Energy Meteorological In-Situ Big Data Analytics

Morris Riedel, Jonas Berndt, Charlotte Hoppe, Hendrik Elbers

Institute of Energy and Climate Research (IEK-8), Forschungszentrum Jülich GmbH, Jülich, Germany
University of Iceland, Reykjavik, Iceland
Rhenish Institute for Environmental Research at the University of Cologne, Cologne, Germany

Background & Objective

The stochastic nature of weather imposes wind and solar power as an uncertain source of electrical energy. Stable power grid management and energy trade on stock markets call for improvement of probabilistic wind and solar power forecasts. The major potential lies in the improvement of the underlying weather forecast.

Data Mining Application Methodology

A data mining application shall identify the relationship between observation compliance and perturbation techniques. This information serves as a basis to improve the further generation of ensemble members within a particle filter algorithm.

Meteological Ensemble

- An ultra large ensemble version of the Weather Research and Forecast model (WRF) as part of ESSAI ( Ensemble for Stochastic Integration of Atmospheric Simulation), which provides a comprehensive probability density evolution of the model state
- Computational efficient implementation on the JUQUEEN, which realizes communication between the ensemble members by introducing a second stage of MPI parallelism
- Initial values and lateral boundary values from the global ECMWF and GFS ensembles
- A broad variety of state-of-the-art techniques of uncertainty representation within the model (SKEBS – Stochastic Kinetic-Energy Backscatter Scheme, SPPT – Stochastic Perturbed Parameterization Tendency, perturbation of surface values, etc.)

Particle Filter/Smooother Approach

An ensemble size of (O)1000 member offers the possibility to apply a Sequential Importance Resampling Filter (SIRF, or Sequential Monte Carlo method) as a novel, non-linear data assimilation technique in atmospheric science.

Particle filtering consists of representing the initial density of the model state by an ensemble of size N

\[ p(\chi) = \frac{1}{N} \sum_{i=1}^{N} \delta(\chi - \chi_i) \]

By applying Bayes’ theorem, we estimate the posterior density of the model state, given the observations \( \mathcal{D} \)

\[ p(\chi|\mathcal{D}) = \frac{p(\mathcal{D}|\chi)p(\chi)}{\int p(\mathcal{D}|\chi)p(\chi) d\chi} \]

Whereas the weights are given by

\[ w_i = \frac{1}{N} \delta(\chi - \chi_i) \]

Each particle member gets a contact weight with respect to the observations

Minimize ensemble variance to negligible members with least weights and spawn members with higher weights.

Energy Meteorological In-Situ Analytics Application Loop

- The coupled forecast-analysis system combines the meteorological forecast, particle filtering and data mining in one application loop. Due to high computational demands, special focus is set on computational efficiency. The particle filter as well as the piSVM1.2.1 shall be embedded in the WRF ensemble setup, with inter-member communication exclusively at particle filter resampling steps.

In-Situ Analytics Application Loop

The in-situ analysis loop for offshore data mining and retrieval consists of the following steps:

1. Off-line Data Mining
2. Observational Data Processing
3. In-situ data collection
4. Particle Filter
5. Particle Smoother

Off-line Data Mining

A medium-size ensemble of 256 members is classified and featured due to temporal compliance with measurement window observation (target forecast variable) over historical periods of different atmospheric conditions. This training phase of the SVM serves as the basis for the on-line data mining. Beforehand, the relationship between the particle filter resampling interval and forecast skill has to be determined empirically (connected to feature engineering), yet an unresolved issue in atmospheric data assimilation.

Observational data processing unit

Observations of high temporal frequency are processed to capture the temporal evolution of the atmosphere within an appropriate particle filter resampling period (off-line and on-line).

- **In-situ** pressure observations as an indication of the time evolution of synoptic conditions = various measuring masts evaluated at hub height for a sufficient boundary compliance
- **Satellite imagery** to be included for the most sensitive phenomena of frontal system movement and convective system development

The obstacle of filter degeneracy in high-dimensional systems has been overcome. For approaches like particle mesh, localization and/or sampling from a proposal density.

Data Preparation & Initial Studies

- The first dataset available is providing wind tower features with overall 3584 (100%) samples
- We follow on approach to take 1/5 (717, 20%) for validation and 1/5 (717, 20%) for testing
- This leaves training data with 2150 (60%) samples in order to create a model of the data
- Cross-validation is performed to optimize parameters, detailed accuracy results will follow

Acknowledgements

The authors gratefully acknowledge the Gauss Centre for Supercomputing (GCS) for providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS’s share of the supercomputer JUQUEEN at Jülich Supercomputing Centre (JSC). GCS is the alliance of the three national supercomputing centres: HLRS (Universität Stuttgart, 36C (Forschungszentrum Jülich), and LRZ (Bayerische Akademie der Wissenschaften), funded by the German Federal Ministry of Education and Research (BMBF) and the German State Ministries for Research of Baden-Württemberg (MWW), Bayern (SMBWFK) and Nordrhein-Westfalen (MIWF).