

Smart4RES

Data-assimilation in local atmospheric models for short-term renewable energy forecasting

D2.4 Data-assimilation in local atmospheric models

WP2, T2.4

Version V2.0

Authors:

Remco Verzijlbergh, WHIFFLE

Jakob Maljaars, WHIFFLE

Peter Baas, WHIFFLE



Disclaimer

The present document reflects only the author's view. The European Climate, Infrastructure and Environment Executive Agency (CINEA) is not responsible for any use that may be made of the information it contains.

Technical references

Project Acronym	Smart4RES
Project Title	Next Generation Modelling and Forecasting of Variable Renewable Generation for Large-scale Integration in Energy Systems and Markets
Project Coordinator	ARMINES – MINES ParisTech
Project Duration	November 2019 – April 2023

Deliverable	D2.4 Data-assimilation in local atmospheric models
Dissemination level ¹	PU
Nature ²	R
Work Package	WP 2 – Next Generation of Weather Forecasting Models for RES Purpose
Task	T2.4 – Data-assimilation in local atmospheric models for short-term renewable energy forecasting
Lead beneficiary	WHIFFLE
Contributing beneficiary(ies)	WHIFFLE
Reviewers	Quentin Libois (MF), Simon Camal (ARMINES)
Due date of deliverable	31 January 2023 (M)

1 PU = Public

PP = Restricted to other program participants (including the Commission Services)

RE = Restricted to a group specified by the consortium (including the Commission Services)

CO = Confidential, only for members of the consortium (including the Commission Services)

2 R = Report, P = Prototype, D = Demonstrator, O = Other

Document history		
V	Date	Description
0.1	12/01/2023	First version containing contributions from all partners
1.0	07/02/2023	Revised version after review
2.0	14/02/2023	Final version published by the Coordinator

Executive summary

This Deliverable report presents the work developed by Whiffle in the framework of Task 2.4 ('Towards innovative data assimilation in LES based weather models') from the Smart4RES project. The aim of this Task was to improve short-term forecasts with LES by using local observations as formulated by KPI 1.1 ('10% RMSE reduction of wind speed and radiative variables for LES forecasts a few hours ahead'). A data assimilation method based on Ensemble Kalman filtering was developed and tested in this task. To facilitate numerical experiments, a single-column model that is conceptually close to Whiffle's operational Large-Eddy Simulation (LES) model was developed and tested against observations and LES results. First, a proof-of-concept study with data-assimilation using numerical simulations of a stable atmospheric boundary layer was performed. Secondly, data-assimilation experiments using observations from the Cabauw meteorological met-mast were performed and compared against forecasts without data-assimilation. Results show that the EnKF is a suitable framework for data-assimilation in which both local atmospheric state variables as well as large-scale pressure gradients can be estimated. Over the first few forecast hours, a reduction of RMSE and MAE between 25% (immediately after data-assimilation) and 5-10% after 2 hours was observed. An initialisation experiment of a nocturnal fog layer qualitatively showed that using observations enhances the representation of the fog layer. Future work will focus on the implementation of the Ensemble Kalman filter in Whiffle's operational LES model.

Table of contents

I	Introduction	6
II	Data-assimilation	8
II.1	The Ensemble Kalman Filter	9
II.2	Unscented Kalman filter	10
III	Model and observations	11
III.1	Single column model	11
III.1.1	Governing equations	11
III.1.2	Closure	12
III.1.3	Boundary conditions	12
III.1.4	Large scale forcings	13
III.1.5	Variables used for data-assimilation	14
III.2	Observations	14
III.3	Validation of the single column model	14
III.3.1	GABLS	14
III.3.2	Cabauw	15
IV	Experimental setup and results	16
IV.1	GABLS case	16
IV.2	Dynamic model with real observations	19
IV.3	Initialisation of LES fields with tower measurements	22
V	Discussion and outlook for data assimilation in LES forecasting	24
VI	Conclusions and recommendations	25
VI.1	Recommendations for further research	25

LIST OF TABLES

LIST OF FIGURES

Figure 1	Schematic view of data assimilation in a prediction and filtering framework. Figure from Carrassi et al. (2018).	8
Figure 2	The Cabauw tower.	15
Figure 3	Comparison of GABLS results between the GRASP LES model (blue lines) and the single column model (red lines) for the first 9 hours of simulation. The LES results represent horizontal averages. The different red and blue shades correspond to different simulation times: from 1 hour (light red/blue) to 9 hours (dark red/blue)	16
Figure 4	Comparison of single-column model with LES and observations on two days in January 2018.	17
Figure 5	Overview of the GABLS data assimilation experiment with measurements. Panels show the vertical profiles of u-component wind speed and temperature.	18
Figure 6	Schematic representation of the ensemble member initialisation process. When the assimilation starts at $t_{assimilation}$, each ensemble member represents a model integration until a different lead time.	19
Figure 7	Overview of one day (2018-01-17) in the SCM data assimilation experiment with Cabauw measurements. Panels show the vertical profiles of u-component wind speed and temperature.	20
Figure 8	Mean absolute error averaged and root mean square over 365 runs in the year 2018 as a function of forecast horizon and height.	21
Figure 9	Relative mean absolute error and root mean square error of the run with data assimilation compared to the run without data assimilation for different heights and lead times.	21
Figure 10	Vertical profiles of relative humidity as observed and modeled by LES. Top left: original LES run without data assimilation, driven by ERA5 tendencies. Top right: LES run with initial conditions from tower measurements. Bottom left: using tower measurements and a dryer soil. Bottom right: Tower observations (only until 200m).	22
Figure 11	Diagram illustrating scalability of EnKF approach with LES models. Each LES model is run on a separate GPU node in a cloud environment. Model states x_i are sent to the headnode, which performs the EnKF and broadcasts the Kalman gains K back to the nodes.	24

Acronyms

ECMWF European Centre for Medium-Range Weather Forecasts.

EnKF Ensemble Kalman Filter.

KPI Key Performance Indicator.

LES Large-Eddy Simulation.

NWP Numerical Weather Prediction.

SCM Single-column model.

1. Introduction

Short-term forecasting of solar and wind energy is important for trading in renewable energy sources (RES) and grid stability. Trade on the day-ahead markets happens roughly between 12h and 36h in advance and is therefore based on forecasts that were generated even earlier, approximately between 24h and 48h in advance. For these forecast horizons, numerical weather prediction (NWP) based forecasts have the highest quality and thus form the basis of those wind and solar energy forecasts. However, the NWP based forecasts have lower skill than data-driven methods that use observations on short-time intervals, say from one minute to several hours ahead.

One of the reasons for this is the computation time needed to produce the NWP based forecast. The computation time is needed for two sequential steps: 1) the data-assimilation process that uses observations to estimate the initial condition of the NWP model and 2) the forward integration over the forecast horizon. As a result, there is a lag of several hours between the first available forecast timestep and the last measurement that were taken into account in the observations Yang et al. (2022). At the moment the NWP forecast becomes available, the information that was used to produce it is already several hours old and the first forecast timesteps are not that close to reality.

Assimilation of measurement data into a local atmospheric model - one of significantly lower dimension and possibly higher resolution - could have the benefit of a faster assimilation time and thus produce a forecast aware of the current atmospheric conditions. In comparison with purely statistical models like the VARX models (vector autoregressive models with exogenous inputs, see e.g. Messner and Pinson (2019)), using a physics based model has the benefit that the governing equations of atmospheric physics provide correlations between model variables that are hard to learn from data only. As an example, the relation between wind speed at 10m and wind speed at 100m is strongly dependent on atmospheric stability.

The use of high-resolution local atmospheric models as a basis for RES forecasting also has clear benefits over the large-scale NWP models because they can capture relevant physics like local terrain details, wind farm wake effects and low clouds and fog. As a result, Large-Eddy Simulation (LES) has emerged as a flow modelling and forecasting technique for renewable energy purposes (currently mostly for wind energy) Gilbert et al. (2020); Schepers et al. (2021); Verzijlbergh (2021); Baas et al. (2022).

Data-assimilation in atmospheric models has been one of the backbones of operational numerical weather prediction and as a consequence it is a well-established scientific field with a long history Kalnay (2002); Bauer et al. (2015). At the same time, however, it is a highly specialised and technical field where a relatively small number of experts are active. Possibly this is one of the reasons why data assimilation has not been studied much in the context of renewable energy forecasting. Today's approaches to data-assimilation in operational weather forecast centres are mostly based on either variational approaches abbreviated with 3D-VAR and 4D-VAR (where time is included as dimension) for example for the ECMWF (ECMWF, 2021), UK Met-Office (Milan et al., 2020) and Meteo-France's global model ARPEGE (Courtier et al., 1994). Ensemble based approaches are mostly used in higher-resolution models, see e.g. Gustafsson et al. (2018) for an overview on this. Recently, some operational weather models started using hybrid models that combine ensemble based and variational methods, like ICON global (DWD, 2023) and the GFS model (Kleist and Ide, 2015).

This document is organized in the following way: chapter 2 presents the background on data assimilation in general and the data assimilation algorithm used in this work: the ensemble Kalman filter. Chapter 3 presents a single column atmospheric model that is used for experimentation instead of an LES model. The use of simplified model was preferred because it allows for rapid experimentation, can be easily interpreted conceptually and does not require a complex high-performance infrastructure. The single-column model should not be considered a full-blown weather model, but only a simplified model that captures the momentum and temperature evolution. Clouds and humidity effects are ignored in this model. Chapter 4 presents results of a number of case studies: a proof-of-concept study with the GABLS case (a stable atmospheric boundary layer), a dynamic case where tendencies from the ECMWF high-resolution forecast are used as dynamic boundary conditions and actual observations are used.

The work has been performed as part of the Smart4RES project Camal et al. (2021). Eventually, the data assimilation methods described in the document will be implemented in a large-eddy simulation (LES) model and possibly included in the operational forecast model of Whiffle. To develop and analyse the algorithms, a simpler and faster single-column model was used instead of a full-blown atmospheric LES. This report describes the initial experiments carried out with the single-column model. Partly due to the simplicity of the single column model, these experiments should not be considered to be representative for an operational forecast context, so quantitative error scores should be understood as such.

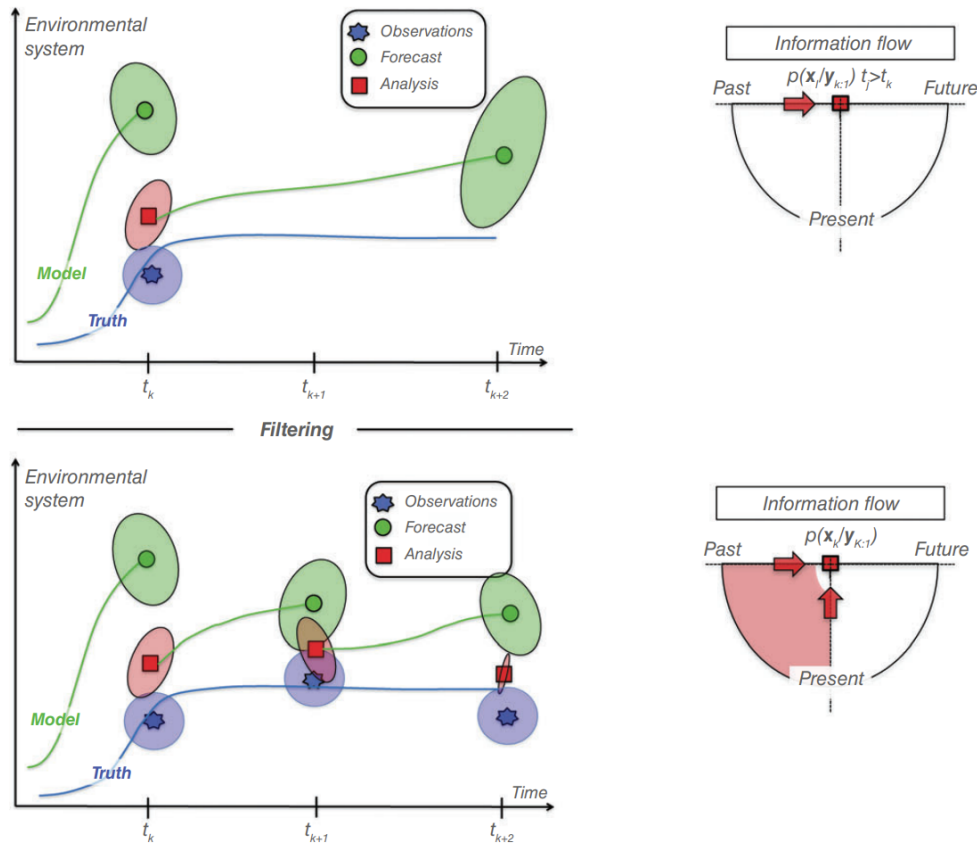


Figure 1 Schematic view of data assimilation in a prediction and filtering framework. Figure from Carrassi et al. (2018).

II. Data-assimilation

Data-assimilation, a term mostly used within the geosciences, refers to the process of finding the state of a model given observed data. In the field of control theory it is widely known as state-estimation, or how to estimate the state of a dynamical system given a set of (possibly noisy) measurements of a part of the system.

Data assimilation in NWP can thus also be seen as the process in which the initial conditions for the model are estimated. The result of this process, i.e. the initial condition is known as the *analysis*, and it is the starting point for the integrations forward in time that form the predictions. When the process of data assimilation is carried out sequentially whereby new measurements are constantly assimilated, this process is often referred to as *filtering*. Figure 1 shows schematically the data assimilation process in prediction and filtering.

Two closely related data assimilation algorithms have been explored in this research: the Unscented Kalman filter and the Ensemble Kalman Filter (EnKF), with an emphasis on the latter. We start with a brief description of the EnKF and later describe how the Unscented Kalman filter differs from it. For a more rigorous treatment of EnKF in the context of the geosciences, the reader is referred to Evensen (2006) and Houtekamer and Zhang (2016). For recent overviews and background about data assimilation in general, we refer to Evensen et al. (2022) and Carrassi et al. (2018).

II.1 The Ensemble Kalman Filter

The traditional Kalman Filter was developed in the field of control theory. The original Kalman Filter assumes linear systems in which both the time dynamics and the observations can be expressed as linear functions of the state variables. Several variations of the Kalman Filter have been proposed that can handle non-linear dynamics and observations operators, of which one gained particular traction in the geoscience community: the Ensemble Kalman Filter (EnKF).

The essence of a Kalman Filter is to apply a correction to a model state by multiplying a difference in observed state and model state:

$$\mathbf{x}^{update} = \mathbf{x} + \mathbf{K}\mathbf{y} \quad (1)$$

in which $\mathbf{x}$ denotes the vector of all model state variables, and $\mathbf{y} = \mathbf{z}_{mod} - \mathbf{z}_{obs}$ is the difference between the model and the observation in observation space, see below. The strength of the correction, or Kalman gain $\mathbf{K}$, depends on how much observation error is assumed, how much error the model prediction has and how strong the correlations are between observed and unobserved variables. It is given by:

$$\mathbf{K} = \mathbf{P}_{xz} \mathbf{P}_z^{-1} \quad (2)$$

with $\mathbf{P}_{xz}$ the cross-variance matrix and $\mathbf{P}_z$ the model covariance matrix (or background error matrix). In the Ensemble Kalman Filter, the cross-variance matrix and model covariance matrix are estimated based on the differences between a finite set of N model ensemble members $\mathbf{x}_i$). The covariance matrices can be derived from the ensemble members as:

$$\bar{\mathbf{z}} = \frac{1}{N} \sum_{i=0}^N \mathbf{z}_i \quad (3)$$

$$\mathbf{P}_z = \frac{1}{N-1} \sum_{i=0}^N (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T + \mathbf{R} \quad (4)$$

Here $\mathbf{R}$ is an appropriate covariance matrix representing measurement noise. An ‘observation’ from an ensemble member represents the state variables of that ensemble member passed through the observation operator:

$$\mathbf{z}_i = h(\mathbf{x}_i) \quad (5)$$

The use of an arbitrary function as an observation operator is one of the attractive features of the EnKF, since it allows for observations that are a non-linear function of the model state variables. A relevant example in the context of RES forecasting would be the use of wind power observations, which depend in a non-linear way on wind speed and other atmospheric variables.

The cross-variance matrix $\mathbf{P}_{xz}$ is computed from the ensemble members:

$$\mathbf{P}_{xz} = \sum_{i=0}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{z}_i - \bar{\mathbf{z}})^T \quad (6)$$

To apply the EnKF, a sequence of a predict and update steps is performed. In the predict step, we run all ensemble members through the process model operator. In our case, this means that

for all our ensemble members, we perform forward integrations of the atmospheric model. Then, the ensemble members are passed through the observation operator and the Kalman gain is computed.

To summarize, we have the following algorithm:

1. Integrate ensemble members:

$$\mathbf{x}_i = f(\mathbf{x}_i) + \mathbf{Q} \quad (7)$$

2. Pass ensemble members through observation operator:

$$\mathbf{z}_i = h(\mathbf{x}_i)$$

3. Compute covariance, cross-variance and Kalman gain:

$$\begin{aligned} \mathbf{P}_z &= \sum_{i=0}^N (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T + \mathbf{R} \\ \mathbf{P}_{xz} &= \sum_{i=0}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{z}_i - \bar{\mathbf{z}})^T \\ \mathbf{K} &= \mathbf{P}_{xz} \mathbf{P}_z^{-1} \end{aligned}$$

4. Update all ensemble members:

$$\mathbf{x}_i = \mathbf{x}_i + \mathbf{K}[\mathbf{z}_{\text{obs}} - \mathbf{z}_i + \mathbf{v}_R] \quad (8)$$

The term $\mathbf{v}_R$ in the last step denotes a random measurement error imposed on the measurements, with the effect that the observations are also treated as a stochastic process. The process noise $\mathbf{Q}$ has been set to zero in this research, because the way the ensemble members are constructed introduces enough uncertainty.

As shown in Figure 1, the above sequence can be repeated in a filtering context or performed once when the aim is to find an initial condition for a prediction.

The choice of ensemble members has not been discussed yet. This is a crucial step, since the spread in ensemble members will determine the model covariance matrix and thus the Kalman gain. By choosing ensemble members that are very close to each other, the EnKF will assign a very high weight to the model predictions and will discard the observations almost completely. Alternatively, when the ensemble members are too different, there will be little correlation between the observations and the unobserved model variables and the EnKF will at best only adjust the model variables that are directly observed. We describe two different strategies for choosing the ensemble member perturbations in the experimental setup section.

II.2 Unscented Kalman filter

The Unscented Kalman Filter (UKF) Van Der Merwe (2004) is closely related to the EnKF. It, too, uses ensemble members to estimate the model covariance and the cross-variance between model and observations. The two main differences are the way in which the ensemble members are created and the fact that ensemble members are created again for each update step. Initial experiments have been performed with this algorithm, but it was concluded that the method in which the ensemble members are constructed in the UKF leads to physically unrealistic wind-speed profiles. After applying some heuristic methods to remedy this and because the construction of the ensemble members is an integral part of the UKF, it was decided to abandon experiments with the UKF altogether and focus on the EnKF instead.

III. Model and observations

In this section we describe the single-column model and the observations used in this study. As reference model, we use Whiffle's LES model GRASP, of which a recent description is given by Baas et al. (2022).

III.1 Single column model

Single column models are used in atmospheric sciences as conceptually simple and computationally cheap replacements for full-blown models. They allow for rapid experimentation of changes to a model, for example a new physics parameterization. Single-column models describe the evolution of an atmospheric column and have no spatial coordinate. Large-scale processes like pressure gradients or the advection of momentum or temperature thus have to be supplied as external forcings to the model. LES models, when run for small domains with periodic boundary conditions, are conceptually similar to a single column model, see Neggers et al. (2012).

III.1.1 Governing equations

In the formulations below, following conventions in the atmospheric community, terms with overbars and prime variables represent turbulent transport that is not resolved explicitly by the model but needs a parameterization. We start with the governing equations for the u- and v-momentum components and temperature θ , that are given by:

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho} \frac{\partial p}{\partial x} + f v - \frac{\overline{w'u'}}{\partial z} = f(v - v_g) - \frac{\overline{w'u'}}{\partial z} \quad (9)$$

$$\frac{\partial v}{\partial t} = -\frac{1}{\rho} \frac{\partial p}{\partial y} - f u - \frac{\overline{w'v'}}{\partial z} = -f(u - u_g) - \frac{\overline{w'v'}}{\partial z} \quad (10)$$

$$\frac{\partial \theta}{\partial t} = -\frac{\overline{w'\theta'}}{\partial z} \quad (11)$$

with f the Coriolis parameter and ρ the air density. In the second equality on the right side, we defined the geostrophic wind components:

$$u_g = -\frac{1}{\rho} \frac{\partial p}{\partial y} \quad (12)$$

$$v_g = \frac{1}{\rho} \frac{\partial p}{\partial x} \quad (13)$$

The geostrophic wind is the wind resulting from a pressure gradient in the absence of friction. The rate of change of the momentum components is thus given by the balance between the geostrophic forcing and the vertical divergence of the turbulent fluxes $\frac{\overline{w'u'}}{\partial z}$ and $\frac{\overline{w'v'}}{\partial z}$.

III.1.2 Closure

The turbulent fluxes are modeled with the eddy-viscosity concept that relates the local fluxes to the local gradient (see e.g. Holtslag and Boville (1993)):

$$\overline{w'u'} = -K_m \frac{\partial u}{\partial z} \quad (14)$$

$$\overline{w'v'} = -K_m \frac{\partial v}{\partial z} \quad (15)$$

$$\overline{w'\theta'} = -K_h \frac{\partial \theta}{\partial z} \quad (16)$$

The values for the eddy-viscosities for momentum K_m and heat K_h (not to be confused with the symbol K for Kalman gain) are parameterized as:

$$K_m = K_h = l_c^2 S f(\text{Ri}) \quad (17)$$

where the length scale l_c is given by a height dependence and a critical length scale $\lambda_c = 30 (1 + \exp(1 - z/1000))$ that has a value of 30m in the mixed layer (assumed to be 1000m high) and decrease exponentially in higher air layers:

$$\frac{1}{l_c} = \frac{1}{\kappa z} + \frac{1}{\lambda_c} \quad (18)$$

where $\kappa = 0.4$ is the Von Karman constant and the absolute wind shear S given by:

$$|S|^2 = \left(\frac{\partial u}{\partial z} \right)^2 + \left(\frac{\partial v}{\partial z} \right)^2 \quad (19)$$

The dependence on the Richardson number Ri according to Holtslag and Boville (1993) is:

$$f(\text{Ri}) = \begin{cases} [1 + 10\text{Ri}(1 + \text{Ri})]^{-1} & \text{if } \text{Ri} \geq 0 \\ [1 - 18\text{Ri}]^{1/2} & \text{if } \text{Ri} < 0 \end{cases} \quad (20)$$

with the gradient Richardson number, a measure of atmospheric stability that expresses the ratio of buoyancy over shear, defined as:

$$\text{Ri} = \frac{g}{\theta} \frac{\partial \theta}{\partial z} \frac{1}{S^2} \quad (21)$$

III.1.3 Boundary conditions

For the lower boundary conditions, the well-known universal flux-profile relations from the Monin-Obukhov similarity theory are used:

$$\frac{\partial u}{\partial z} = \frac{u_*}{kz} \phi_m(z/L) \quad (22)$$

$$\frac{\partial \theta}{\partial z} = \frac{\theta_*}{kz} \phi_h(z/L) \quad (23)$$

with u_* the friction velocity and θ_* a scaling parameter for temperature. The universal functions $\phi_m(z/L)$ and $\phi_h(z/L)$ are empirically determined functions of the dimensionless number z/L in which L represents the Obukhov length.

Integration of the flux-profile relations leads to:

$$u(z) = \frac{u_*}{k} \left[\log \left(\frac{z}{z_0} \right) - \Psi_m(z/L) \right] \quad (24)$$

$$\theta(z) = \theta_0 - \frac{\theta_*}{k} \left[\log \left(\frac{z}{z_{0,h}} \right) - \Psi_h(z/L) \right] \quad (25)$$

$$(26)$$

We use settings common in LES models for the integrated flux-gradient functions Ψ_m and Ψ_h , i.e. Businger-Dyer for unstable cases and Beljaars-Holtstlag (1991) for stable cases, see e.g. Heus et al. (2010).

The surface fluxes can then be computed as:

$$\overline{u'w'}_0 = -C_M |u_1| u_1 \quad (27)$$

$$\overline{v'w'}_0 = -C_M |u_1| v_1 \quad (28)$$

$$\overline{\theta'w'}_0 = -C_H |u_1| (\theta_1 - \theta_0) \quad (29)$$

where the subscript 1 refers to the first model level and the transfer coefficients for momentum C_M and heat C_H given by:

$$C_M = \frac{\kappa^2}{\left[\log \left(\frac{z}{z_0} \right) - \Psi_m(z/L) \right]^2} \quad (30)$$

$$C_H = \frac{\kappa^2}{\left[\log \left(\frac{z}{z_0} \right) - \Psi_m(z/L) \right] \left[\log \left(\frac{z}{z_{0,h}} \right) - \Psi_h(z/L) \right]} \quad (31)$$

The Obukhov length L is found iteratively like described in Heus et al. (2010).

III.1.4 Large scale forcings

The information of the large-scale weather evolution is given to the single column model by prescribing dynamic tendencies. This method is identical to how LES or single column models are often run, see e.g. Neggers et al. (2012), Schaalkwijk et al. (2015) and Baas et al. (2018). The governing equations are appended as follows:

$$\frac{\partial u}{\partial t} = f(v - v_g) - \frac{\overline{u'w'}}{\partial z} + S_u^{adv} - \frac{1}{\tau}(u - u^{LS}) \quad (32)$$

$$\frac{\partial v}{\partial t} = -f(u - u_g) - \frac{\overline{v'w'}}{\partial z} + S_v^{adv} - \frac{1}{\tau}(v - v^{LS}) \quad (33)$$

$$\frac{\partial \theta}{\partial t} = \frac{\overline{\theta'w'}}{\partial z} + S_\theta^{adv} - \frac{1}{\tau}(\theta - \theta^{LS}) \quad (34)$$

where the variables u^{LS} , v^{LS} and θ^{LS} denote the values of the wind components and temperature from the large-scale model. The large-scale advection terms S_u^{adv} , S_v^{adv} and S_θ^{adv} are diagnosed from the large-scale model. The last terms $\frac{1}{\tau}(u - u^{LS})$ are nudging or relaxation terms that

prevent the variables from drifting too far away from the large-scale model values. The nudging time-scale τ determines the strength of the relaxation and is set to 6 hours in the present study.

Furthermore, the surface temperature tendency is also taken from the large-scale model:

$$\frac{\partial \theta_0}{\partial t} = \frac{\partial \theta_0^{LS}}{\partial t} \quad (35)$$

III.1.5 Variables used for data-assimilation

With the definition of the single-column model, we can connect the model state variables to the EnKF algorithm described in the previous chapter. In theory, the EnKF could also be used for parameter estimation, but we have not pursued that in the current research. However, the geostrophic wind components have been included in the data-assimilation algorithm. This means that the vector $\mathbf{x}$ consists of the atmospheric variables u , v and θ on all height levels extended with the geostrophic forcings u_g and v_g on all levels:

$$\mathbf{x} = [u_1 \quad \cdots \quad u_{N_z} \quad v_1 \quad \cdots \quad v_{N_z} \quad \theta_1 \quad \cdots \quad \theta_{N_z} \quad u_{g,0} \quad \cdots \quad u_{g,N_z} \quad u_{g,0} \quad \cdots \quad v_{g,N_z}] \quad (36)$$

III.2 Observations

The Cabauw meteorological tower provides high-quality and well documented atmospheric observations, see Bosveld et al. (2020). Figure 2 shows a picture of this 200m high tower.

A variety of advanced atmospheric and soil measurements are available, but most relevant for the current study are the tower measurements of wind speed and temperature. In addition, a dataset of ‘observed’ geostrophic winds is available that have been computed by fitting a pressure gradient field using neighbouring meteorological stations, see Bosveld (2020) for more information on this procedure. The observed geostrophic winds are particularly relevant because they are the main drivers of the wind speeds. Neggers et al. (2012) also used the Cabauw facility for experimenting with single-column atmospheric models.

III.3 Validation of the single column model

Because the single-column model is a conceptual and highly simplified model it is not intended for use in an operational setting. Rather, it is used as a substitute for more complex atmospheric LES model that allows rapid experimentation on different data-assimilation approaches. Nevertheless, it is insightful to look at its performance in relation to the more advanced LES model.

III.3.1 GABLS

The GABLS1 case is a well-known intercomparison in the LES community, see Beare et al. (2006) and Kosovic and Curry (2000). It describes a stably stratified arctic boundary layer that is characterized by a low-level jet and a strong vertical gradient in wind direction. The case has a constant geostrophic wind forcing of 8 m/s, an initial potential temperature profile of a constant 265 K over the first 100m (i.e. a well-mixed layer) topped by a stably stratified layer with a temperature gradient of 1K/km. At the surface, a cooling rate of 0.25 K/hour is applied. For the GABLS



Figure 2 The Cabauw tower.

case, the Blackader lengthscale λ_c is set to a small value of 6.0. The domain height was set to 400m with 128 vertical levels, yielding a resolution of 3.125m

The results of the SCM as well as LES results (with run settings identical to Beare et al. (2006)) for the GABLS case are shown in Figure 3. The SCM shows a quite comparable evolution of the boundary layer as the GRASP LES model, especially in the momentum components. The temperature mixing is markedly less strong in the SCM. This could be due to a higher aerodynamic resistance at the surface and/or due to less turbulent mixing.

III.3.2 Cabauw

As a second validation case, the SCM has been compared with observations from the Cabauw metmast. In this case, dynamical tendencies from the ECMWF high-resolution forecast have been used in both the LES and the SCM. The LES has been run with a horizontal resolution of 100m, a vertical resolution of 25m and a domain height of 3200m. The SCM has been run with a vertical resolution of 32m and a domain height of 4096m. Figure 4 shows the evolution of the vertical u-component wind speed profiles for the SCM, the LES and the tower observations for two typical days in January 2018.

The single-column model shows similar profiles as the LES and the observations. In other cases (not shown here), it was found that the single-column model has difficulties in representing strong shear cases. Also, cooling of the atmosphere near the ground is not well captured by the SCM. Overall, the similarity between the single-column model and the LES is considered sufficient for the purpose in this study: to perform preliminary data assimilation experiments for short-term forecasting with a local atmospheric model.

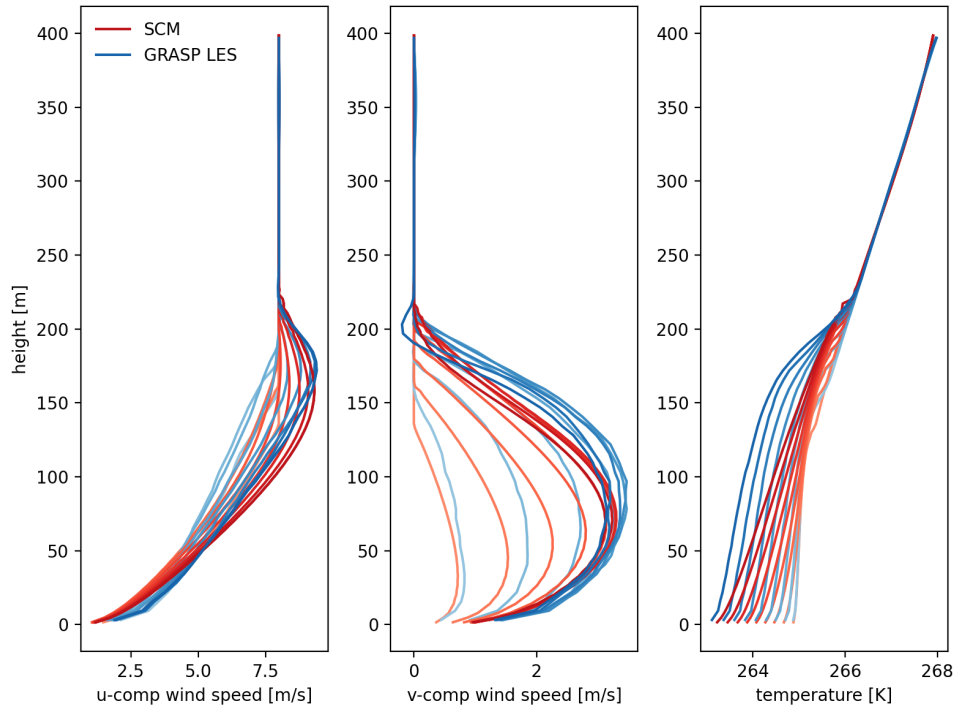


Figure 3 Comparison of GABLS results between the GRASP LES model (blue lines) and the single column model (red lines) for the first 9 hours of simulation. The LES results represent horizontal averages. The different red and blue shades correspond to different simulation times: from 1 hour (light red/blue) to 9 hours (dark red/blue)

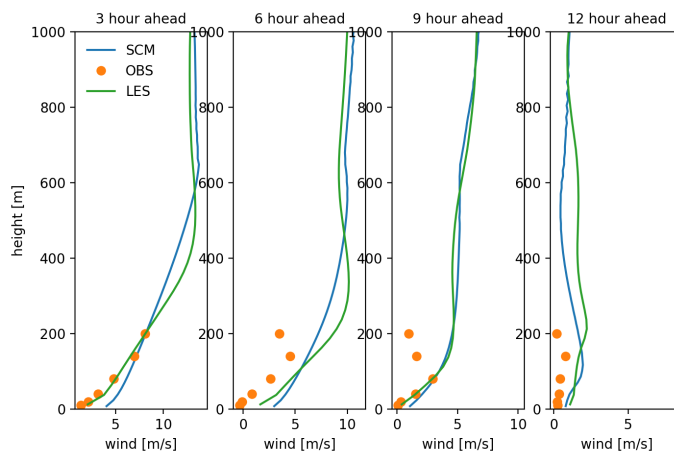
IV. Experimental setup and results

IV.1 GABLS case

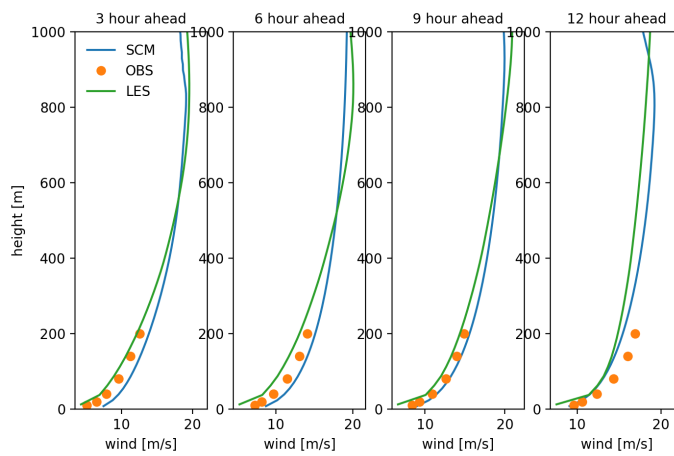
A data assimilation experiment using the GABLS case is set-up as follows: a reference case a geostrophic wind of 6 m/s (instead of the original 8 m/s) is run. Since the geostrophic forcing is constant in time, the only dynamics in the case is the evolution of the boundary layer and the cooling of the surface, see also figure 3. This run is treated as the 'reality' from which observations are sampled and used as input for the EnKF. The ensemble members ($N = 10$) are created by perturbing the original geostrophic forcing of 8 m/s by an extra component drawn from a normal distribution with $\mu = 0$ and $\sigma = 1$ m/s. In the experiment, a virtual metmast samples observations of u , v and θ from the reference run from the lowest model level at 3.125m to the 10th level at 30m height. A sequence of 6 predict-update steps has been performed, each prediction step with a horizon of 600 seconds.

Figure 5 shows the results of the GABLS experiment for the u-component of the wind and the temperature. It can be seen in the top right panel that the EnKF already assimilates the measurements from the reference profile correctly: the ensemble mean is very close to the reference profile from which the measurements were sampled. Secondly, one observes that the ensemble spread collapses. This already happened in the first step of the EnKF sequence (not shown here) and as a result, during the EnKF sequence, the updates become more insensitive to the observations.

It is also noteworthy that the EnKF manages to adapt the geostrophic wind very well, even though it was not part of the observed variables. This can be seen by the constant wind speed



(a) 2018-01-04



(b) 2018-01-16

Figure 4 Comparison of single-column model with LES and observations on two days in January 2018.

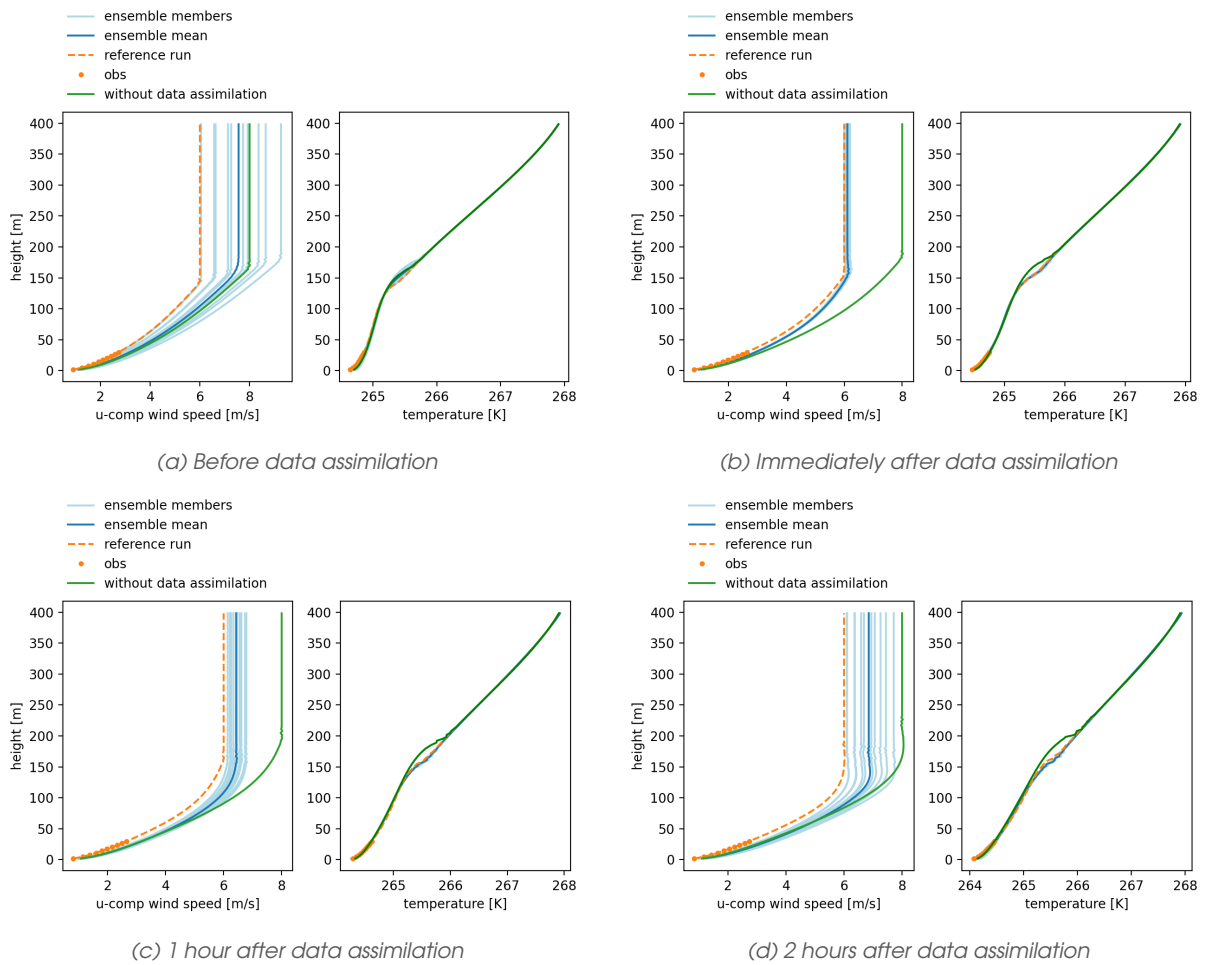


Figure 5 Overview of the GABLS data assimilation experiment with measurements. Panels show the vertical profiles of u-component wind speed and temperature.

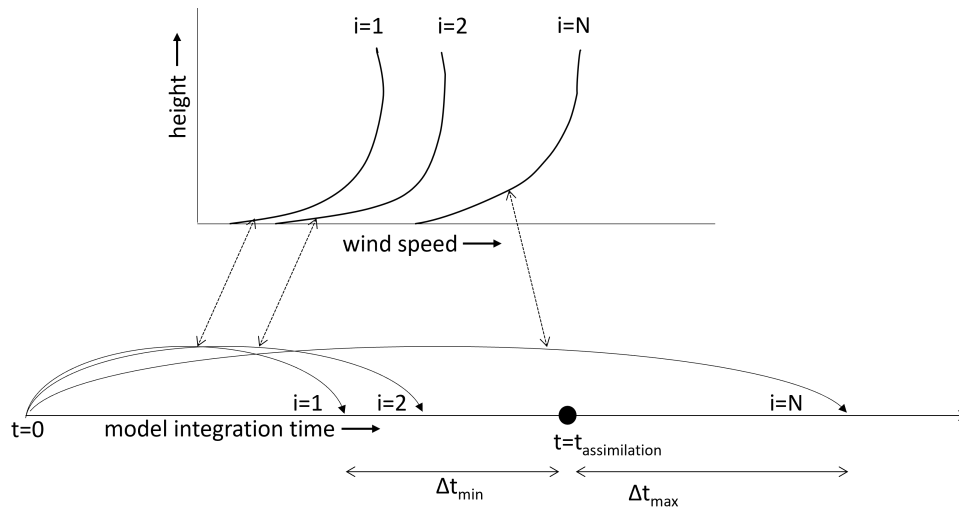


Figure 6 Schematic representation of the ensemble member initialisation process. When the assimilation starts at $t_{\text{assimilation}}$, each ensemble member represents a model integration until a different lead time.

above roughly 100m which is adjusted to the correct value of 6 m/s. Apparently, the correlation between the lower level variables and the geostrophic forcings was strong enough to adjust the unobserved variables correctly.

The GABLS results provide a proof-of-concept for data assimilation in a single-column model with an EnKF. It should, however, be acknowledged that this is a relatively simple case in which the reality was artificially constructed using the same model and there was little atmospheric dynamics.

IV.2 Dynamic model with real observations

As a more realistic experiment, we used the single-column model with the large-scale tendencies from the ECMWF high-resolution forecast and assimilated real Cabauw observations. The experiment was set-up as follows.

Ensemble members were created by running the single column $N = 10$ times with different integration times around the assimilation time at which the observations are available. This process is illustrated schematically in Figure 6.

The rationale behind this choice is that one of the largest sources of uncertainty in a local forecast model comes from the large-scale conditions, for example in the exact timing or location of passing weather systems. The ensemble thus represents all forecasted states of the atmosphere between two given moments in time. We started the forecast at 9UTC and used spin-up times distributed uniformly between 30 minutes and 3.5 hours ahead. The mean spin-up time was thus 2 hours. We then assimilated the Cabauw tower measurements from 11UTC and used this in a single update step, i.e. not a sequence of predict-update steps like the GABLS case. We used the tower measurements of u , v and θ and the ground 'measurements' of u_g and v_g ¹. After the update step, we set the model state to the ensemble mean and performed forward integrations of this model only. The reason why the individual ensemble members are not integrated further in time is that we are mainly interested in the best estimate of the atmospheric state at the assimilation time and to save computational cost.

¹Note that this is different than the GABLS case, in which no observations of the geostrophic wind were used

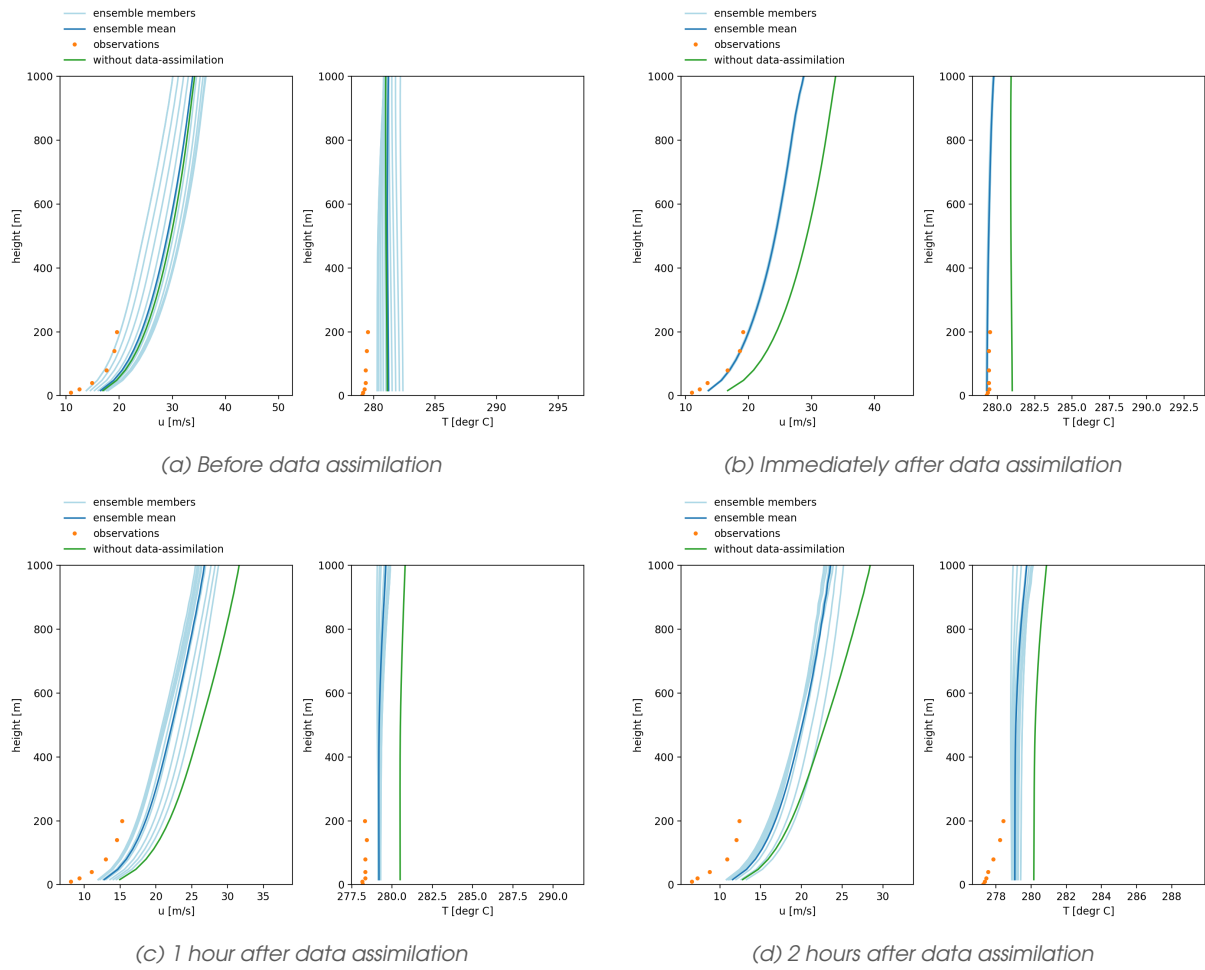


Figure 7 Overview of one day (2018-01-17) in the SCM data assimilation experiment with Cabauw measurements. Panels show the vertical profiles of u-component wind speed and temperature.

Because the u_g and v_g components are external inputs to the SCM, but they are modified by the EnKF, we need to specify a rule for their evolution in time. Keeping them constant after assimilation is not correct, but immediately overwriting them with the original values based on the large-scale model is also not correct. It was decided to relax them slowly towards the original values:

$$\frac{du_g(t)}{dt} = \frac{1}{\tau_{u_g}} (u_g^{LS}(t) - u_g(t))$$

where a time-scale of $\tau_{u_g} = 1$ hour has been used. This is a model parameter that deserves better tuning in future experiments.

As benchmark to compare the results with, we also performed a reference model run of a single-column model without data assimilation.

Figure 7 shows an overview of one run with data assimilation. The ensemble members, created as described above, are shown as the 10 light-blue lines. They can be seen to reflect different wind conditions that were forecasted in the 3 hours after 9.30UTC. It is good to recall that the spread between these ensemble members determines how much the EnKF adjusts the wind profiles to the observed ones: the more spread, the more weight is assigned to the observations.

Shortly after the assimilation step, the wind speed and the temperature are closer to the observations. After two more hours, the wind speed is very close again to the reference forecast

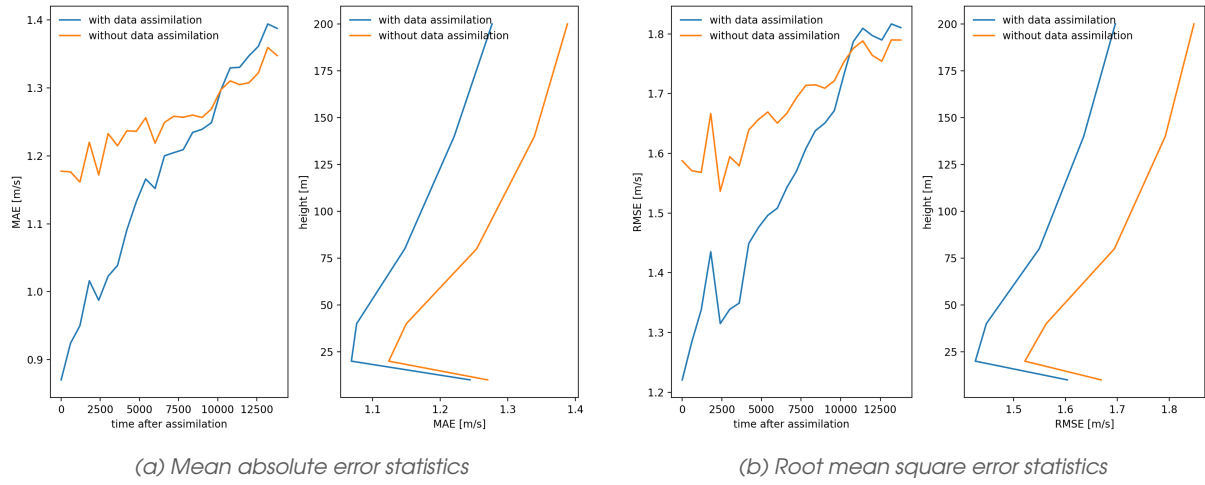


Figure 8 Mean absolute error averaged and root mean square over 365 runs in the year 2018 as a function of forecast horizon and height.

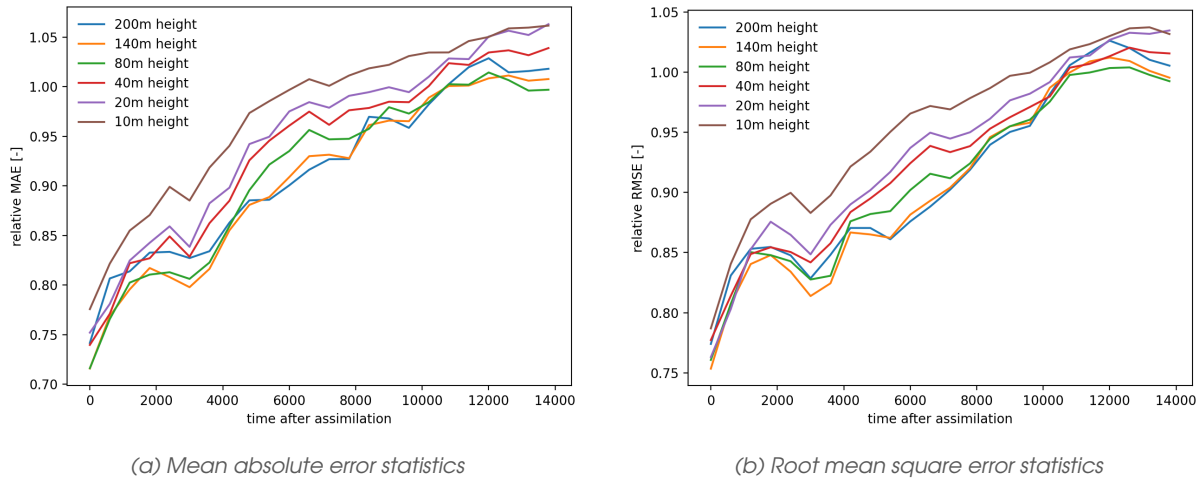


Figure 9 Relative mean absolute error and root mean square error of the run with data assimilation compared to the run without data assimilation for different heights and lead times.

without data assimilation. This is likely the result of the relaxation of the driving u_g, v_g components that has been described above.

To get statistically more meaningful results, we repeated the experiment described above for all days in the year 2018 and the compiled error statistics for this period. Figure 8 shows the mean absolute error and root mean square error of wind speed (averaged over all heights) as a function of lead time in the left panel. It is clear that the assimilation step reduces the forecast error for the first hours. The dependence of the forecast error on height is shown in the right panel of Figure 8. The near-surface winds have the highest errors, this is most likely due to the simple representation of the surface in the single column model. The reduction of MAE and RMSE by the data assimilation is present for all heights, but the relative reduction is largest for higher air layers.

We also computed the relative MAE and relative RMSE of the runs with data assimilation w.r.t the runs without data assimilation. Figure 9 shows that the reduction of the errors for most heights is around 25% and then decreases to roughly 5%-10% after two hours. It can also be observed that the relative reduction in MAE and RMSE is lowest for the 10m winds.

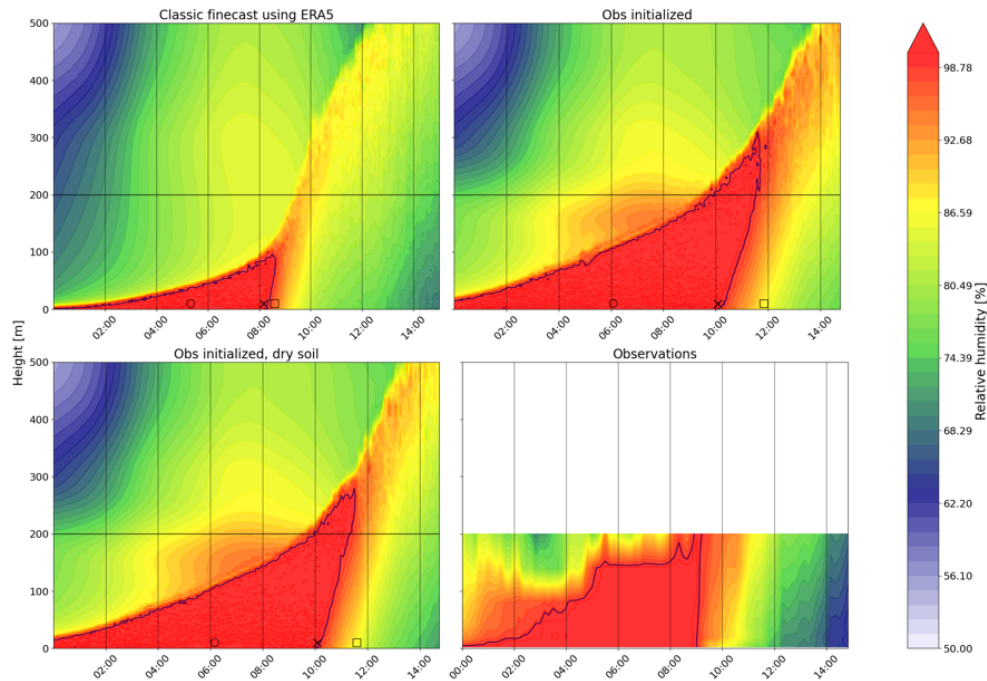


Figure 10 Vertical profiles of relative humidity as observed and modeled by LES. Top left: original LES run without data assimilation, driven by ERA5 tendencies. Top right: LES run with initial conditions from tower measurements. Bottom left: using tower measurements and a dryer soil. Bottom right: Tower observations (only until 200m).

IV.3 Initialisation of LES fields with tower measurements

A third case presented in this study is relevant for solar energy rather than for wind energy purposes. It was performed with the full physics LES model that Whiffle employs for operational forecasting in the context of the research described in Van Soest (2023). Furthermore, it differs from the previous cases because the data assimilation was performed in a heuristic way by simply changing the initial profiles of the LES using the observed values and leaving values higher than the tower height unchanged. In principle, the same data-assimilation strategy based on EnKF could be used here as well, but this is left for future work. Observations of wind speed, temperature and specific humidity taken at 22UTC the day before have been used to initialize the LES. The case that was run features the formation of nocturnal radiation fog as described by Maronga and Bosveld (2017).

Figure 10 shows the evolution of the fog layer as modeled and observed. The original LES run in the top left panel used ERA5 boundary conditions, but did not use any tower measurements. The fog layer is already present in this run, but it is more shallow than the observations. The top right panel shows the LES results when the tower measurements at 22UTC were used to initialize the model. This was done by setting the horizontal averages of the LES run to the tower measurements and interpolating between the tower measurement heights.

Using the tower observations as initial conditions clearly leads to a deeper and more persistent fog layer that matches the observations better. The sensitivity of the fog layer to soil moisture content was tested by altering the soil conditions, but the results do not change much for a dryer soil (bottom left panel).

The obtained results are of practical relevance in the light of solar power forecasting. The moment of dissipation of nocturnal fog layers has a large influence on solar power output and is notoriously difficult to forecast, see e.g. Boutle et al. (2018) and Maronga and Bosveld (2017).

Using an observation from 22UTC to improve the forecast of the moment of dissipation of the fog layer the next day is operationally possible and has a large practical relevance.

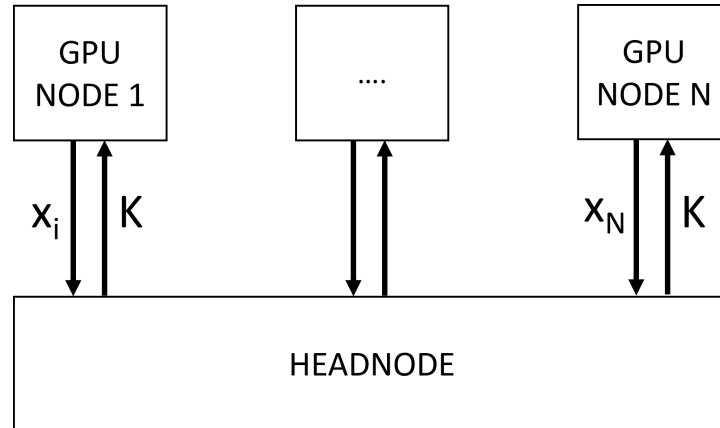


Figure 11 Diagram illustrating scalability of EnKF approach with LES models. Each LES model is run on a separate GPU node in a cloud environment. Model states x_i are sent to the headnode, which performs the EnKF and broadcasts the Kalman gains K back to the nodes.

V. Discussion and outlook for data assimilation in LES forecasting

The ultimate goal of this research is to assimilate different types of atmospheric measurements in local LES based weather models used for renewable energy forecasting. For wind energy, measurements of wind speed and power from wind turbine sensors, scanning LiDARs, profiling LiDARS as well as more traditional sources like met-masts, radio-soundings and automatic weather stations can be used. For solar energy, in addition to the traditional measurements (satellite, tower measurements, weather stations), one could use observations from all-sky camera's, pyranometers and solar panels themselves. As shown in this research, EnKFs have some attractive properties for non-linear models: they are conceptually simple and have modest computational requirements. The non-linear observation operator provides a lot of flexibility to represent the various types of atmospheric measurements described above.

A large difference between an operational LES model and the single-column model used in this research is the amount of grid points and thus state-variables. In the simulations with the single column model, only 128 vertical points were used for three state variables. More realistic LES sizes are $256 \times 256 \times 128$ grid points and have 6 state variables (horizontal winds u and v , vertical velocity w , temperature θ , specific humidity q_t and pressure p) amounting to approximately 50 million degrees of freedom. More research is needed to get insights in a good EnKF design in such cases. As a first step, this could be done with the aim to estimate the mean wind profiles and geostrophic forcings, very similar to the set-up of this research and essentially treating the LES as a single-column model. Furthermore, localisation is a technique that is often used when EnKF are employed in large geoscientific models Houtekamer and Zhang (2016).

Another attractive feature for operational data assimilation with EnKF is that it is straight-forward to parallelize the ensemble member integrations over different nodes. Since Whiffle runs its operational LES model on graphics processing units (GPUs), the nodes can be separate GPU nodes in a cloud computing environment. Figure 11 illustrates this schematically. It requires only a limited amount of data-transfer between the headnode that performs the EnKF algorithm and the GPU nodes.

Since short-term RES forecasts are nowadays mostly produced with techniques like VARX, it would be useful to compare LES runs with DA against such a benchmark. Using local and recent observations are always expected to benefit a forecast; the question is if there is significant benefit in using data assimilation or one can resort to simpler methods.

VI. *Conclusions and recommendations*

This report describes a set of data assimilation experiments carried out with a single-column atmospheric model and measurements from the Cabauw tower. It should be regarded a proof-of-concept study to test concepts that can later be implemented in an operational LES based forecast model.

The Ensemble Kalman Filter (EnKF) has been used as the algorithm for data assimilation. The results demonstrate the feasibility of this approach. In a proof-of-concept with a stylized atmospheric boundary layer, the EnKF algorithm led to a correct adjustment of the atmospheric variables to the profiles of the simulation from which ‘measurements’ were sampled.

In a more realistic setting, significant improvements in the short-term forecast were observed: from 30% right after the assimilation step to roughly 5% after two hours. The sensitivity of the forecast improvements for several EnKF design choices has, however, not been systematically assessed in this research.

A third case showed how tower measurements can be used to directly initialize the LES state variables in a case that features the formation of nocturnal fog. The forecasted evolution of the fog layer improved markedly by using these observations. This is a promising results in the light of solar power forecasting, for which the break-up of fog layers remains a critical issue.

Although the work presented here may be considered a succesful proof-of-concept, the eventual goal is to assimilate observations in an operational LES model. A number of recommendations for future research to achieve this goal are formulated next.

VI.1 Recommendations for further research

In parallel to implementing the EnKF in the LES model, more experimentation with the SCM model is deemed a useful next step. In particular, we recommend to systematically study the EnKF filter design by testing different methods to specify the process and observation covariance matrices $\mathbf{Q}$ and $\mathbf{R}$ and relaxation time-scales. Such settings may also be specific for the quality and quantity of the available observations. Data-denial experiments, in which, for example, part of the Cabauw tower observations are withheld, could provide useful insights for operational forecasting. After all, the availability of tall tower measurements in the close proximity of wind farms is highly unlikely. Nevertheless, it could be that further experiments provide more insights in the added value and commercial viability of such measurements, or possibly cheaper alternatives like profiling LiDARs or drones.

A second line of future work should focus on the implementation of a data assimilation algorithm in the operational LES model. As a first step, this can be done in a setting similar to the one used in this study: by treating the LES as a single column model and only trying to adjust horizontal mean profiles. A promising next step is to include SCADA wind farm observations in this setting. Because wind turbines can be modeled directly in the LES, it is straightforward to model an observation operator that mimics the SCADA observations. With respect to solar forecasting, more research into the value of observations for predicting fog layers or stratocumulus clouds is recommended.

More challenging but also more promising is the assimilation of data in large-domain and heterogeneous LES runs. In such a setting, there is no longer a prescribed geostrophic forcing, but pressures and velocities are directly prescribed at the boundary values of the LES. In these settings, the inclusions of satellite or ground-based sky images becomes possible. As discussed in the previous section, data assimilation in large LES models likely brings a range of additional com-

putational challenges, for example with respect to the amount of ensemble members needed. The long experience of several operational weather models with a variety of data assimilation methods will provide inspiration to tackle the future challenges.

References

- Baas, P., van de Wiel, B. J., van der Linden, S. J., and Bosveld, F. C. From near-neutral to strongly stratified: Adequately modelling the clear-sky nocturnal boundary layer at cabauw. *Boundary-Layer Meteorology*, 166:217–238, 2 2018. ISSN 15731472. doi: 10.1007/s10546-017-0304-8.
- Baas, P., Verzijlbergh, R., van Dorp, P., and Jonker, H. Investigating energy production and wake losses of multi-gigawatt offshore wind farms with atmospheric large-eddy simulation. *Wind Energy Science Discussions*, 2022:1–28, 2022. doi: 10.5194/wes-2022-116. URL <https://wes.copernicus.org/preprints/wes-2022-116/>.
- Bauer, P., Thorpe, A., and Brunet, G. The quiet revolution of numerical weather prediction. *Nature*, 525:47, 9 2015. URL <https://doi.org/10.1038/nature14956><http://10.0.4.14/nature14956>.
- Beare, R. J., Macvean, M. K., Holtzlag, A. A., Cuxart, J., Esau, I., Golaz, J. C., Jimenez, M. A., Khairoutdinov, M., Kosovic, B., Lewellen, D., Lund, T. S., Lundquist, J. K., McCabe, A., Moene, A. F., Noh, Y., Raasch, S., and Sullivan, P. An intercomparison of large-eddy simulations of the stable boundary layer. *Boundary-Layer Meteorology*, 118:247–272, 2 2006. ISSN 00068314. doi: 10.1007/s10546-004-2820-6.
- Bosveld, F. C. The cabauw in-situ observational program 2000-present: Instruments, calibrations and set-up. Technical report, KNMI, 2020. URL <https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubTR/TR384.pdf>.
- Bosveld, F. C., Baas, P., Beljaars, A. C., Holtzlag, A. A., de Arellano, J. V. G., and van de Wiel, B. J. Fifty years of atmospheric boundary-layer research at cabauw serving weather, air quality and climate. *Boundary-Layer Meteorology*, 177:583–612, 12 2020. ISSN 15731472. doi: 10.1007/s10546-020-00541-w.
- Boutle, I., Price, J., Kudzotsa, I., Kokkola, H., and Romakkaniemi, S. Aerosol-fog interaction and the transition to well-mixed radiation fog. *Atmospheric Chemistry and Physics*, 18:7827–7840, 6 2018. ISSN 16807324. doi: 10.5194/acp-18-7827-2018.
- Camal, S., Kariniotakis, G., Sossan, F., Libois, Q., Legrand, R., Raynaud, L., Lange, M., Mehrens, A. R., Pinson, P., Pierrot, A., Giebel, G., Göçmen, T., Bessa, R. J., Gouveia, J., Teixeira, L. M. A., Neto, A., Santos, R. M., Mendes, G., Nouri, B., Lezaca, J., Verzijlbergh, R., Deen, G., Sideratos, G., Vitellas, C. A., Sauba, G., Eijgelaar, M., and Petit, S. Smart4res: Next generation solutions for renewable energy forecasting and applications with focus on distribution grids. *CIRE2021 - The 26th International Conference and Exhibition on Electricity Distribution*, 2021:2899–2903, 2021.
- Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9:1–50, 2018. ISSN 17577799. doi: 10.1002/wcc.535.
- Courtier, P., Thépaut, J.-N., and Hollingsworth, A. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994. doi: <https://doi.org/10.1002/qj.49712051912>. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712051912>.
- DWD. Icon data assimilation. https://www.dwd.de/EN/research/weatherforecasting/num_modelling/02_data_assimilation/data_assimilation_node.html, 2023. Accessed: 2023-02-07.
- ECMWF. ifs documentation – cy47r3. part ii: Data assimilation, 2021. URL https://www.ecmwf.int/sites/default/files/elibrary/2021/81269-ifs-documentation-cy47r3-part-ii-data-assimilation_1.pdf.
- Evensen, G. *Data Assimilation: The Ensemble Kalman Filter*. Springer Berlin Heidelberg, 2006. ISBN 9783540383017.

- Evensen, G., Vossepoel, F., and Leeuwen, J. P. V. *Data Assimilation Fundamentals*. Springer Cham, 2022. ISBN <https://doi.org/10.1007/978-3-030-96709-3>. doi: <https://doi.org/10.1007/978-3-030-96709-3>. URL <https://link.springer.com/bookseries/15201>.
- Gilbert, C., Messner, J. W., Pinson, P., Trombe, P. J., Verzijlbergh, R., van Dorp, P., and Jonker, H. Statistical post-processing of turbulence-resolving weather forecasts for offshore wind power forecasting. *Wind Energy*, 23:884–897, 2020. ISSN 10991824. doi: 10.1002/we.2456.
- Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., Mile, M., Hamdi, R., Lindskog, M., Barkmeijer, J., Dahlbom, M., Macpherson, B., Ballard, S., Inverarity, G., Carley, J., Alexander, C., Dowell, D., Liu, S., Ikuta, Y., and Fujita, T. Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, 144:1218–1256, 4 2018. ISSN 1477870X. doi: 10.1002/qj.3179.
- Heus, T., Van Heerwaarden, C. C., Jonker, H. J. J., Siebesma, A. P., Axelsen, S., Van Den Dries, K., Geoffroy, O., Moene, A. F., Pino, D., De Roode, S. R., and Vila-Guerau de Arellano, J. Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications. *Geoscientific Model Development*, 3(2):415–444, 2010. doi: 10.5194/gmd-3-415-2010. URL <http://www.scopus.com/inward/record.url?eid=2-s2.0-79954567650-1&partnerID=40-1&md5=e6e3068d3425dc9113a57a0f7138530c>.
- Holtlag, A. A. M. and Boville, B. A. Local versus nonlocal boundary-layer diffusion in a global climate model. *Journal of Climate*, 6:1825–1842, 1993.
- Houtekamer, P. L. and Zhang, F. Review of the ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144:4489–4532, 2016. ISSN 15200493. doi: 10.1175/MWR-D-15-0440.1.
- Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2002. doi: 10.1017/CBO9780511802270.
- Kleist, D. T. and Ide, K. An osse-based evaluation of hybrid variational-ensemble data assimilation for the ncep gfs. part ii: 4dvar and hybrid variants. *Monthly Weather Review*, 143:452–470, 2015. ISSN 15200493. doi: 10.1175/MWR-D-13-00350.1.
- Kosovic, B. and Curry, J. A large eddy simulation study of a quasi-steady, stably stratified atmospheric boundary layer. *J. Atmos. Sci.*, 57:1052–1068, 2000.
- Maronga, B. and Bosveld, F. C. Key parameters for the life cycle of nocturnal radiation fog: a comprehensive large-eddy simulation study. *Quarterly Journal of the Royal Meteorological Society*, 143, 2017.
- Messner, J. W. and Pinson, P. Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *International Journal of Forecasting*, 35:1485–1498, 10 2019. ISSN 01692070. doi: 10.1016/j.ijforecast.2018.02.001.
- Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M., Halloran, G., Kelly, G., Li, D., Maycock, A., Payne, T., Piccolo, C., Stewart, L., and Wlasak, M. Hourly 4d-var in the met office ukv operational forecast model. *Quarterly Journal of the Royal Meteorological Society*, 146:1281–1301, 4 2020. ISSN 1477870X. doi: 10.1002/qj.3737.
- Neggers, R. A. J., Siebesma, A. P., and Heus, T. Continuous single-column model evaluation at a permanent meteorological supersite. *Bulletin of the American Meteorological Society*, 93(9): 1389–1400, 2012. ISSN 00030007. doi: 10.1175/BAMS-D-11-00162.1.
- Schalkwijk, J., Jonker, H. J. J., Siebesma, A. P., and Bosveld, F. C. A Year-Long Large-Eddy Simulation of the Weather over Cabauw: an Overview. *Monthly Weather Review*, 143:828–844, 2015. doi: 10.1175/MWR-D-14-00293.1.

- Schepers, G., Dorp, P. V., Verzijlbergh, R., Baas, P., and Jonker, H. Aeroelastic loads on a 10mw turbine exposed to extreme events selected from a year-long large-eddy simulation over the north sea. *Wind Energy Science*, 6:983–996, 2021. ISSN 23667451. doi: 10.5194/wes-6-983-2021.
- Van Der Merwe, R. Sigma-point kalman filters for probabilistic inference in dynamic state-space models, 2004.
- Van Soest, M. The Role of Soil Moisture in Land-Atmosphere Interactions - A Study using Large Eddy Simulations and the Tiled ECMWF Scheme for Surface Exchanges over Land. Master's thesis, Utrecht University, 2023.
- Verzijlbergh, R. A. Atmospheric flows in large wind farms. *Europhysics News*, 52:20–23, 2021. ISSN 14321092. doi: 10.1051/epn/2021502.
- Yang, D., Wang, W., Gueymard, C. A., Hong, T., Kleissl, J., Huang, J., Perez, M. J., Perez, R., Bright, J. M., Xia, X., van der Meer, D., and Peters, I. M. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality, 6 2022. ISSN 18790690.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 864337