# REQUEST FOR AN EMI R&D PROJECT 2027–2029

| | |
|---|---|
| **Country/Organisation:** | SPAIN |
| **Principal Investigator[1]:** | Fidel González Rouco (IGEO, CSIC-UCM) |
| **Affiliation:** | IGEO (CSIC-UCM) |
| **Address:** | C/Doctor Severo Ochoa, 7 – planta 4<br>Ciudad Universitaria<br>28040 - Madrid |
| **Other researchers:** | Félix García-Pereira (Universidad Politécnica de Madrid, UPM), Cristina Vegas-Cañas (IGEO, CSIC-UCM), Jorge Navarro Montesinos (CIEMAT), Elena García Bustamante (CIEMAT), Sara Madera (CIEMAT), Luana S. Cardoso (Lisbon University) |
| **Project Title:** | SMileAtYou-CIMAs: *Approaching a "seamless" strategy through improved soil realism and convection-permitting simulations over Iberian Mountain Areas*. |

To make changes to an existing project please submit an amended version of the original form

| | | |
|---|---|---|
| If this is a continuation of an existing project, please state the computer project account assigned previously. | **SP** …………. | |
| Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.) | 2027 | |
| Would you accept support for 1 year only, if necessary? | YES X | NO ☐ |

| **Computer resources required for project year:** | | **2027** | **2028** | **2029** |
|---|---|---|---|---|
| High Performance Computing Facility | [SBU] | 95,500,000 | 95,500,000 | 95,500,000 |
| Graphics Processing Unit Cluster-A | [GBU] | | | |
| Graphics Processing Unit Cluster-B | [GBU] | | | |

[1] The Principal Investigator will act as contact person for this EMI R&D Project and, in particular, will be asked to register the project, provide annual progress reports of the project's activities, etc.

[2] These figures refer to data archived in ECFS and MARS. If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year etc.

[3]The number of vGPU is referred to the equivalent number of virtualized vGPUs with 8GB memory.

| | | | | |
|---|---|---|---|---|
| Accumulated data storage (total archive volume)[2] | [GB] | | | |

| EWC resources required for project year: | | 2027 | 2028 | 2029 |
|---|---|---|---|---|
| Number of vCPUs | [#] | | | |
| Total memory | [GB] | 350,000 | 350,000 | 350,000 |
| Storage | [GB] | | | |
| Number of vGPUs[3] | [#] | | | |

*Continue overleaf.*

## Extended abstract

*All EMI R&D Project requests should provide an abstract/project description including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used. The completed form should be submitted/uploaded at https://www.ecmwf.int/en/research/special-projects/special-project-application/special-project-request-submission.*

*Following submission by the relevant Member State the EMI R&D Project requests will be published on the ECMWF website and evaluated by ECMWF and its Scientific Advisory Committee. The requests are evaluated based on their scientific and technical quality, and the justification of the resources requested. Previous EMI R&D Project reports and the use of ECMWF software and data infrastructure will also be considered in the evaluation process.*

*Requests exceeding 5,000,000 SBU should be more detailed (3-5 pages).*

### 1. ABSTRACT / PROJECT DESCRIPTION

The SMileAtYou-CIMAs initiative represents a critical advancement in regional climate modeling by integrating two highly complementary research lines: the refinement of physical realism through land-surface processes and the deployment of massive, ultra-high-resolution convection-permitting simulations (CPS) over complex mountainous terrain (Coppola *et al.*, 2020; Lucas-Picher *et al.*, 2021; Giorgi and Prein, 2022).

The project is built on three fundamental pillars that define its scientific value:

1. Climatic Nature: Unlike short-term forecasting, this project will help to generate massive multidecadal simulations (30-year historical periods and future CMIP6 scenarios), which are essential for understanding long-term climate variability and trends.
2. Unprecedented Realism: By utilizing a custom-modified version of the model (WRFdeep), the project incorporates a 1,400 m deep soil component (Steinert *et al.*, 2021). This ensures hydro-thermodynamic consistency between the global driving model (MPI-ESM) and the regional model, allowing for realistic land-air interactions across all relevant time scales.
3. Very high Resolution in Complex Terrain: The project focuses on the Iberian Central System, where elevations reach 2,500 m. Traditional models fail to capture the steep

altitudinal gradients and local phenomena of these areas (Barry, 2002; Giorgi *et al.*, 1996). By reaching kilometric scales (1 km) and resolving convection explicitly, the project provides the high-fidelity data needed for climate change adaptation in sensitive mountain ecosystems (Barry, 2002). Results, in collaboration with AEMET, will contribute to address current gaps in adaptation information for Mountain Areas within National Adaptation Plans in Spain.

## 2. SCIENTIFIC PLAN

### 2.1 BACKGROUND: Accomplishments and Progress to Date

The CIMAs-SMileAtYou project relies on an exceptional observational framework that is unique in its quantity, altitude, and quality, providing a robust foundation for evaluating ultra-high-resolution climate simulations. Central to this framework is the GuMNet (Guadarrama Monitoring Network), which consists of 10 automatic weather stations specifically designed to monitor the atmosphere, surface, and subsurface in the extreme high-altitude environments of the Iberian Central System, reaching elevations up to 2,500 meters.

This network provides rare data on temperature, humidity, wind, radiation, and snow depth, supplemented by specialized soil measurements from boreholes up to 20 meters deep and trenches. The quantity and spatial density of the data are further enhanced by the strategic integration of multiple institutional databases, including AEMET, the SAIH hydrological systems of the Duero and Tagus basins, and Portuguese records from IPMA and SNIRH. This comprehensive cross-border integration ensures a high-density sampling of precipitation and temperature that effectively captures the complex orographic gradients and thermal subsurface regimes of the peninsula (García-Pereira *et al.*, 2024).

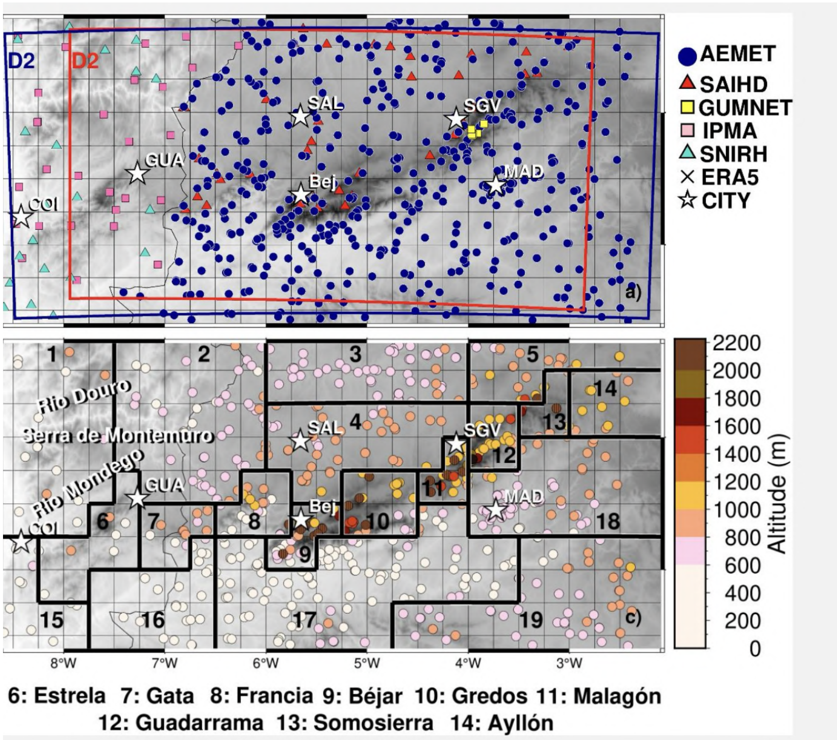

Figure 1. Observational data spatial distribution: institution and altitude. Final figure split into 9 sections.

Unlike standard regional climate approaches like CORDEX (Gutowski Jr. *et al.*, 2016; Solman *et al.*, 2021), this project prioritizes non-interpolated products, preserving the integrity of local station information to avoid the significant uncertainties often introduced by interpolation in rugged, mountainous terrain.

The WRF model (v4.6.1) (Skamarock *et al.*, 2019) is the primary regional climate model employed by the CIMAs project to simulate atmospheric conditions over the Iberian Central System. It is configured using a multi-level nesting strategy that transitions from a 4 km resolution to an ultra-high 1 km resolution domain specifically focused on rugged terrain. This setup incorporates advanced physical schemes, including the Noah-MP land surface model (Niu *et al.*, 2011), Thompson microphysics, and the RRTM radiation scheme. A defining feature of its application in this project is the use of Convection-Permitting (CP) configurations at the 1 km scale, where cumulus parameterizations are disabled to allow the model to resolve convective processes explicitly.

Evaluation results indicate that WRF provides consistent improvements across all temperature ranges compared to the ERA5 reanalysis (Hersbach *et al.*, 2020). It is particularly reliable during the winter season and adds significant value at high-altitude stations by more accurately capturing the steep thermal gradients characteristic of mountain environments.

The HCLIM model is used within the project as a comparative regional climate tool to assess model sensitivity and uncertainty at high resolutions (4 km and 1 km). While HCLIM is instrumental for multi-model evaluation, its performance reveals distinct climatic tendencies compared to WRF and observational data. Analysis shows that HCLIM tends to generate a slightly more extreme climate and exhibits a systematic cold bias in its annual temperature averages. Specifically, the model has been found to underestimate temperatures during the winter period. Despite these biases, HCLIM still shows improved skill in mountain regions compared to the ERA5 reanalysis. However, skill score evaluations (such as the Brier Score) suggest that WRF generally adds more value than HCLIM at high elevations, as HCLIM often exhibits higher levels of temperature overestimation in certain intervals at the 1 km scale.

The transition from a 4 km to a 1 km horizontal resolution represents a strategic trade-off between physical realism and computational complexity. The primary advantage of the 1 km simulation is its superior ability to represent the rugged orography of the Iberian Central System, capturing steep altitudinal gradients that are often smoothed out at coarser resolutions. This increased detail results in a more accurate representation of temperature anomalies at high-altitude stations and the identification of precipitation maximums at peak elevations, which the 4 km resolution tends to underestimate. Furthermore, the 1 km scale allows for Convection-Permitting (CP) setups, where cumulus parametrizations are disabled to let the model resolve convective processes explicitly, leading to a more physically consistent simulation of extreme events.

However, this higher resolution also introduces significant disadvantages and technical challenges. A major drawback identified in the sources is the tendency of the 1 km resolution to overestimate total accumulated precipitation and the number of wet days, often predicting rain on days that observations record as dry. Interestingly, the analysis shows that increasing the resolution does not inherently decrease the Root Mean Square Error (RMSE) in the precipitation field, as the temporal variability of errors is often shared across all resolutions and the driving ERA5 reanalysis. Finally, the computational cost of 1 km simulations is massive; the increase in grid points and the necessity of smaller time steps demand significantly more high-performance computing (HPC) units and vast amounts of data storage compared to the 4 km runs.

The evaluation of simulation performance highlights the significant added value of the high-resolution regional models (WRF and HCLIM) over the ERA5 reanalysis  particularly when resolving the complex terrain of the Iberian Central System. As demonstrated in the analysis of mean temperature skill scores across various altitudinal ranges, there is a clear positive trend in the skill of the regional models relative to elevation.

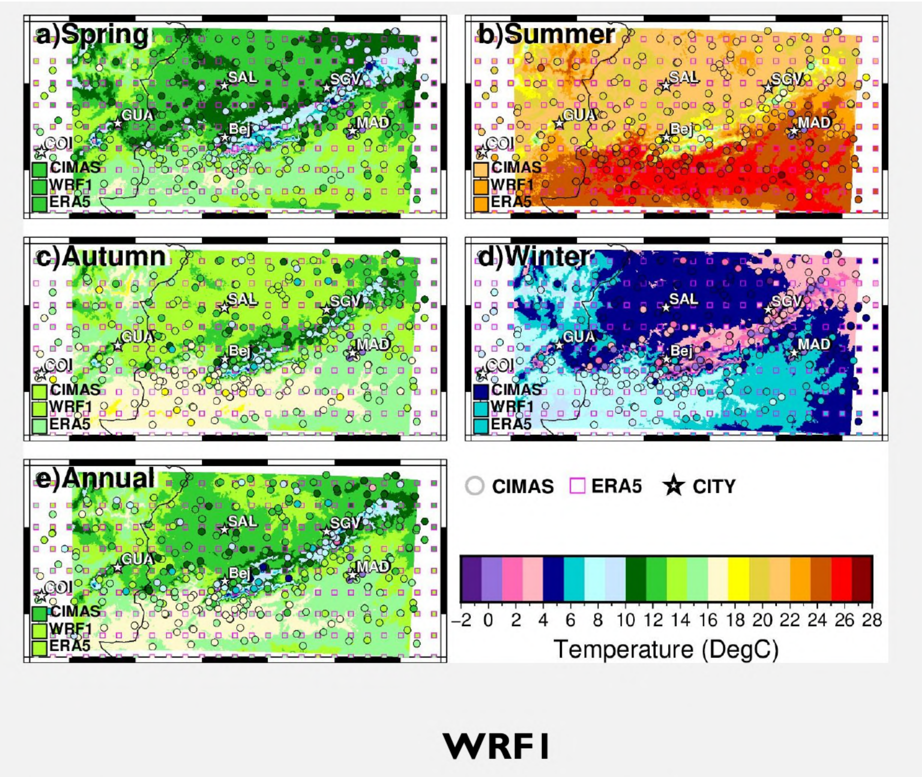

Figure 2. Mean seasonal temperature of WRF at 1 km spatial resoultion  (2010-2014) over the Central System as testing domain for the convection permitting scheme.

While global reanalyses like ERA5 perform adequately in plateau and low-altitude regions, they often struggle to capture the sharp thermal gradients of mountain environments, leading to systematic biases such as a notable warm bias at high altitudes. In contrast, the regional simulations demonstrate a marked improvement in performance as elevation increases; specifically, the WRF1 and HCLIM1 configurations (1 km resolution) show substantial added value over ERA5 at high-altitude stations, with the exception of the summer season. Furthermore, at low and middle altitudes, the regional models exhibit improved skill primarily during winter. Notably, these results are consistent across both 4 km and 1 km horizontal resolutions, reinforcing the conclusion that high-resolution modeling is essential for reducing systematic biases in rugged terrain.

The comparison between the WRF and HCLIM regional models within the framework of the CIMAs project reveals significant differences in their physical behavior and statistical accuracy. While the WRF model shows consistent improvement across all temperature ranges compared to the ERA5 reanalysis the HCLIM model tends to generate a slightly more extreme climate and exhibits a systematic cold bias in its annual averages.

Regarding seasonal variability, WRF is more reliable during winter, a period in which HCLIM tends to underestimate temperatures. However, skill analysis (Brier Score Skill) indicates that HCLIM shows certain improvements in winter compared to other periods, while WRF demonstrates superior performance in summer and spring at low and medium elevations. A fundamental result of the evaluation is that, in high mountain areas, the WRF model provides greater added value than HCLIM for correcting reanalysis biases, establishing itself as the preferred option for capturing thermal gradients in complex orographic terrain.

This evaluation provides a strong scientific justification for the project's transition toward massive convection-permitting simulations, as standard products fail to accurately represent the specific climatic features of high-mountain areas that are most vulnerable to the impacts of global warming.

The combined effort of SMileAtYou and CIMAs has established a robust foundation at the ECMWF facilities:

- Model Customization: The team successfully modified the WRF code to include 4D files for aerosol depth, enabling the simulation of both stratospheric and tropospheric natural/anthropogenic forcings.
- Observational Integration: Through the CIMAs agreement with AEMET, the project has integrated the GuMNet (Guadarrama Monitoring Network) with institutional databases (AEMET, IPMA, SNIRH). This provides a high-altitude benchmarking dataset for temperature, precipitation, wind, and snow depth that is unique in the Iberian Peninsula.

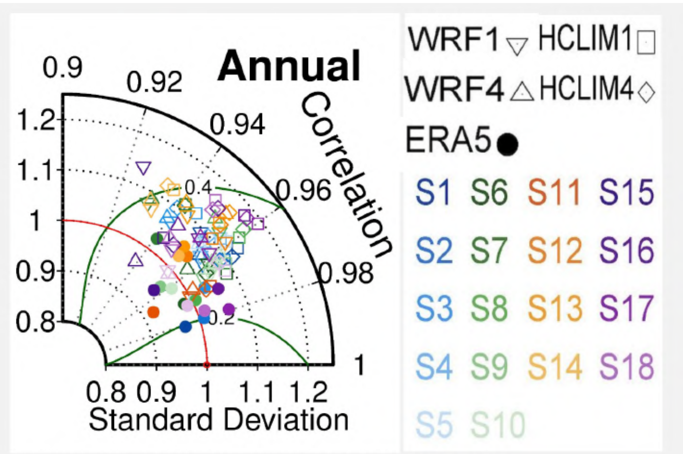

Figure 3. Taylor diagram of mean annual temperature across 19 sections.

- Initial Convection-Permitting Tests: Sensitivity experiments at 1 km resolution over the Sierra de Guadarrama have demonstrated that increasing resolution significantly improves the representation of altitudinal precipitation gradients (Greciano-Zamorano, 2023a, 2023b). These tests have allowed the team to optimize CPS configurations (e.g., Kain-Fritsch and Grell-Freitas schemes) to reduce systematic biases such as the overestimation of wet days.
- Production Progress: Historical downscaling simulations at 4 km and the production of the 1 km Nest Down simulations, have been completed in the period 2009-2025 and are currently advancing through the observational period.

## 2.2 FUTURE WORK PLAN: 2027–2029 Tasks

The upcoming phase transitions from technical validation to the generation of massive climate datasets:

- 1 km Production (CIMAs): The project will execute full-scale production runs at 1 km resolution for the historical period (1989–2009) and future projections for the Iberian Central System and the wider Peninsula.

## Planned Simulation Sequence and Technical Specifications

The project will execute a series of integrated simulations designed to provide a continuous and high-resolution climatic record for the Iberian Peninsula and its surrounding regions. These runs are categorized by their driving conditions and spatial focus:

### I. ERA5-Driven Historical Reconstructions

These simulations utilize ERA5 reanalysis (Hersbach *et al.*, 2020) as boundary conditions to provide the most accurate reconstruction of the recent past:

- *Regional Scale (Europe & Iberian Peninsula):* A nested two-domain configuration with 12 km resolution (European domain) and 4 km resolution (Iberian Peninsula domain), covering the full historical period from 1989 to 2009.
- *Local High-Resolution (Central System):* A single-domain simulation focused on the Central System of Spain at an ultra-high 1 km resolution, covering the period from 1989 to 2025.
- *Regional High-Resolution (Iberian Peninsula):* A single-domain simulation covering the entire Iberian Peninsula at 1 km resolution, spanning the historical period from 2009 to 2025.

### II. MPI-ESM Coupled Climate Simulations

These simulations utilize the coupled WRF-MPI framework (including the deep-soil configuration) for both historical validation and future projections:

- *Regional Historical (Europe & Iberian Peninsula):* A two-domain setup (12 km Europe / 4 km Iberian Peninsula) covering the historical window from 1989 to 2009.
- *Regional Future Projections (Europe & Iberian Peninsula):* A two-domain setup (12 km Europe / 4 km Iberian Peninsula) following the SSP3-7.0 scenario (and alternative scenarios to be determined) from 2015 onwards.
- *Local Historical Convection-Permitting (Central System):* A high-resolution 1 km domain for the Central System, covering the 2009–2015 historical period.
- *Local Future Convection-Permitting (Central System):* A high-resolution 1 km domain for the Central System following the SSP3-7.0 scenario from 2015 onwards.
- *Regional Historical Convection-Permitting (Iberian Peninsula):* An ultra-high resolution 1 km simulation for the full Iberian Peninsula during the 2009–2015 period.
- *Regional Future Convection-Permitting (Iberian Peninsula):* An ultra-high resolution 1 km simulation for the full Iberian Peninsula under the SSP3-7.0 scenario from 2015 onwards.

### III. Additional Scenario Explorations

- *Multi-Scenario Sensitivity Runs:* The project will initiate additional coupled simulations for the Europe/Iberian (4 km), Central System (1 km), and Iberian Peninsula (1 km) domains exploring climate scenarios alternative to SSP3-7.0 to address uncertainty in future climate risks.

- *ALL-Forced Historical Reconstruction:* Launching a comprehensive simulation (1990–present) incorporating all external forcings (GHG, solar variability, LULC, and 4D aerosols) driven by ERA5 (Hersbach *et al.*, 2020).
- *CMIP6 Future Scenarios:* Downscaling future scenarios (e.g., SSP3-7.0 or others) using the WRFdeep-MPI coupled framework to evaluate how improved soil realism affects climate change signals.
- Hazardous Extreme Events Analysis: Utilizing the Convection-Permitting outputs to investigate the intensity and frequency of extreme events such as floods, windstorms, and droughts in mountain areas (García-Bustamante *et al.*, 2021; Jiménez *et al.*, 2013).
- Impact Assessment for Services: The data will be used to develop indicators for renewable energy resources (wind/solar) (Hahmann *et al.*, 2020), water resource management, and wildfire risk, contributing directly to the Spanish National Plan for Adaptation to Climate Change (PNACC-2) (PNACC2, 2020).
- Comparison with current CMIP/CORDEX and Destination Earth Simulations on Climate Variability and Hazardous Extreme events, exploring the behaviour of the different simulations for accurately reproducing relevant elements of complex orography regions climate.

## 3. JUSTIFICATION OF THE COMPUTER RESOURCES REQUESTED

The demand for resources is driven by the massive nature of the simulations and the move to kilometric scales:

- HPC (95,500,000 SBUs/year): This allocation is necessary to sustain the 1 km Convection-Permitting (CP) setup. Moving from the standard 9 km resolution to 1 km increases the number of grid points by a factor of 81, and the integration time step must be significantly reduced to maintain numerical stability.
- Data Storage (350,000 GB / 350 TB): The high-resolution outputs for multidecadal runs generate enormous datasets. Saving high-frequency (hourly) variables for extreme event analysis and the 4D analysis of aerosol/soil dynamics requires substantial archival volume.
- Technical Complexity: The inclusion of modified code for the external forcings, specialy the aerosol integration, increases the computational cost per grid point compared to standard WRF configurations.

## 4. TECHNICAL CHARACTERISTICS OF THE CODE

The project relies on a highly customized and optimized version of the Weather Research and Forecasting (WRF) model (Skamarock *et al.*, 2019):

- WRFdeep Architecture: Modified to include a deep land module for a more realistic distribution of soil temperature and moisture, ensuring consistency with the driving MPI-ESM.
- Nesting Strategy: A three-level nesting approach (9 km -> 3 km -> 1 km). In the 1 km domain, the cumulus parametrization is disabled, allowing the model to resolve convective processes explicitly (Convection Permitting).
- Physics:

    - *LSM:* Noah-MP with deep-soil configuration (Niu *et al.*, 2011).
    - *Microphysics:* Thompson *et al.* (2008) scheme.
    - *PBL:* Improved Mellor-Yamada Nakanishi and Niino Level 2.5.
- Parallelization: The code is optimized for high-parallelization on the ECMWF HPC facility, utilizing typically 512+ cores and I/O quilting (dedicated cores for data writing) to handle the massive data volumes produced by 1 km runs.

## REFERENCES


Barry, R.G. (2002) 'Past and potential future changes in mountain environments: a review', en *Mountain Environments in Changing Climates*. Londres: Routledge.

Coppola, E. *et al.* (2020) 'A first-of-its-kind multi-model convection permitting ensemble for investigating next-generation regional climate projections', *Climate Dynamics*, 55, pp. 3–16.

García-Bustamante, E. *et al.* (2021) 'Expected recurrence of extreme winds in the northwestern Sahara', *Energies*, 14(21), p. 7091.

García-Pereira, F. *et al.* (2024) 'Thermodynamic and hydrological drivers of the subsurface thermal regime in central Spain', *SOIL*, 10(1), pp. 125–143.

Giorgi, F. *et al.* (1996) 'Elevation dependency of the surface climate change signal: a model study', *Journal of Climate*, 9(11), pp. 2874–2901.

Giorgi, F. y Prein, A.F. (2022) 'Populated regional climate models with convection-permitting resolution', *PLOS Climate*, 1(1), p. e0000013.

Greciano-Zamorano, E. (2023a) *Evaluation of the precipitation variability over the Sierra de Guadarrama*. Tesis de Máster inédita. Universidad Complutense de Madrid (UCM).

Greciano-Zamorano, E. (2023b) *Analysis of extreme weather variables over the Central System*. Tesis de Máster inédita. Universidad Complutense de Madrid (UCM).

Gutowski Jr., W.J. *et al.* (2016) 'WCRP Coordinated Regional Climate Downscaling Experiment (CORDEX): a sentiment for regional climate science', *Geoscientific Model Development*, 9(10), pp. 3421–3431.

Hahmann, A.N. *et al.* (2020) 'The making of the New European Wind Atlas – Part 1: Model climate', *Geoscientific Model Development*, 13(10), pp. 4303–4319.

Hersbach, H. *et al.* (2020) 'The ERA5 global reanalysis', *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999–2049.

Jiménez, P.A. *et al.* (2013) 'Analysis of the long-term surface wind speed variability over the Iberian Peninsula', *Climate Dynamics*, 40(7), pp. 1711–1723.

Lucas-Picher, P. *et al.* (2021) 'Convection-permitting modeling with regional climate models', *WIRES Climate Change*, 12(5), p. e731.

Niu, G.Y. *et al.* (2011) 'The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements', *Journal of Geophysical Research: Atmospheres*, 116(D12).

PNACC2 (2020) *Plan Nacional de Adaptación al Cambio Climático 2021-2030*. Madrid: Ministerio para la Transición Ecológica y el Reto Demográfico.

Skamarock, W.C. *et al.* (2019) *A Description of the Advanced Research WRF Version 4*. NCAR Technical Note NCAR/TN-556+STR. Boulder, Colorado: National Center for Atmospheric Research.

Solman, S. *et al.* (2021) *The future scientific challenges for CORDEX*. CORDEX White Paper.

Steinert, N.J. *et al.* (2021) ‘Increasing the depth of a land surface model. Part I: Impacts on subsurface variability and vertical heat fluxes’, *Journal of Hydrometeorology*, 22(12), pp. 3191–3209.