitation for the RegCM5 Convection Permitting simulation (orange), the RegCM5 12 km simulation (red), ERA5 (blue) and high resolution observations (black) from 5 regions (Great Britain, Germany, France, Switzerland and Italy). Each figure represents the distribution based on all the data available over the domain and time interval investigated. The bin size resolution is 0.5 mm/hr. The insert figure in each panel shows a breakdown of the three lowest precipitation intensity bins in the main panel, using a bin size resolution of 0.1 mm/hr.
Finally, Figure 12 shows the daily cycle of five precipitation statistics for the same five regions of Figure 11 and the JJA season. The explicit representation of convection successfully reproduces both the phase and amplitude of the diurnal cycle in most statistics and regions. The daily cycles for DJF, SON and MAM are reported in Figure S10 for completeness.

![Diurnal cycles JJA](image)

Figure 12: Diurnal cycles for mean precipitation (first column), precipitation intensity (second column), precipitation frequency (third column), p99 (fourth column) and p99.9 (fifth column) in JJA for 5 regions in Europe: Great Britain (top row), France (second row), Germany (third row), Switzerland (fourth row) and Italy (bottom row). The same figures for DJF, SON and MAM are shown in the Supplementary material.

### Summary and Outlook

The Regional Climate Modeling system (RegCM) has evolved significantly since its inception, with versions such as RegCM4 and RegCM4-NH playing pivotal roles in climate research and participating in international projects such as CORDEX. These models, however, require relatively small time steps, and thus present limitations especially when applied at CP resolutions. The recently developed RegCM5 incorporates the dynamical core from the non-hydrostatic weather prediction model MOLOCH's to enhance model speed and stability. This paper aims to comprehensively evaluate the performance of RegCM5, focusing on convection-parametrized and convection-permitting scales across various CORDEX-CORE domains, and including for the first time a pan-European domain at convection-permitting resolution. The assessment encompasses temperature biases, precipitation patterns, monsoon circulations, extratropical and tropical cyclone tracks, and the model's ability to explicitly simulate convective events.

The evaluation of RegCM5 shows important improvements in addressing challenges posed by higher resolutions, offering improved capabilities for understanding climate dynamics and
projections. The model demonstrates good performance in capturing temperature patterns, precipitation distributions, and monsoon circulations across various regions. The introduction of RegCM5's pan-European convection-permitting domain shows improved representation of daily and hourly precipitation distribution and diurnal cycle compared to the convection parametrized model version and illustrates the possibility to reach such resolution for larger model domains.

The model is currently available for use by the RegCM community and other prospective users. In this paper we have used for the different domains, model configurations that can be adopted as starting points for optimizing the model performance for different applications. Being a new development, the model needs to be further tested, and in this regard the contribution and feedback from the broader model community is essential. We are currently further improving the model capabilities, for example updating the land surface scheme CLM, the PBL scheme and including a two moment 6 hydrometeors microphysical scheme, and fine tuning some of the model’s available physics options. We are also planning to develop a model version usable on GPU-based computing architectures. We expect that RegCM5 will be the basic model version used by the RegCM community and maintained by the ICTP development team over the next several years.

Data Availability Statement

The RegCM5 model code is available at the web site: https://zenodo.org/record/7548172#.Y8gVV7TMKUk.

The data used in this work can be found at the following web sites:

http://www.euro4m.eu/datasets.html (EURO4M-APGD),

References


NCAR, Boulder, CO


26


https://doi.org/10.1007/s00382-020-05494-x


Supplementary Figures.
**Figure S1**: Cloud fraction vertical profiles for DJF (a) and JJA (b). The period covered is 2000-2009, except for the European domain (MED, WCE and NEU regions): 1980-2010.

**Figure S2**: Cloud ice vertical profiles for DJF (a) and JJA (b). The period covered is 2000-2009, except for the European domain (MED, WCE and NEU regions): 1980-2010.
Figure S3: Taylor diagrams for the maximum temperature. Symbols represent seasons and colors are the subregions of a specific domain.
Figure S4: Taylor diagrams for the minimum temperature. Symbols represent seasons and colors are the subregions of a specific domain.
Figure S5: Taylor diagrams for precipitation for the remaining domains. Symbols represent seasons and colors are the subregions of a specific domain.

Figure S6: Taylor diagrams for precipitation (left panel) and temperature (right panel) with available regional datasets. (Note: For East Asia, results for RFE and ESB are not shown since APHRO and CN05.1 only cover a small portion of these subregions). Symbols represent seasons and colors are the subregions of a specific domain.
Figure S7: Wind field at 500 hPa (upper panel) and 200 hPa (lower panel).
Figure S8: Probability density function distributions of the daily temperature [°C] for the 10 regions investigated in the European domain. Each panel shows the distribution estimated from combining all available data for the years 2000-2004 for RegCM5 CP (orange), RegCM5 12km (red), ERA5 (blue) and observations (black). Details about the observational datasets for each region can be found in table 2.
Figure S9: Precipitation intensity, wet frequency and P99.9 seasonal bias for hourly REGCM5 CP (panel a) and REGCM5 12km (panel b) versus high resolution observations. The seasonal biases are the same as in Figure 10, but using 0.1 mm/hr as threshold for the minimum precipitation in the REGCM5 simulations.
Figure S10: Diurnal cycles for mean precipitation (first column), precipitation intensity (second column), precipitation frequency (third column), p99 (fourth column) and p99.9 (fifth column) in DJF (panel a), MAM (panel b) and SON (panel c). In each panel, the results are shown for the following 5 regions in Europe: Great Britain (top row), France (second row), Germany (third row), Switzerland (fourth row) and Italy (bottom row).
Figure S1: Cloud fraction vertical profiles for DJF (a) and JJA (b). The period covered is 2000-2009, except for the European domain (MED, WCE and NEU regions): 1980-2010.

Figure S2: Cloud ice vertical profiles for DJF (a) and JJA (b). The period covered is 2000-2009, except for the European domain (MED, WCE and NEU regions): 1980-2010.
Figure S3: Taylor diagrams for the maximum temperature. Symbols represent seasons and colors are the subregions of a specific domain.
Figure S4: Taylor diagrams for the minimum temperature. Symbols represent seasons and colors are the subregions of a specific domain.
Figure S5: Taylor diagrams for precipitation for the remaining domains. Symbols represent seasons and colors are the subregions of a specific domain.

Figure S6: Taylor diagrams for precipitation (left panel) and temperature (right panel) with available regional datasets. (Note: For East Asia, results for RFE and ESB are not shown since APHRO and CN05.1 only cover a small portion of these subregions). Symbols represent seasons and colors are the subregions of a specific domain.
Figure S7: Wind field at 500 hPa (upper panel) and 200 hPa (lower panel).
Figure S8: Probability density function distributions of the daily temperature [°C] for the 10 regions investigated in the European domain. Each panel shows the distribution estimated from combining all available data for the years 2000-2004 for RegCM5 CP (orange), RegCM5 12km (red), ERA5 (blue) and observations (black). Details about the observational datasets for each region can be found in table 2.
Figure S9: Precipitation intensity, wet frequency and P99.9 seasonal bias for hourly REGCM5 CP (panel a) and REGCM5 12km (panel b) versus high resolution observations. The seasonal biases are the same as in Figure 10, but using 0.1 mm/hr as threshold for the minimum precipitation in the REGCM5 simulations.
Figure S10: Diurnal cycles for mean precipitation (first column), precipitation intensity (second column), precipitation frequency (third column), p99 (fourth column) and p99.9 (fifth column) in DJF (panel a), MAM (panel b) and SON (panel c). In each panel, the results are shown for the following 5 regions in Europe: Great Britain (top row), France (second row), Germany (third row), Switzerland (fourth row) and Italy (bottom row).