Plan Overview

A Data Management Plan created using DMPonline

Title: High-resolution modelling of atmospheric flows over complex terrain

Creator: Isadora Limas Coimbra

Principal Investigator: José Palma

Contributor: Jesus Monteiro, Vasco Batista, Isadora Coimbra, João Manuel Viana Parente Lopes

Affiliation: University of Porto (Universidade do Porto)

Template: DCC Template

ORCID iD: 0000-0002-5223-0057

Project abstract:

Atmospheric flows are characterised by a daily cycle, a highly unsteady complex interaction among, for instance, topography, flow separation, forest drag, and thermal stratification. These phenomena have been captured extensively in the Perdigão-2017 measurement campaign, whose dataset is an important validation tool for numerical weather prediction and fluid dynamics codes aiming to model the wind flow. The main objective is to model the flow's micro-scale behaviour during the 46 days of the Intensive Operation Period (IOP) of Perdigão-2017. Results from the Weather Research and Forecasting (WRF), obtained in a previous CPCA project, will be supplied to VENTOS/M, an atmospheric flow simulation code featuring the appropriate physics modelling capabilities for microscale level of interest to wind energy. The second objective is to develop and evaluate new methodologies to model the effects of forestry and surface cover on the atmospheric flow. The high spatial and temporal resolution of these simulations will allow the capture of more intricate flow phenomena and require extensive computational resources.

ID: 198793

Start date: 06-12-2024

End date: 31-12-2025

Last modified: 05-03-2026

Grant number / URL: 2024.14673.CPCA.A3

Copyright information:

imply that the creator(s) endorse, or have any relationship to, your project or proposal

High-resolution modelling of atmospheric flows over complex terrain

Data Collection

What data will you collect or create?

The project uses both existing datasets and newly generated data.
Existing datasets used as inputs include:

- Shuttle Radar Topography Mission (SRTM), with approximately 30 m of horizontal resolution, which is available in the following file formats: Digital Terrain Elevation Data (DTED®), Band interleaved by line (BIL) and Georeferenced Tagged Image File Format (GeoTIFF).
- CORINE Land Cover 2012 and 2018, which is available both in raster (GeoTiff format) and vector (ESRI and SQLite geodatabase format) products.
- Military charts and Airborne Lidar Survey data for the Perdigão site.
- Perdigão field-campaign measurements, including meteorological tower data and lidar measurements, used for model validation. These are storage in NetCDF format.

These datasets are open access, with no constraints on their use.
New data created by the project include:

- VENTOS/M and VENTOS/LES simulation outputs in .nc (Network Common Data Form - NetCDF), including embedded metadata and variables (see the “Documentation and metadata” section). These contain time-dependent wind-flow fields, stored as 5-minute averages. NetCDF is an open, machine-independent format that stores file information, metadata and variables, and is widely used in the scientific community.
- Model configuration and setup files, including simulation parameters, run scripts, and supporting files required to reproduce the simulations.

How will the data be collected or created?

Data will be generated by numerical simulations, mainly VENTOS/M (for uRANS simulations) and VENTOS/LES (for large-eddy simulations). Simulations will be executed on HPC using MPI domain decomposition, and intermediate restart/checkpoint files will be produced to manage wall-time limits and ensure safe continuation of long runs.

To support a consistent FAIR workflow, the created data will be organised using a clear and reproducible folder structure, separating:

- raw simulation outputs (original model results),
- post-processed datasets (derived products used for analysis/validation),
- run configuration and scripts (inputs required to reproduce each simulation),
- key documents and archives (notes, logs, and supporting material).

Documentation and Metadata

What documentation and metadata will accompany the data?

Simulation outputs will be stored in NetCDF (.nc) format, with metadata embedded in the files (e.g., variable names and units, time information, grid/domain details).
To ensure traceability and reuse, the dataset will also include:

- A clear file naming convention and folder structure (e.g., organised by model, case study, and simulation period).
- README.txt files at the top level and within key folders, describing the directory structure, file contents, and how each dataset was produced.
- Run documentation (e.g., main configuration files, scripts, and selected logs) to enable reproducibility.

Ethics and Legal Compliance

How will you manage any ethical issues?

No major ethical issues are expected, as the project data are either generated by the project team (numerical simulations) or obtained from publicly available datasets, which will be properly referenced and cited.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

The contents of the project will be the property of FEUP (Faculdade de Engenharia da Universidade do Porto).

Storage and Backup

How will the data be stored and backed up during the research?

Data will be stored in the local project storage after simulations and post-processing.
Backups will be performed weekly: project data will be transferred via SSH to a local NAS. The NAS uses RAID 6 to provide disk-level redundancy and protect against hardware failures.
When relevant, curated datasets (e.g., subsets supporting publications and validation products) will also be transferred to the Perdigão storage/repository for dissemination, in accordance with applicable access rules.

How will you manage access and security?

Access and security will be managed through institutional authentication and role-based permissions. Project data will be stored in restricted project directories on the local storage, with access granted only to authorised team members. Data transfers (e.g., to the NAS or repository) will be performed via secure protocols (SSH/SCP/RSYNC over SSH).

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

The main long-term value data to be retained and preserved are:

- Final simulation datasets, including the key NetCDF output files needed to reproduce analyses and support publications.
- Model configuration and run files (input parameters, run scripts, coupling settings, and essential logs) required to ensure reproducibility.
- Post-processed/validation-ready datasets, such as collocated outputs at measurement locations, extracted time series and vertical profiles, and derived metrics/figures used in reports and publications.

Documentation and metadata (READMEs, workflow descriptions, and version information).

What is the long-term preservation plan for the dataset?

Long-term preservation will rely on curated archiving, redundant storage, and clear documentation:

- The data catalogue and documentation (READMEs, folder structure, metadata descriptions, processing scripts, and configuration files) will be stored on a local NAS (for version control and long-term traceability).
- The large simulation outputs (NetCDF files) will be preserved on the local NAS and additionally backed up to external offline hard disks to provide an extra protection layer (offline copy).
- Project documents (reports and supporting material) will be preserved on the local NAS and in Overleaf (for the LaTeX sources of manuscripts/theses).

Where applicable, curated subsets supporting publications (and any shareable derived products) will also be deposited in the Perdigão repository (or an institutional repository), in line with access rules and any restrictions on third-party data.

Data Sharing

How will you share the data?

The data results, analysis, and simulations will be shared through journal articles, presentations, and reports.

Are any restrictions on data sharing required?

For the time being, data restrictions are not clear.

Responsibilities and Resources

Who will be responsible for data management?

The Principal Investigator (PI) will be responsible for overall data management, including defining the storage structure, access rules, backup strategy, and decisions on long-term preservation and sharing. In addition, all researchers involved in the project will follow this DMP, ensuring clear documentation of the adopted procedures.

What resources will you require to deliver your plan?

To deliver this DMP, the project will require:

- Storage and backup capacity: sufficient space for raw and processed simulation outputs, intermediate files (restart/checkpoints), and curated subsets (multi-TB scale), including local project storage for long-term holding.
- Data transfer capability: secure transfer tools (SSH/SCP/rsync) and adequate network bandwidth for regular backups and transfers to local storage.
- Human time: staff time to curate folders, maintain READMEs/metadata, track versions, prepare validation-ready subsets, and prepared data supporting publications.