Cloud and Solar Power Prediction within the Helmholtz Analytics Framework

Dorit Jerger[1], Daniel Todt[1], Jonas Berndt[1], Hendrik Elbern[2], Kai Krajsek[1]

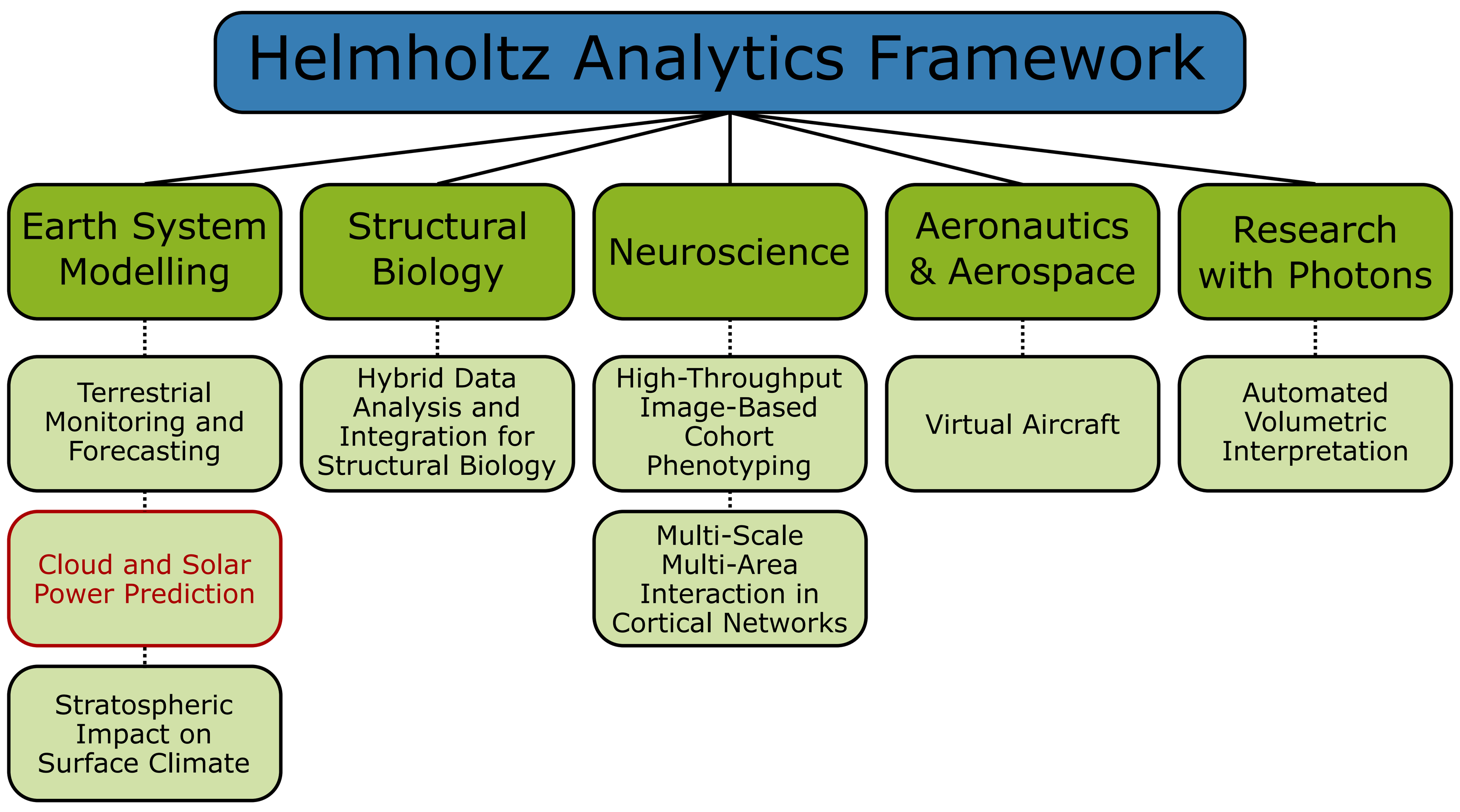
[1]Forschungszentrum Jülich [2]Rhenish Institute for Environmental Research at the University of Cologne

Introduction

In a pilot project a cooperation between data scientists and domain scientists from different research fields has been formed with the goal to develop methods of Scientific Big Data Analysis (SBDA) to tackle problems of great scientific interest and in areas where large data volumes occur. In the sub-project "Cloud and Solar Power Prediction" a large ensemble forecast set is investigated using methods of supervised machine learning. Cloud observations from satellite data will be assimilated by means of sequential importance resampling filters and smoothers.

Helmholtz Analytics Framework

- 6 Helmholtz centres, 5 research fields, 8 use cases
- exchange of SBDA techniques among use cases
- generalisations and standardisations of methods and tools ⇒ advanced analytics toolset

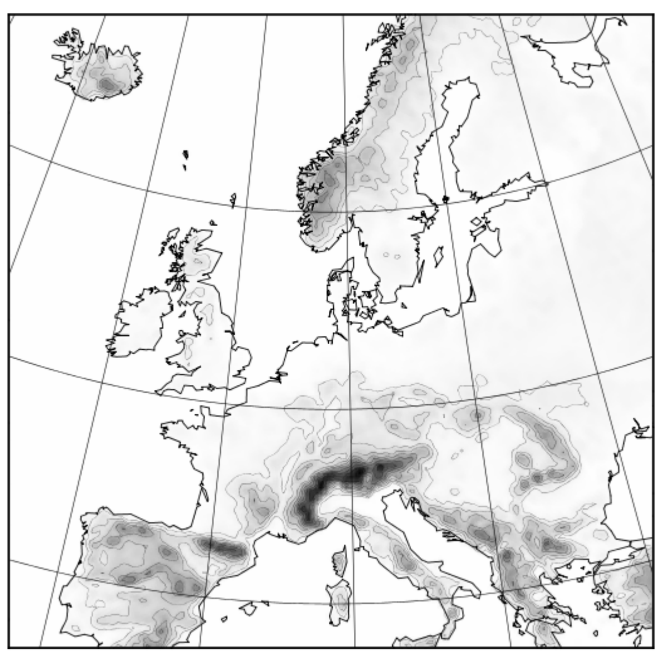


Cloud and Solar Power Prediction

- objective: improve predictability of clouds and solar energy
- support power grid stability
- provide better price formation at energy exchanges

Data Generation

- ensemble version of the WRF model has been developed
- ensemble forecast set with **512 members** over 6 months (01/04/15 – 30/09/15) on a 20km resolution grid generated
- 184 days with forecast horizon of 48 hours and 1 hour forecast step
- GEFS reforecast dataset consisting of 1 control member and 10 perturbed members used as initial and boundary values
- Stochastically Perturbed Parameterisation Tendency (SPPT) and Stochastic Kinetic Energy Backscatter (SKEB) Schemes used to represent model uncertainty

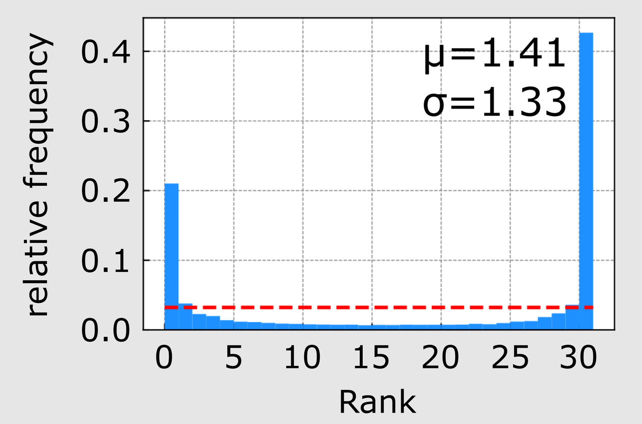


Model domain (orographic view)

First Steps in Ensemble Calibration Using Deep Neural Networks

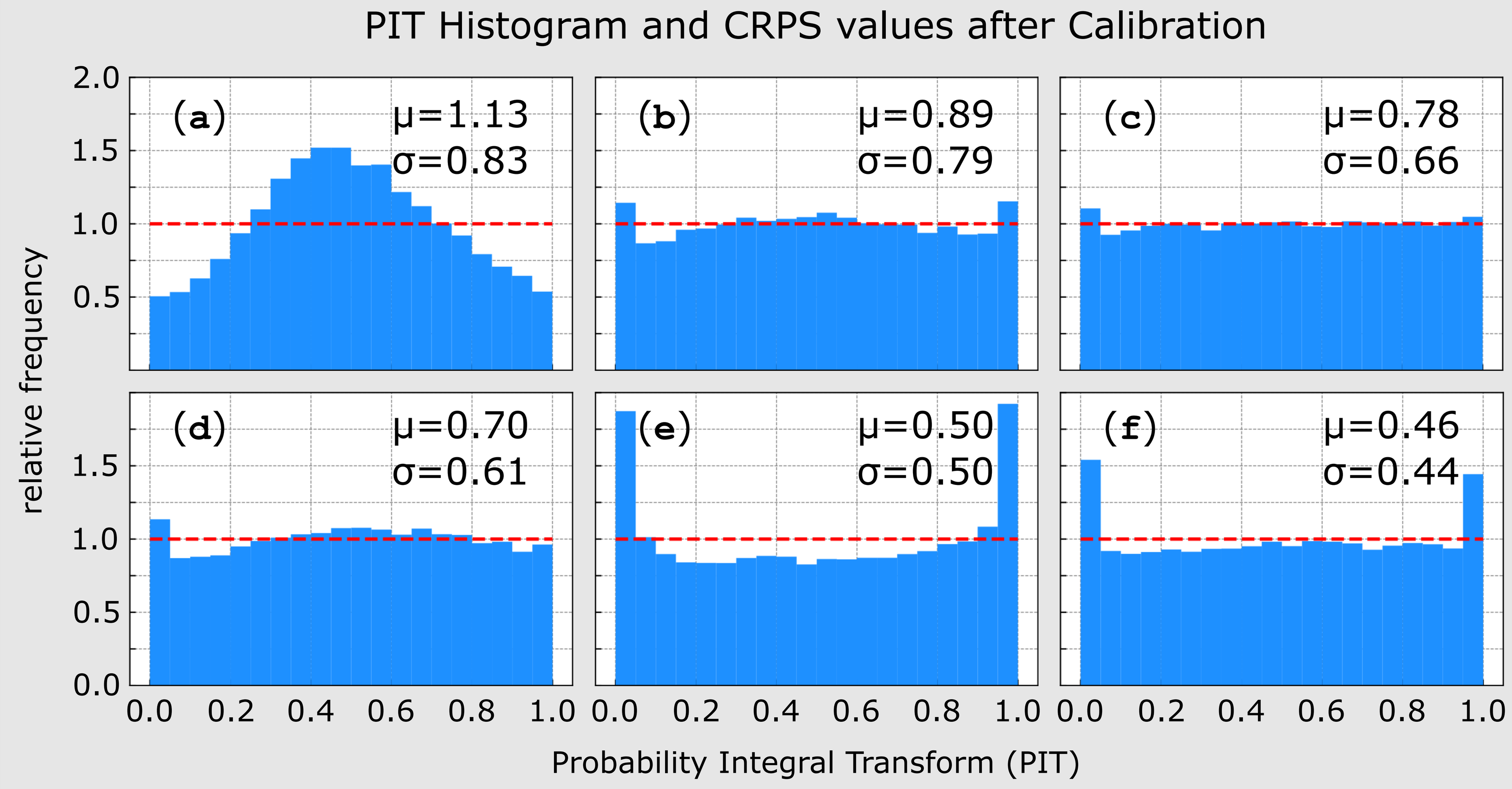
- 2 metre temperature calibrated
- a draw of 30 ensemble members used for calibration
- observations taken from NCEP ADP Global Upper Air and Surface Weather Observations
- influence of surrounding model grid values investigated

Talagrand Diagram and CRPS values of uncalibrated ensemble



- Continuous Ranked Probability Score (CRPS) used as loss function with mean μ and standard deviation σ
- the following calibration methods have been implemented:
  (**a**) Ensemble Method Output Statistics (EMOS)
  (**b**) linear neural network with no hidden layer
  (**c**) same as (**b**) and including surrounding model grid values
  (**d**) deep neural network with up to 4 hidden layers
  (**e**) same as (**d**) and including surrounding model grid values
  (**f**) convolutional neural network with 2 convolutional layers and including surrouding model grid values

PIT Histogram and CRPS values after Calibration



Next Step: Particle Filter

Work plan:

- investigation of different combinations of disturbed model variables (using different levels of complexity)
- implementation of various supervised machine learning approaches (e.g. support vector machines) and testing for their use as resampling method
- comparison of methods with other assimilation techniques

Acknowledgements

The project is funded by the Helmholtz Association Initiative and Networking Fund. Additional funding was provided by the European Union's Horizon 2020 research and innovation programme.