Study and implementation of mesoscale weather forecasting models in the wind industry

Bénédicte Jourdier
Abstract

As the wind industry is developing, it is asking for more reliable short-term wind forecasts to better manage the wind farms’ operations and electricity production. Developing new wind farms also requires correct assessments of the long-term wind potentials to decide whether to install a wind farm at a specific location. This thesis is studying a new generation of numerical weather forecasting models, named mesoscale models, to see how they could answer those needs. It is held at the company Maïa Eolis which operates several wind farms in France. A mesoscale model, the Weather Research and Forecasting model (WRF), was chosen and used to generate high resolution forecasts based on lower resolution forecasts from NCEP’s Global Forecasting System.

The stages for implementation of daily forecasts for the company’s wind farms were: explore and configure the model, automate the runs, develop post-processing tools and forecasts visualization software which was intended to be used by the management team. WRF was also used to downscale wind archives of NCEP’s Final Analysis and determine the possibility to use these in assessing wind potentials. Finally the precision of the model in both cases and for each wind farm was assessed by comparing attained data from the model with real power production.
Table of Contents

Abstract ........................................................................................................................................................................... 2
Acknowledgements ....................................................................................................................................................... 4
1 Introduction .......................................................................................................................................................... 5
2 Numerical Weather Prediction ........................................................................................................................... 6
   2.1 Mesoscale models........................................................................................................................................ 6
   2.2 Microscale wind.......................................................................................................................................... 7
3 Objectives .............................................................................................................................................................. 8
4 The Weather Research and Forecast Modeling System (WRF).................................................................... 9
   4.1 Programs composing WRF ..................................................................................................................... 9
   4.2 Input data .............................................................................................................................................. 11
   4.3 Output data ........................................................................................................................................... 11
   4.4 User-defined parameters ...................................................................................................................... 12
      4.4.1 Horizontal grid ........................................................................................................................ 12
      4.4.2 Vertical “eta” coordinates ...................................................................................................... 13
      4.4.3 Physics choices ........................................................................................................................ 14
5 Implementation of daily operational forecasts ..............................................................................................15
   5.1 Parameterization of WRF .................................................................................................................... 15
      5.1.1 Computation issues ................................................................................................................. 16
      5.1.2 Vertical profile ......................................................................................................................... 16
      5.1.3 Physics ...................................................................................................................................... 16
   5.2 A script for daily runs with GFS ........................................................................................................... 17
   5.3 Post-processing the output files .......................................................................................................... 18
   5.4 Forecasts visualization .......................................................................................................................... 19
      5.4.1 Graphics with NCL ................................................................................................................ 19
      5.4.2 Graphical Interface ................................................................................................................. 20
6 Long-term wind potentials ................................................................................................................................21
   6.1 Parameterization of WRF .................................................................................................................... 21
   6.2 Realizations .......................................................................................................................................... 22
7 Evaluation ............................................................................................................................................................23
   7.1 Methodology ....................................................................................................................................... 23
   7.2 Results ................................................................................................................................................ 24
8 Conclusion ...........................................................................................................................................................28
List of acronyms ..........................................................................................................................................................29
List of figures ...............................................................................................................................................................30
Acknowledgements

This thesis was held at the company Maïa Eolis between September 2011 and February 2012. I wish to thank Nicolas Girard and Olivier Texier for giving me the opportunity to study such an interesting subject at Maïa Eolis, in the heart of the wind industry. I also thank the people I’ve worked with: Olivier, Stéphanie, Sophie, Julien and the rest of the Lyon’s team for their dynamism and joviality during those six months.

I also thank Claire Vincent at Risø Laboratory for kindly answering some of my questions about WRF; and my KTH supervisor, Thomas Nordgreen, for his advice.
1 Introduction

The European Union has targeted a share of 20% renewable energy in its final energy consumption in 2020. The development of wind energy which nowadays is a mature technology is planned to take a large part in achieving that goal. But wind energy has a major shortcoming in the random nature of its production, solely based on the wind conditions.

Predicting wind is an important issue for managing the electricity production of a wind farm and integrating it in the electricity network. It is also important to have access to reliable wind forecasts for planning the budget, to look out for potentially dangerous situations with very high wind or to plan the semiyearly maintenance operations on days without wind to avoid losing production.

A second issue for the wind industry is the prospection of new sites suitable to build wind farms on. A profitability study requires a precise estimate of the future energy production. The usual way for estimating the wind potential is in-site measurements from a mast. The mast has to be installed and to measure during at least one year; this is thus a long process. Moreover those masts are expensive and the measuring instruments are neither very precise nor very reliable since they are subject to faults, freezing, possibly battery problems, and the measures are correlated with near-by weather stations to estimate the long-term potential but this introduces errors.

Therefore there is a need for forecasting short-term and long-term wind with more accuracy and new numerical methods may be a solution. Numerical weather prediction is continuously improving along with the development of computers, and new models (named mesoscale models) have appeared recently. As the name mesoscale refers to, those models deal with distances from some kilometers to some hundreds kilometers, less than usual global weather prediction models. Using a limited domain, that may be defined by the user, those models offer high resolution for a specific location and the lower demand for computing power allows small companies or individuals to use them.

One company operating in the wind industry, interested in those models is the French company Maïa Eolis, this is also the location where this thesis is performed. Today Maïa Eolis operates 99 wind turbines for a total installed capacity of 199 MW divided in 18 wind farms in North and North-East France and several other projects are under development throughout France.

The company is mainly divided in four poles:

- Development teams, who prospect new sites, study the projects feasibility and deal with the long administrative processes to obtain the construction permits.
- Construction teams, who deal with everything related to the installation of the wind turbines.
- Management and maintenance teams for the operational farms.
- And expertise teams, who assist the other teams in scientific matters, especially concerning electricity network and wind (measurements in prospecting sites, wind modeling, sitting optimization, production forecasts and monitoring etc.).

This thesis is a research study in the wind expertise pole. The aim of the thesis is to study the mesoscale models to internally develop short-term wind forecasts to be used by the management team and long-term potential predictions for use in the development process.
2 Numerical Weather Prediction

The basic principles of numerical weather prediction (NWP) were defined a century ago (Bjerknes, 1904) but there was no computer to implement them at the time. Numerical predictions started only in the 50s with the first computers. As computing power has reached higher capacities more complex and higher resolution NWP models have been developed. Still wind forecasts from the common large models are not precise enough for the wind industry. New methods for downscaling the wind to regional and then local scales have appeared.

The NWP models are computer programs that solve the fluid dynamics equations which are ruling the atmosphere: conservation of mass, momentum and energy in a three-dimensional grid. A model also includes many parameterization schemes that take into account sub-grid physical processes that cannot be resolved by the numerical model, for example the formation of raindrops (Stensrud, 2007).

An initial state of the atmosphere is defined with aid of observations from land stations, measures from ships, balloons and other equipments, and nowadays also satellites. The variables e.g. temperature, pressure, wind speed are moved forward in time step by step by the model. The initial state is not exact which makes long-range predictions impossible even if the model were perfect. This is because for nonperiodic flow systems such as the atmosphere, slightly different initial conditions make the system evolve towards completely different states (Lorenz, 1963).

Several global weather models exist. They require supercomputers, thus they are only run by national or supranational meteorological organizations, such as the European Centre for Medium-Range Weather Forecasts (ECMWF), the UK Met Office, the French weather service Météo-France or the US National Oceanic and Atmospheric Administration (NOAA). Nowadays the best resolutions of operational models are about 0.5° (i.e. approx. 40 km).

2.1 Mesoscale models

A mesoscale model is quite similar to a global model but is generally limited to an area of some hundred square kilometers. Therefore the precision can be increased, without demanding too much computing time; the horizontal resolution is some kilometers (or even below a kilometer). The initial and boundary conditions necessary for input to the mesoscale model are given by a global NWP model.

Many mesoscale models have been developed in laboratories around the world (Giebel, 2006). Some important examples are the following:

- Meso NH, developed by two laboratories of the French scientific research center (CNRS), including the weather service Météo-France. It has a high resolution over France, but is not freely available.
- MM5, developed by the National Center for Atmospheric Research (NCAR) and Penn State University, whose development and support have been stopped for about 6 years (because MM5 was merged with WRF; see below).
- The Weather Research and Forecasting (WRF) model, developed jointly by NCAR, NCEP (National Centers for Environmental Predictions) and several other agencies and laboratories. It is freely available online and is used world-wide by scientists as well as
companies and individuals. With the merging of MM5 with WRF, there are now 2 versions of WRF:

- WRF-NMM (Nonhydrostatic Mesoscale Model), mainly used for operational weather forecasting.
- WRF-ARW (Advanced-Research WRF), more complex so slower to run; it is aimed for atmospheric research.

- The Karlsruhe Atmospheric Mesoscale Model (KAMM), developed by Karlsruhe University (Germany) and used at Risø Laboratory but not available online.
- The Regional Atmospheric Modeling System (RAMS), developed originally by Colorado State University and Mission Research Corporation and now by the company ATMET. It is freely available online.

Thus two freely available and interesting models were found: WRF and RAMS. WRF is more widely used and maybe more up-to-date (new releases at least every year); there are much more literature about WRF than about RAMS. It is also maybe a little more precise, for example Chan (2009) or Steeneveld (2011) found satisfactory results with both models but little better with WRF.

Therefore it was decided to install and handle WRF first but unfortunately there was no time to try and use RAMS afterwards.

### 2.2 Microscale wind

The resolution of a mesoscale model is limited to some kilometers so the model may not accurately predict the wind at a turbine location (at a microscale level, i.e. below a kilometer). Different kinds of models are available to locally downscale the wind. Two such models are described here: linear models and CFD (Computational Fluid Dynamics) models (Girard, 2006).

- **Linear models**

  Linear models interpolate wind data between close locations. The calculations are very rapid but they do not take into account any non-linear effects. This is not satisfactory in complex terrain.

  A linear model commonly used in the wind industry is WAsP (Wind Atlas Analysis and Application Program) developed by Risø Laboratory, Denmark.

  In several studies wind data were downscaled from global to local scales by using first a mesoscale model and then coupling the mesoscale data with WAsP. The KAMM-WAsP method was developed at Risø Laboratory (Frank, Rathmann, Mortensen and Landberg, 2001); other used MM5 or WRF (Berge, Bredesen and Mollestad, 2007).

- **CFD models**

  CFD models solve very precisely the fluid dynamics equations in 3 dimensions. They can deal with non-linear effects and possibly thermal stratification, radiations... The issue is that the calculations are very complicated and take long time.


3 Objectives

This study aims to use mesoscale models, assess their precision and the possibility to use them at Maïa Eolis for two final objectives:

1. attain operational wind forecasts to better manage the existing wind farms owned by the company
2. predict wind potentials to evaluate new sites under prospection.

In order to achieve the objectives, the study will involve the following steps:

- Installing, handling the mesoscale model, adjusting the parameters.
- Running mesoscale wind forecasts driven by two kind of global forecasts (in relation of the two objectives):
  1. Daily operational forecasts
  2. Archives of forecasts over several years
- Using the resulting data in possibly three different ways:
  - directly,
  - coupling with WAsP,
  - coupling with Météodyn WT.
- Assessing the results by comparing with wind measurements from Maïa Eolis’ measure masts and operating wind turbines.

This process is not linear but interactive since results from a step can be used to improve previous ones.
4 The Weather Research and Forecast Modeling System (WRF)

A major part of present study was to understand the model, install it and to create tools to handle it in the easiest way possible. For this purpose, much help was provided by the User's Guide (NCAR, 2012), by the online tutorial (UCAR, 2011) and the technical notes that provided more details about the internal equations of the model (Skamarock et al, 2008).

In this part we describe WRF’s chain of programs, the input and output data along with their specific formats and some of the parameters entered by the user are explained.

4.1 Programs composing WRF

WRF is not one program but a set of programs: one core program is computing the simulation itself and there are many other programs that may be used to process the input or output data.

There are 2 versions of the core program: NMM and ARW. The first decision was to use NMM because it is lighter and supposedly faster for operational forecasts. However a bug in Ubuntu made it impossible to work with the distributed memory option, which enables the program to be operated on the 8 cores of the computer’s processors. Therefore WRF was installed once again with the ARW core which could be operated with the 8 cores instead of only one.

Among all the other programs, only the basic ones are used in this study. Those are the three programs of the pre-processing system, the initialization program and the common post-processor.

A brief description of the program chain and a flowchart are presented below:

- **Geogrid** defines the grid according to the parameters given by the user and extracts the needed geographic data
- **Ungrib** extracts the meteorological fields from the global forecasts’ files for the run dates
- **Metgrid** interpolates the meteorological fields from ungrib on the domain from geogrid
- **Real** creates a file with the initial conditions and another with the boundary conditions
- **WRF-ARW** is the model core, it uses the physics and dynamics options specified by the user

- **NCL** (NCAR Command Language) is not included in WRF but is software developed by NCAR as well, and useful to extract and/or plot data with the WRF output files which are in the netCDF format.
Figure 1: Flowchart of the WRF-ARW programs as they are used in the study.
4.2 Input data

The input data needed for WRF are of two types:

- Static geographic data from the USGS (United State Geological Survey) stored in a database and downloaded once with WRF: soil altitude, soil type, roughness and other information. There are several possible resolutions; the best is 30” arc.

- Meteorological data from a global NWP model for the whole time period of the WRF simulation (they are used as boundary conditions). Two kinds of data from the NOAA global model GFS (Global Forecasting System) can be downloaded freely:
  - GFS data with a 0.5°-resolution (NCEP, updated daily a). Those are operational forecasts commonly used worldwide. Each run is giving forecasts for the upcoming 8 days. A new run is issued every six hours. The runs are available for downloading during one day and are not archived. You can use one GFS run as input for WRF to create an 8-day simulation at a higher resolution.
  - FNL (Final) data with a 1°-resolution (NCEP, updated daily b). Those are analyses, this means that observational data are assimilated by a model and projected on the 1°-resolution grid. The data assimilation model is the same than the one used to create the initial state for the GFS model (the first time (0-hour) of each forecasts run) but the FNL data are created a little later, with more observational data. The archives of these analyses are available from 1999 so that a range of archived data may be used as input in WRF.

Those global forecasts used as input files are in GRIB format. This is a standard format for meteorological data, in a binary form, defined by the World Meteorological Organization (WMO, 1994). More information about this format and how to handle it are given in annex A.

4.3 Output data

The WRF-ARW output files – “wrfout” files – contain more than a hundred variables, including wind components, temperature, pressure, moisture, rain, cloud cover etc. and most variables have 3 or 4 dimensions:

- 2 horizontal dimensions covering the grid defined by the user;
- Possibly a vertical dimension, either on all the vertical levels in the atmosphere (there can be several dozens of levels) or on all the soil levels (probably 4 or 5 levels);
- And a time dimension (the time interval between the outputs is defined by the user, for example 10 or 60 minutes).

The files are in netCDF (network Common Data Form). This is a machine-independent format for sharing array-oriented scientific data. It is self-describing, which means that the file contains not only the data but also information about it. It is a convenient format to handle many multi-dimensional fields. More details about this format and how to handle it are given in annex A.
4.4 User-defined parameters

Parameters are written in two text files named “namelist.wps” and “namelist.input” and are read respectively by the preprocessing programs and the “real” program. Apart from the simulation start and end dates and information about the input files, important parameters are the ones that define the domain, and the choices made among the available physics schemes.

4.4.1 Horizontal grid

In order to establish a horizontal grid the user defines a limited domain centered on a chosen latitude-longitude location by giving the number of points in the west-east and the south-north directions, the resolution (which is, the distance between the grid points), and the type of map projection used (presently Lambert Conformal recommended for mid-latitudes).

The resolution (or grid spacing) is an important parameter that has large influence on the computing time. Indeed, dividing the grid spacing by two implies four times more cells but also divides the simulation time-step by two, which in the end multiplies the computing time by about 8.

A commonly used option to decrease the computing time is the nesting: smaller domains of higher resolutions are included inside the previous ones, like Russian dolls. It enables to get a higher resolution in just one or some parts of a larger grid. Below is an example of the horizontal grid cells of a main domain (dark blue) and a nest (light blue) with a ratio of 3 (most common ratio).

![Figure 2: Scheme of a 10-by-9-point horizontal grid with a 13-by-10-point nest (ratio of 3 between the main domain and the nest grid spacing)](image)

This grid is just an example illustrating the concept of nesting but it would not be suitable for any simulation because there are not enough grid points.

First, it could be problematic to properly interpolate the input data from a low-resolution model (for example if all the grid points of the domain correspond to just one or a few grid points of the large NWP model).
Second, Skamarock (2004) showed that the kinetic energy spectrum of a mesoscale model with a grid spacing “dx” is well matching reality over wavelengths of 7 dx but rapidly deteriorating below this limit. This means that the features smaller than about 6-7 times the grid spacing (called the “effective resolution”) are not resolved properly. So a 10-point grid is too small compared to this effective resolution. Consequently, a balance must be found between grid size, resolution and computing time.

4.4.2 Vertical “eta” coordinates

WRF uses a special kind of vertical levels, named eta-levels, to vertically divide the atmosphere into computation cells. Those levels are neither altitude-fixed nor pressure-fixed.

The vertical coordinate $\eta$ is pressure-dependent. It is always 1 at the earth’s surface (it follows the topography) and decreases towards 0 at the top of the atmosphere according to the equation:

$$\eta = \frac{p - p_{top}}{p_{surf} - p_{top}}$$

Where $p$ is the hydrostatic component of the pressure, varying between:

- $p_{surf}$ at the surface of the earth and
- $p_{top}$ at the top of the atmosphere (a constant pressure fixed by the model, typically 50 hPa, which corresponds to a 20 km altitude)

![Figure 3: Non-scaled scheme of eta-levels over a mountain](image)

If the user only defines the number of levels, a default set of eta-coordinates will be provided by the “real” program. But the user may also give explicit values for the set of eta-levels which will be used. It is to be noted that these values given to WRF will correspond to the levels dividing the computation cells (i.e. the cells’ bottom and top boundaries). In the output files, most variables will be given, not on these eta-levels but on the cell centers in the middle between these levels.
4.4.3 Physics choices

There are 5 groups of physics schemes in WRF:

- The microphysics scheme explicitly solves the distribution of water in the atmosphere between vapor, clouds, rain, snow…
- The cumulus parameterization adds the sub-grid-scale effects of convective clouds (for domains whose grid spacing is greater than 5 km)
- The planetary boundary layer scheme represents the vertical fluxes from sub-grid-scale turbulence
- The short-wave and long-wave radiation schemes take into account the presence of clouds, carbon dioxide and ozone to calculate the absorption, reflection, emission of radiation, and therefore the heating or cooling of the surface and the atmosphere.
- The surface and land-surface models are the key point, taking into account the information from the other schemes to compute heat and moisture exchanges at the surface and the friction velocity. Some of them take into account some vegetation processes and temperature and moisture profiles in the soil.

For each scheme, WRF allows choices between several models. Only some combinations of choices make sense. For example the 3 schemes dealing with the surface (planetary boundary layer, surface, and land-surface schemes) are closely related, thus they cannot be chosen independently.

Those 3 schemes have a large influence on the surface fields, particularly on the wind in the lower atmosphere which is studied here. Draxl (2010) studied 7 configurations of those 3 parameters and found that 3 out of 7 were modeling the diurnal cycle in atmospheric stability more accurately, thus better modeling the wind shear in the lower atmosphere.

Of course the performance of one configuration is site-related, so a good choice of physics schemes at one place may not be the best one for another place.
5 Implementation of daily operational forecasts

The first objective of this study was to implement daily forecasts for the company’s wind farms to be used by the management and maintenance teams. Therefore the main tasks were to choose adequate parameters for the WRF simulations, then to automate daily runs and finally to exploit the WRF output files.

5.1 Parameterization of WRF

A grid for the WRF simulation was designed to include all the company’s wind farms (cf. black dots in figure 4). The final main domain turned out as a 80-by-60-point grid with 12-km resolution, covering Northern France. It is complemented in its center by a 4-km resolution nest (91 x 73 points), which contains all the wind farms.

Figure 4: Map of the main domain (12-km resolution) and of the inner nest “d02” (4-km resolution). The black dots show the locations of Maïa Eolis’ wind farms.
5.1.1 Computation issues

The grid size and resolution strongly influence the computation time which limited the possible choices of grids. The first intention was to go down to a higher resolution with three levels of nest (18, 6 and 2-km resolutions) which is illustrated in several publications, but the computation time was too important.

The following table gives the computation time and size of the output files for the grid defined in the previous section, with or without using the nest. It can be seen that the simple configuration with one domain and one nest is demanding much time and storage space for only 8 days of simulation. The computation time was the determining factor in the choice of this configuration.

<table>
<thead>
<tr>
<th></th>
<th>Only the main 12 km resolution grid</th>
<th>12 km main domain + 4 km nested domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computation time</td>
<td>2 h</td>
<td>7 h</td>
</tr>
<tr>
<td>(with 8 CPU)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRF output file size</td>
<td>1.6 Gb (1 hour interval)</td>
<td>24 Gb (10 minute interval)</td>
</tr>
</tbody>
</table>

Table 1: Approximate computation time (computer-dependent) and output file size (dependent on some parameters) for an 8-day WRF run on the previously defined domain.

As far as daily forecasts are concerned, a 7-hour run is still possible even though not very operational since you want the forecasts as soon as possible and for future, not past times! But running for years of archived data would be impossible in those conditions.

5.1.2 Vertical profile

In this study, the default set of eta-levels was modified to have more and thinner layers in the lower atmosphere (below ca 200 meters), aiming for a better resolution of the wind profile around the wind turbines heights. The seven first eta-levels in the modified set are spaced with 0.003, so that the six first layers are about 30-meter thick.

5.1.3 Physics

It would have been interesting to compare several sets of parameters, but this would have required issuing months of simulation with several sets and there was neither time nor space to issue and store so many simulations. The set applied in this study was chosen according to the recommendations from the User’s Guide for grids with similar resolution to ours.

Moreover the grid covers a large area, with 18 diverse sites of interest (some close to the coast, others far in-land, with more or less mountainous terrains), so that even if a set of parameters was better for one site, it may not be the best one for the other sites.
5.2 A script for daily runs with GFS

Once the parameters used in WRF were defined, the task was to automate the runs for daily forecasts. A script was written in bash (Ubuntu’s shell) and is launched automatically every night. It goes through the following steps:

1. Downloading GFS files from the last 8-day run
   The date of the last available run is automatically calculated. With the “wget” function, the 65 files of the run are downloaded from NCEP’s ftp (total size: 3.5 Gb). With a system of loops in the script, the files are downloaded in batches and the download status is checked. If there is a problem, if the files are not finished downloading after 8 minutes, the download is stopped and launched again.

2. Preparing to run WRF
   The script creates a new folder and creates the links to the WPS, real and wrf programs, as well as to several tables and files necessary to start WRF.

3. Importing “namelist” files with the chosen parameters and editing their date
   This is done using another script created by the author and named edit_namelist. This script takes as arguments the starting and ending dates and some grid parameters; and its execution creates the adequate namelists.wps and namelist.input files. The main script has just to substitute the dates in a specified edit_namelist (with good grid and physics parameters) and execute it. The script also looks for a restart file and modifies the namelist accordingly. Restart files from a previous simulation enables the model to start the new simulation with already small-scale variance. Otherwise the high resolution physics features are developed in the first 6 to 24 hours of the run and these first hours have to be discarded.

4. Running the geogrid, ungrib, metgrid, real and finally wrf programs
   Each program is launched one after the other.

5. Post-processing the wrfout files
   See next paragraph on post-processing.

6. Finally archiving the files.
   Since the output files are very large, it is required to cut them to keep only some of the variables and the lowest levels in the atmosphere. But this has to be done correctly so that the files can still be read.

For now the script is launched once a day, at 11 pm, so that the forecasts are ready on the following morning. It was thought to perform 2 runs per day but considering the size of the GFS data it is not possible to download them during the day without considerably slowing down the company’s network. A solution might be studied in the future to bypass this problem.
5.3 Post-processing the output files

NCAR Command Language (NCL) is used to extract the desired data from the netCDF WRF output files. This interpreted language with built-in functions enables the user to write scripts that execute a series of actions. In this study the typical steps in a script are to:

1. **Load a netCDF file and extract some variables**

   The wanted variables are typically the horizontal components of wind, the temperature and pressure. Some WRF-specific NCL functions enable to easily retrieve these fields from the WRF output files, even when they need to be un-staggered or rotated from the model grid. Other pieces of information such as rain, moisture, and cloud fraction could also be extracted.

2. **Read a text file with information on the wind farms**

   For each farm, the text file specifies the location (latitude-longitude), the turbines’ type and height.

3. **Extract time series of the previous variables at each location**

   For the specified latitude-longitude locations, each variable is extracted along time and along the lowest vertical levels.

4. **Interpolate the variables to the turbines’ height**

   The wind and other variables are needed at a fixed height above ground and not at a vertical level whose altitude varies in time. The hub height could be different for each location and was read in the text file with information about the farms. A specific output variable from WRF is used: the geopotential (or gravitational potential energy in m².s⁻²) which once divided by the standard gravity at mean sea level (ca 9.806 m.s⁻²) gives the geopotential height. This height is equivalent to the geometric height for altitudes small compared to the Earth radius. Given the terrain elevation, the height above ground of each atmospheric level is calculated. Then the variable at the fixed hub height is linearly interpolated from the values on the two closest levels.

5. **Calculate the power production**

   Given the wind speed at hub height and the turbine’s type (read in the text file), the power in kW is calculated with a formula approximating the turbine’s power curve.

6. **Plot or print these data for further use.**

   There are many possible uses of the extracted data. One of them is to directly plot the variables using NCL’s many graphical functions. Another one is to write the data in a table in CSV format.

   In the daily script, as soon as WRF has finished, two NCL scripts are used, one to create plots of the forecasts in a PDF file, and the other to write the output data in a CSV file.
5.4 Forecasts visualization

The purpose of those forecasts is to be used by the operation and maintenance team. Therefore 2 ways for visualizing the data were developed. The first one uses NCL to create graphs. The second one uses the CSV files from NCL to upload the predictions in a database and use them later.

5.4.1 Graphics with NCL

With one of the developed NCL scripts, the final output file is a PDF file containing plots of the predicted wind speed at hub height, wind direction, turbine output and temperature at hub height (useful for assessing frost risk), with one page per wind farm. An example of a page is given below.

![Example of the output plots of the daily forecasts for one wind farm (made with NCL).](image)

From top to bottom: wind speed at hub height (in m/s); wind direction; power output for one wind turbine (in kW); temperature at hub height (in Celsius).
5.4.2 Graphical Interface

The previous method is useful but not sufficient when it comes to comparing those forecasts with other datasets as: the actual power production once it is recorded, as well as another set of forecasts provided by an external consultant. Therefore, a second visualization tool was developed, using the company’s database and the CSV file created daily by the second NCL script. 10-minute averaged power production data and forecasts data are regularly imported to the database. This CSV file is written so that it is ready for being uploaded to the database.

The GUI tool in Matlab was used to develop stand-alone software that can be used on any PC.

First, this software connects to the company’s database and retrieves the acronyms and number of wind turbines for each wind farm and prepares to plot the data. (Therefore the software doesn’t need to be updated whenever there is a new wind farm).

Then the user may choose either a run from one of the prediction models, or a period of time with the actual power production and then the software retrieves those data from the company’s database and plot them. With a system of checkboxes, the user may choose to make each curve visible or invisible.

![Figure 6: Screenshot of the software. There is one tab per wind farm.](image)

The actual power production from one wind turbine is shown in yellow.

One WRF run is shown in dash line (colors from green (first day of run) to orange (8th day)).
6 Long-term wind potentials

The second objective of this study was to assess the use of a mesoscale model to evaluate wind potential.

6.1 Parameterization of WRF

For this matter, the grid cannot be as precise as it was in the first part because the computing time would be too important. Moreover the meteorological data used as input parameters are the FNL data which have a resolution of 1°, instead of 0.5° for the GFS data previously used. So the use of two levels (36 then 12 km) will downscale the wind step by step. Finally the main domain is a 50-by-40-point grid of 36-km resolution covering France. It is complemented in its center by a 12-km resolution nest (61 x 76 points), which contains all the wind farms (cf. Figure 7).

![Map of the main domain (36-km resolution) and of the inner nest “d02” (12-km resolution) in the case of long-time runs from the FNL archive data. The black dots show the locations of Maia Eolis’ wind farms](image)

Figure 7: Map of the main domain (36-km resolution) and of the inner nest “d02” (12-km resolution) in the case of long-time runs from the FNL archive data. The black dots show the locations of Maia Eolis’ wind farms
The other parameters are almost the same as before: same vertical levels and almost the same physics choices. Only the parameters linked to the grid resolution were changed in comparison to the previous case.

With those choices, the simulation for one month took about 5 hours, which is reasonable.

### 6.2 Realizations

This model was run for 2 months: last December and January so that it could be used in assessing the GFS forecasts too. No other months were run due to a lack of time and lack of storage space for FNL data, for the ongoing simulation and for archiving the simulations’ results.

If simulations could have been run for over a year or more it would have produced wind distributions (i.e. proportion of each wind speed for each wind direction). For each location those long-term distributions could have been used as input for WAsP or Meteodyn to downscale the wind to the scale of the wind farm (microscale level), taking more into account the local topography. The company already has the technology and methodology for doing such a coupling so when the storage issue is taken care of, more runs can be issued easily with help of a script that was written; and then easily coupled.

Unfortunately it could not be done for now so it is impossible to assess the precision of this coupled mesoscale-microscale model and how it could be useful in determining wind potentials. Only the raw mesoscale data could be evaluated along with the GFS-based forecasts, as is described in the following part.
7 Evaluation

A weather forecast model is nothing without a proper assessment of its reliability. If the prediction errors are too large, the forecasts are not useful. After two months of daily runs (or almost), there was enough data to start evaluating the forecasts. Of course these evaluations are site-related.

7.1 Methodology

There are many statistical tools for calculating errors. When it comes to power forecasts, it should take into account that the positive and negative errors do not compensate one another (thus using absolute errors) and that large prediction errors are more problematic than small errors (thus using squared errors).

Therefore the tools used here are the bias (actual production minus prediction), the mean absolute error, and the standard deviation error (mean squared error once the bias is removed). Those tools are described in the annex B. All the errors are normalized by the turbines’ power capacity.

Three different sets of data were evaluated:

- 4-km WRF runs based on 0.5°-GFS forecasts as input: there are 57 runs, each run being an 8-day forecast, and with 10-minute interval output.
- 12-km WRF runs based on 0.5°-GFS forecasts as input: also 57 runs, each run being an 8-day forecast, and with 1-hour interval output.
- 12-km WRF runs with 1°-FNL analysis as input: a long run of 61 days with 1-hour interval output.

Those sets are compared with the 10-minute averaged power production directly transmitted by the wind turbines. Among these measurements:

- The data corresponding to a wind turbine that was stopped for maintenance or another technical reason were removed.
- The data corresponding to a wind turbine that was not completely stopped but not at its full capacity due to some technical reason were not removed. This happened on two wind turbines in two different farms and alters the comparison.

Data from all the wind turbines in a wind farm are averaged to compare to the farm’s predictions.

The errors are calculated for each date in the output forecasts (the intervals are either 10 or 60 minutes). None of the data are smoothed.
7.2 Results

The evolutions of normalized bias, normalized mean absolute error (NMAE) and normalized standard deviation error (NSDE) of the GFS-based runs were plotted versus the forecast horizon. Examples of these graphs are given in Figure 9 (a good case) and Figure 8 (a bad case).

The comparisons of the forecasts with actual power production data show that WRF is almost always over-estimating the wind: the bias is negative for all the studied wind farms. For almost all the wind farms this bias is clearly oscillating with a 1-day period (such as in Figure 9, top panel). This may be due to the physics schemes used which do not well represent the diurnal cycle in the atmosphere stability, which modifies the wind profile (Draxl, 2010).

The main parameter for assessing the results is usually the NMAE. A good model would be expected to have a NMAE at 10% at the first forecast horizons. With the developed model, and for most cases, the NMAE is a little below 20% for the first 4 days and rises in the following days. This is not very good but not bad either. The negative bias and the bad diurnal cycle have not been removed yet. Also the model downscales to the mesoscale – not microscale – level so that some corrections should be made to take into account the local topography and roughness. This will be done by a statistical model developed by the company. But there is not enough data yet to train this model correctly.

In the worst cases, NMAE are more around 30% in the first days, which is too much. These two cases are known for being problematic because they are very close to the shore and the influence of winds over the Channel makes the meteorological conditions very unpredictable there. It is not sure whether those two wind farms will ever be well modeled.

The first comparisons between 4-km and 12-km forecasts runs do not show a big difference in terms of performance. Therefore using more computing time to run at a higher resolution may not be worth it.
Figure 9: Normalized bias, normalized mean absolute error and normalized standard deviation error along the forecast horizon for the GFS runs. The green curves correspond to the run at a 4-km resolution (10-minute intervals between output horizons), the red curves to the 12-km resolution (1-hour interval). Wind farm #17 (one of the best cases)

Figure 8: Idem than Figure 9, for wind farm #8 (one of the worst cases)
Other interesting graphs are the distributions of normalized errors (examples in Figure 10 and Figure 11). They show if the errors are centered on zero and how the errors are grouped or spread which gives more information than the single values of bias, MAE and SDE.

The runs from FNL data may be seen as an optimum. As we see in Figure 10 and Figure 11, they are quite similar to the GFS-based runs for a small horizon forecast (less than a day). Indeed the FNL data are analyses issued by the same data assimilation model than the initial state of the GFS run.

The errors in the FNL-based runs come from the input data (errors in the data assimilation process) and from the WRF model. In the GFS-based runs the errors come from the same reasons (same data assimilation model and WRF) but also from the imprecision in forecasting. Thus the precision declines along time, especially after the 4th day. The following table gives statistical keys for the FNL-based runs performance at each wind farm. This represents the precision of the WRF model (with our parameters) for these locations.

<table>
<thead>
<tr>
<th>Farm</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Mean Absolute Error</th>
<th>Standard Deviation Error</th>
</tr>
</thead>
<tbody>
<tr>
<td># 1 (i)</td>
<td>-6.90%</td>
<td>31.21%</td>
<td>23.13%</td>
<td>31.95%</td>
</tr>
<tr>
<td># 2</td>
<td>-14.74%</td>
<td>17.29%</td>
<td>16.67%</td>
<td>22.71%</td>
</tr>
<tr>
<td># 3</td>
<td>-10.40%</td>
<td>18.84%</td>
<td>16.20%</td>
<td>21.51%</td>
</tr>
<tr>
<td># 4</td>
<td>-7.13%</td>
<td>15.35%</td>
<td>11.70%</td>
<td>16.93%</td>
</tr>
<tr>
<td># 5</td>
<td>-9.95%</td>
<td>16.67%</td>
<td>14.05%</td>
<td>19.41%</td>
</tr>
<tr>
<td># 6</td>
<td>-15.14%</td>
<td>18.19%</td>
<td>17.39%</td>
<td>23.66%</td>
</tr>
<tr>
<td># 7</td>
<td>-9.31%</td>
<td>15.48%</td>
<td>13.02%</td>
<td>18.06%</td>
</tr>
<tr>
<td># 8</td>
<td>-16.23%</td>
<td>28.93%</td>
<td>26.23%</td>
<td>33.16%</td>
</tr>
<tr>
<td># 9</td>
<td>-11.16%</td>
<td>24.62%</td>
<td>20.01%</td>
<td>27.03%</td>
</tr>
<tr>
<td># 10 (i)</td>
<td>-0.83%</td>
<td>34.38%</td>
<td>25.62%</td>
<td>34.38%</td>
</tr>
<tr>
<td># 11</td>
<td>-10.87%</td>
<td>18.38%</td>
<td>14.42%</td>
<td>21.34%</td>
</tr>
<tr>
<td># 12</td>
<td>-4.04%</td>
<td>17.13%</td>
<td>12.79%</td>
<td>17.59%</td>
</tr>
<tr>
<td># 13</td>
<td>-6.09%</td>
<td>17.60%</td>
<td>13.20%</td>
<td>18.62%</td>
</tr>
<tr>
<td># 14</td>
<td>-12.30%</td>
<td>20.11%</td>
<td>17.42%</td>
<td>23.57%</td>
</tr>
<tr>
<td># 15</td>
<td>-9.77%</td>
<td>17.75%</td>
<td>14.12%</td>
<td>20.26%</td>
</tr>
<tr>
<td># 16</td>
<td>-5.57%</td>
<td>17.68%</td>
<td>12.30%</td>
<td>18.53%</td>
</tr>
<tr>
<td># 17</td>
<td>-6.14%</td>
<td>17.81%</td>
<td>12.35%</td>
<td>18.84%</td>
</tr>
<tr>
<td># 18</td>
<td>-6.57%</td>
<td>18.00%</td>
<td>12.57%</td>
<td>19.15%</td>
</tr>
</tbody>
</table>

Table 2: Evaluation of the WRF runs based on FNL archived analysis.
The red lines correspond to the badly represented wind farms.

(i) The data used to evaluate the forecasts at those two wind farms were corrupted because of technical problems on two of the wind turbines. This problem alters the comparisons.
Figure 10: Distribution of normalized errors at the wind farm #17. In grey: 2-month run with FNL data. The histogram is normalized but corresponds to more than 1500 values (62 days with one output per hour). In blue: daily forecasts with GFS data (4-km resolution) and forecast horizon between 0 and 24 hours (day 1), between 24 and 48 hours (day 2), etc. Each histogram has almost 8000 values (144 per day, times 57 runs).

Figure 11: Idem than Figure 10, for wind farm #8.
8 Conclusion

After choosing WRF as the mesoscale model that would be used in the study, the first important task of the thesis was to operate the WRF model. This model is sophisticated and there was sometimes a lack of information, even though the User's Guide is several hundred pages thick. It took a long time to understand it: the installation of the program itself took several weeks at first; and a bug in the model resulted in switching to ARW so that it took more time to install and handle this other version. Certainly these difficulties impacted the time dedicated to actually running and testing the model, with an operational start of the forecasts in late November.

Technical limitations also impacted what could or could not be done. Even though the computer used is more powerful than an average office computer it is still not a super computer so that computing time becomes an issue at high resolution. Moreover there was a lack of computing and storage space which limited the possible number of runs.

Concerning the first objective - operational wind forecasts for the existing wind farms -, forecasts using the WRF mesoscale model were automated and new visualization software was developed. This in order to supply the management and maintenance teams with easy and operative visualizations of weather forecasts. The evaluation of the forecasts showed acceptable results for most of the wind farms. A statistical model will then be used to adjust the WRF output, taking into account the specificities at each site. This coupled WRF + statistical model is expected to show good results.

Concerning the second objective - wind potential to evaluate new sites under prospection -, WRF was parameterized but long simulations were not run due to lack of time and space. Thus it could not be coupled to a microscale model for evaluation of the precision of such a coupling. The first comparison (without coupling) did not show enough precision to let the mesoscale model be the sole mean for evaluation of a site potential. This is because measurements will still be needed to readjust the output of the mesoscale model with a statistical model, with WAsP or with a CFD model. But the mesoscale model could be used to better extrapolate the wind measurements. Nowadays the measurements from a one-year period are correlated with long measurements from a nearby weather station to find the long-term wind potential; but “nearby” sometimes means hundreds of kilometers, which may introduce much error. Correlations with output from a mesoscale model could provide better long-term potentials.
# List of acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFD</td>
<td>Computational Fluid Dynamics</td>
</tr>
<tr>
<td>ECMWF</td>
<td>European Centre for Medium-Range Weather Forecasts</td>
</tr>
<tr>
<td>FNLI</td>
<td>NCEP Final Analysis</td>
</tr>
<tr>
<td>GFS</td>
<td>GFS (Global Forecasting System)</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>KAMM</td>
<td>The Karlsruhe Atmospheric Mesoscale Model</td>
</tr>
<tr>
<td>NCAR</td>
<td>National Center for Atmospheric Research (US)</td>
</tr>
<tr>
<td>NCEP</td>
<td>National Centers for Environmental Prediction (US)</td>
</tr>
<tr>
<td>NCL</td>
<td>NCAR Command Language</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanic and Atmospheric Administration (US)</td>
</tr>
<tr>
<td>NWP</td>
<td>Regional Atmospheric Modeling System</td>
</tr>
<tr>
<td>RAMS</td>
<td>University Corporation for Atmospheric Research</td>
</tr>
<tr>
<td>USGS</td>
<td>United State Geological Survey</td>
</tr>
<tr>
<td>WAsP</td>
<td>Wind Atlas Analysis and Application Program</td>
</tr>
<tr>
<td>WPS</td>
<td>WRF Preprocessing System</td>
</tr>
<tr>
<td>WRF</td>
<td>Weather Research and Forecasting</td>
</tr>
<tr>
<td>WRF-NMM</td>
<td>Nonhydrostatic Mesoscale Model</td>
</tr>
<tr>
<td>WRF-ARW</td>
<td>Advanced Research WRF</td>
</tr>
</tbody>
</table>
List of figures

Figure 1: Flowchart of the WRF-ARW programs as they are used in the study ........................................10

Figure 2: Scheme of a 10-by-9-point horizontal grid with a 13-by-10-point nest (ratio of 3 between the main domain and the nest grid spacing) ........................................................................................................12

Figure 3: Non-scaled scheme of eta-levels over a mountain .....................................................................13

Figure 4: Map of the main domain (12-km resolution) and of the inner nest “d02” (4-km resolution). The black dots show the locations of Maïa Eolis’ wind farms .............................................................15

Figure 5: Example of the output plots of the daily forecasts for one wind farm (made with NCL). From top to bottom: wind speed at hub height (in m/s); wind direction; power output for one wind turbine (in kW); temperature at hub height (in Celsius). ........................................19

Figure 6: Screenshot of the software. There is one tab per wind farm. The actual power production from one wind turbine is shown in yellow. One WRF run is shown in dash line (colors from green (first day of run) to orange (8th day)) ........................................................................................................20

Figure 7: Map of the main domain (36-km resolution) and of the inner nest “d02” (12-km resolution) in the case of long-time runs from the FNL archive data. The black dots show the locations of Maïa Eolis’ wind farms ................................................................................................................21

Figure 8: Idem than Figure 9, for wind farm #8 (one of the worst cases) ................................................25

Figure 9: Normalized bias, normalized mean absolute error and normalized standard deviation error along the forecast horizon for the GFS runs. The green curves correspond to the run at a 4-km resolution (10-minute intervals between output horizons), the red curves to the 12-km resolution (1-hour interval). Wind farm #17 (one of the best cases) ...........................................................................................................25

Figure 10: Distribution of normalized errors at the wind farm #17. In grey: 2-month run with FNL data. The histogram is normalized but corresponds to more than 1500 values (62 days with one output per hour). In blue: daily forecasts with GFS data (4-km resolution) and forecast horizon between 0 and 24 hours (day 1), between 24 and 48 hours (day 2), etc. Each histogram has almost 8000 values (144 per day, times 57 runs) .................................................................................................27

Figure 11: Idem than Figure 10, for wind farm # 8 ......................................................................................27

List of tables

Table 1: Approximate computation time (computer-dependent) and output file size (dependent on some parameters) for an 8-day WRF run on the previously defined domain ........................................................16

Table 2: Evaluation of the WRF runs based on FNL archived analysis. The red lines correspond to the badly represented wind farms ........................................................................................................26
Annex A: Specific file formats

An issue of the thesis was to handle new command language and new data formats, specific to the weather prediction world.

GRIB

The GRIB (GRIdded Binary) form is a standard format for meteorological data, defined by the World Meteorological Organization (WMO, 1994). That is why all the forecasts available on the web are in this format.

Data are coded in binary, not directly readable. There are two versions of GRIB: in GRIB 2 the data are compressed so that the files are lighter.

When using GRIB files as input in WRF, the ungrib program deals with decoding the data to a more easy-to-read form. But you may also want to extract time series of some variables at some specified latitudes, longitudes and vertical levels directly from the GRIB files and without using WRF. In that case you may use the GRIB API tools (ECMWF, 2011). For example, the commands:

- `grib_ls file.grib` : will list the variables along with their keys (e.g. name, level type, level number).
- `grib_get -l lat,lon,1 -w key1=a,key2=b file.grib` : will give the value of the variable identified by the specified keys, on the grid point nearest to (lat,lon).

Another way is to convert the file into netCDF and to use NCL, which would be more powerful.

NetCDF

NetCDF (Network Common Data Form) is a powerful format to handle large multidimensional variables such as the WRF output files.

A netCDF file contains 3 types of objects:

- Dimensions (having a name and a size)
- Variables which are multidimensional arrays (varying along some of the defined dimensions)
- Attributes which are comments (having a name, type and value). They are attached to a variable or to the file itself.

An additional library, the NCO tools, may be used to read, create or modify a netCDF file. NCL (NCAR Command Language) is another very powerful tool for exploiting netCDF files.
Annex B: Tools for evaluating a prediction model

Some tools are proposed for standardizing the models evaluation (Madsen, 2004). The prediction error is defined by:

\[ e(t + k|t) = P(t + k) - \hat{P}(t + k|t) \]

Where \( P(t + k) \) is the measured power at time \((t + k)\).

\( \hat{P}(t + k|t) \) is the power prediction for time \((t + k)\) made at time \(t\).

The evaluation of the model is made with \(N\) forecast runs (\(N\) starting times \(t\)) with the 4 following tools:

- **Bias:**

\[ \bar{e}(k) = \frac{1}{N} \sum_{t=1}^{N} e(t + k|t) \]

- **Mean Absolute Error:**

\[ MAE(k) = \frac{1}{N} \sum_{t=1}^{N} |e(t + k|t)| \]

- **Root Mean Squared Error:**

\[ RMSE(k) = \sqrt{\frac{1}{N} \sum_{t=1}^{N} e(t + k|t)^2} \]

- **Standard Deviation Error:**

\[ SDE(k) = \sqrt{\frac{\sum_{t=1}^{N} [e(t + k|t) - \bar{e}(k)]^2}{N + 1}} \]

Those tools may be normalized by the installed power capacity.
Bibliography


Giebel, G., 2006. Description of NWP, Mesoscale and CFD models. ANEMOS Project, Deliverable 4.1a


Girard, N., 2006. Évaluation de la ressource éolienne à l'aide d'un logiciel CFD. Rapport de stage, Maía Eolis [in French]


