High frequency forecasting for wind energy using statistical and machine learning post-processing methods

I. Schicker

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Outline

1. Introduction
2. Post-processing for wind energy applications / forecasting – classical approach
3. Machine Learning and statistics
4. Methods, data, and tools
5. Examples

Introduction

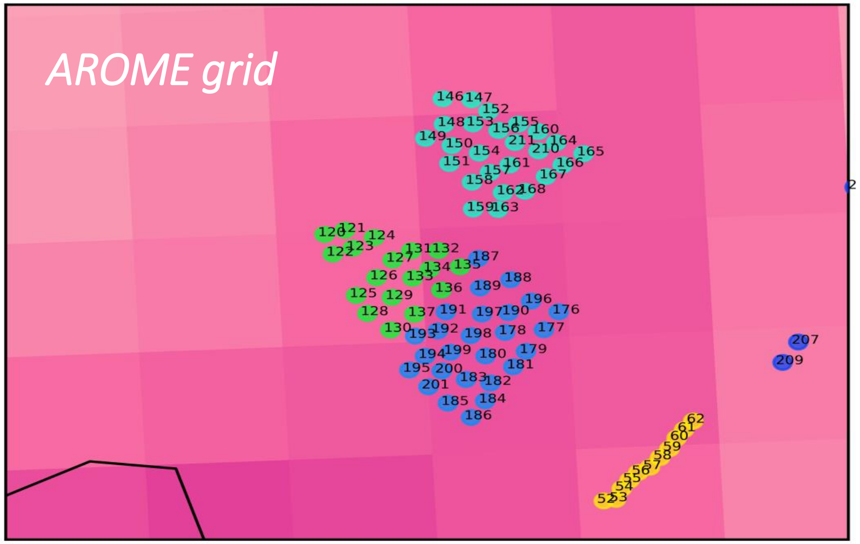
- Renewable and wind energy production is increasing
- Wind energy is highly weather dependent
- Fluctuations in production cause fluctuations in power transmission and consequently can cause grid instabilities and need counteractions to ensure stability
- Energy producers and traders need to submit day(s)-ahead and intra-day production → if above / below fees can arise
- Targeted predictions of expected power need to be as accurate as possible, should also include uncertainty estimation, and with a high temporal frequency
- Often, also meteorological field forecasts are needed besides power forecasts

How to generate targeted predictions? What tools can we use, what could be an additional input?

ZAMG

Numerical weather prediction models

AROME 2.4 x 2.4 km model grid and wind turbines



- grid rather coarse for wind turbines but better than global model (9 x 9 km)
- **models provide forecast of wind speed also at different altitudes (e.g. 80, 100, and 135 m a.g.l.), suitable for nacelle heights → BUT: only wind components, seldom temperature**

Challenges:

- Data accessability of turbine data (model verification, model tuning,...)
- Turbine specs ?
- Data transfer
- Smoothness of forecasts
- „downscaling“ to turbine level
- storage

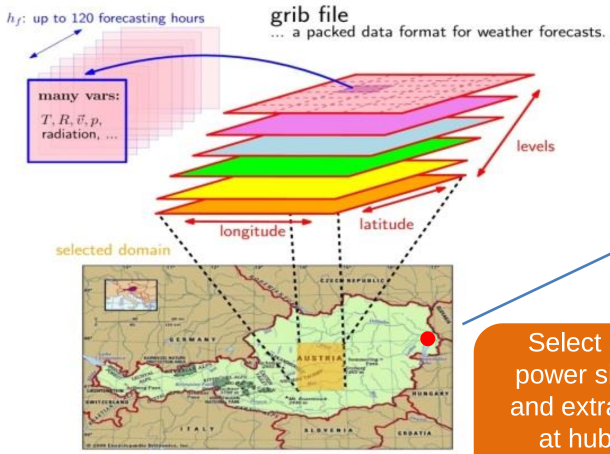
Numerical weather prediction models – wind turbine forecast (?)

Post-processing

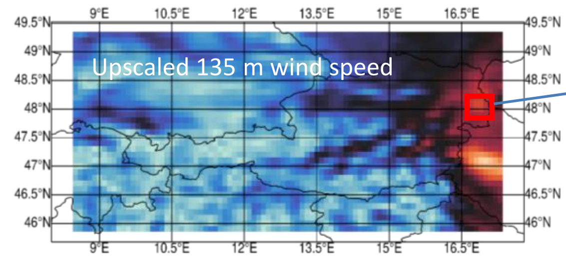
ZAMG

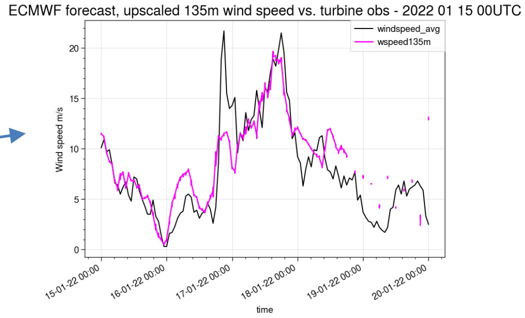
Post-processing for wind energy applications / forecasting – simple approach

NWP model (ECMWF / AROME / COSMO / WRF / …)

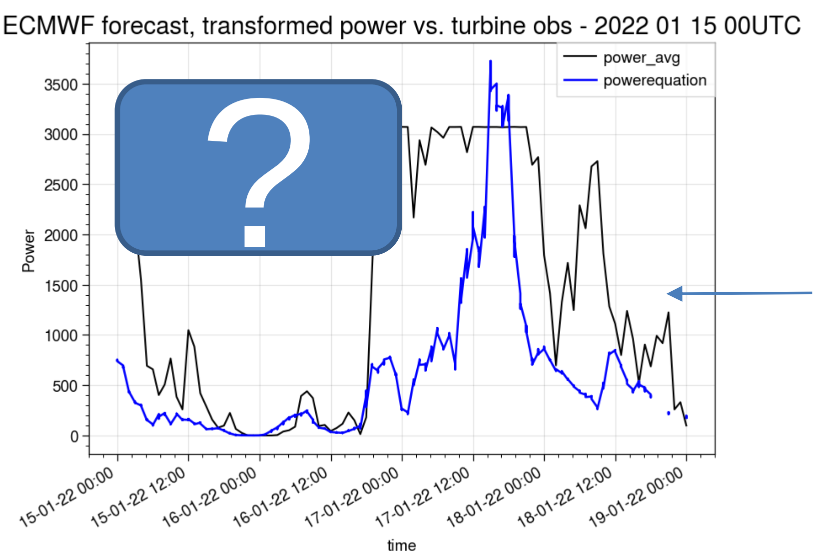


Select a power site and extract at hub height

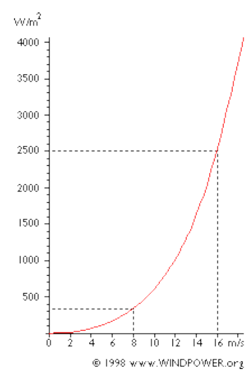




How to get the produced power?



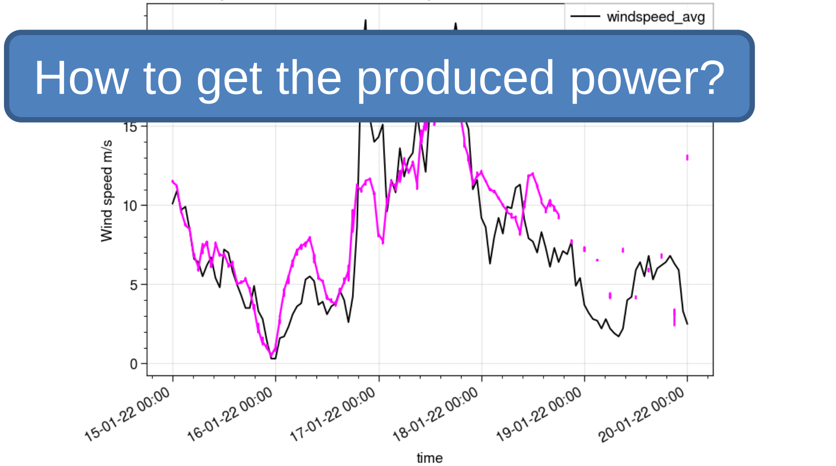
If you don't have a power curve or more than upscaled wind speed…



$$Power = \frac{1}{2}\rho\, v^3 \pi r^2 C_p$$

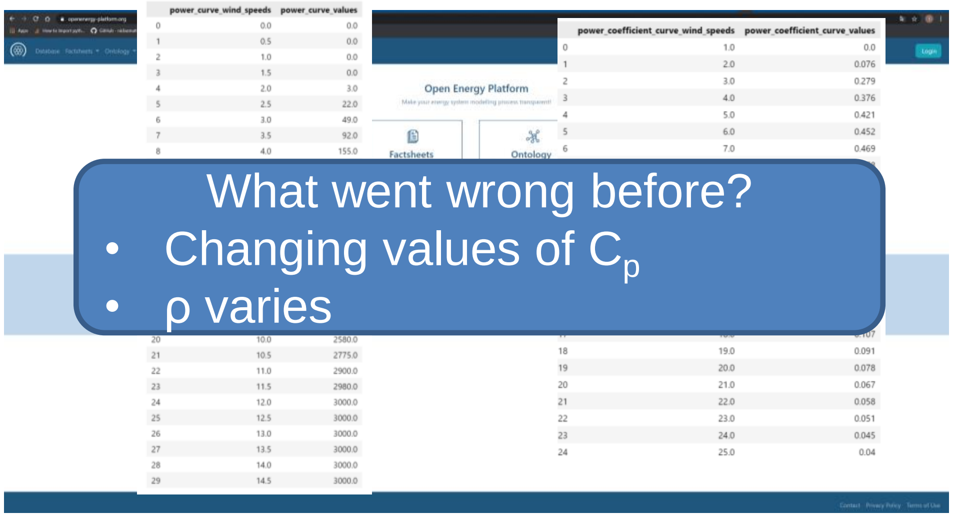
ρ = density of dry air (1.225 kg/m³)
v = velocity of wind (m/s)
r = radius of rotor (m)
C_p = power coefficient (0.4)

Post-processing for wind energy applications / forecasting – simple approach



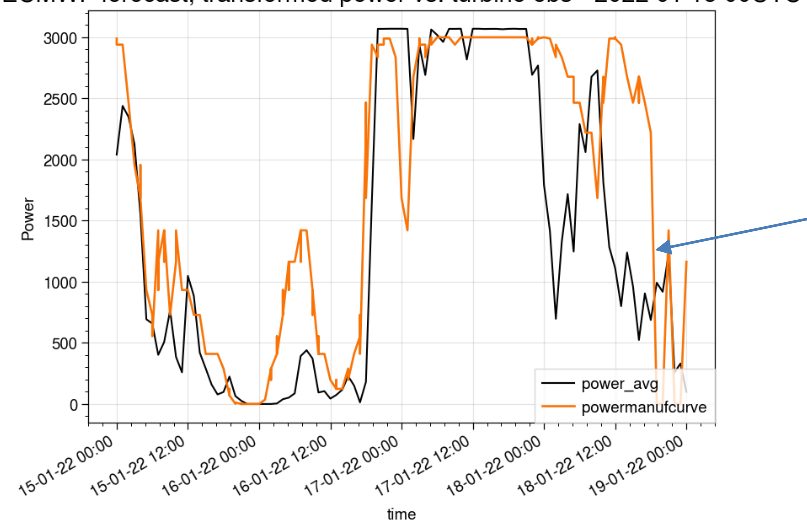
How to get the produced power?

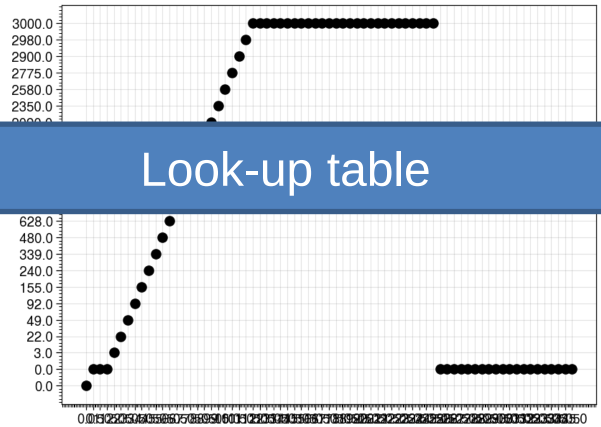
Typically, you know what turbine type you are predicting



What went wrong before?

- Changing values of C_p
- ρ varies



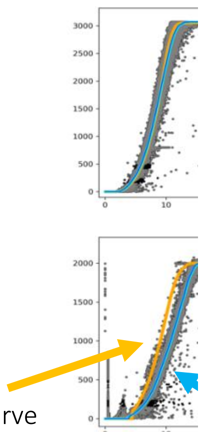


Look-up table

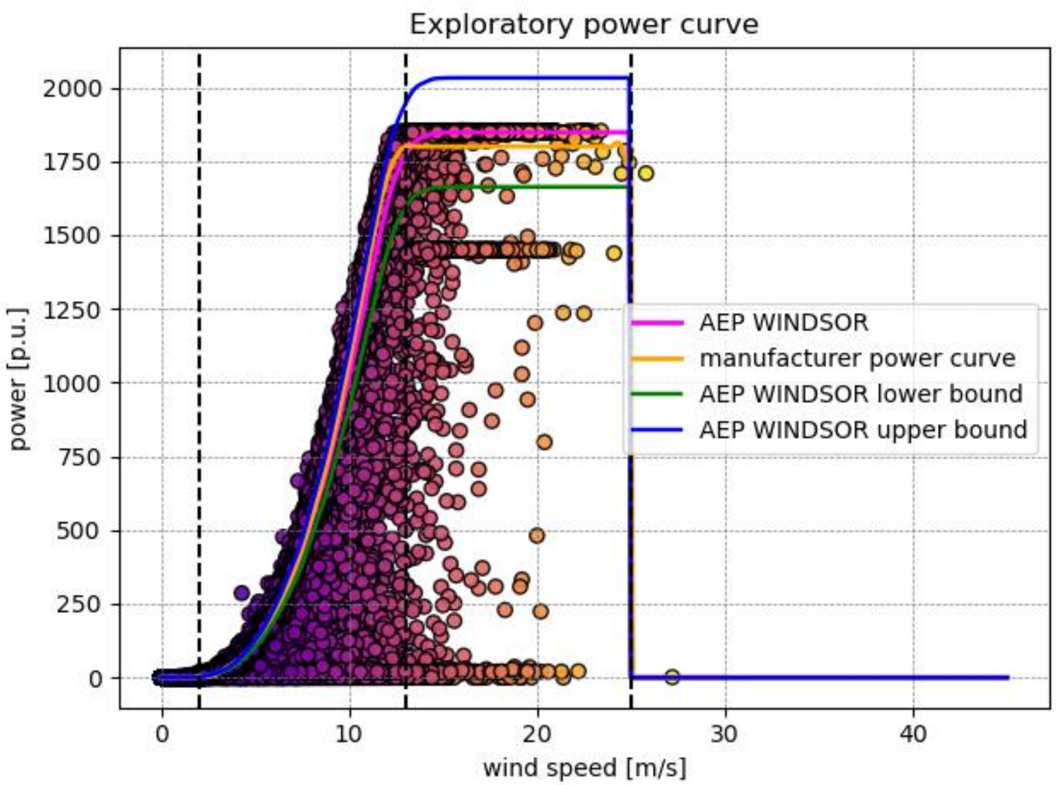
Post-processing for wind energy applications / forecasting – short side note…SCADA data

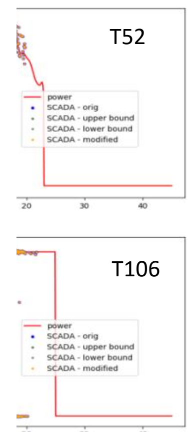
- Input, target, and validation data to any kind of algorithm
- Needs:
  - (basic) quality control algorithms → when did curtailment happen, etc.
  - Periodically updating the annual energy prodction curves

QC wind

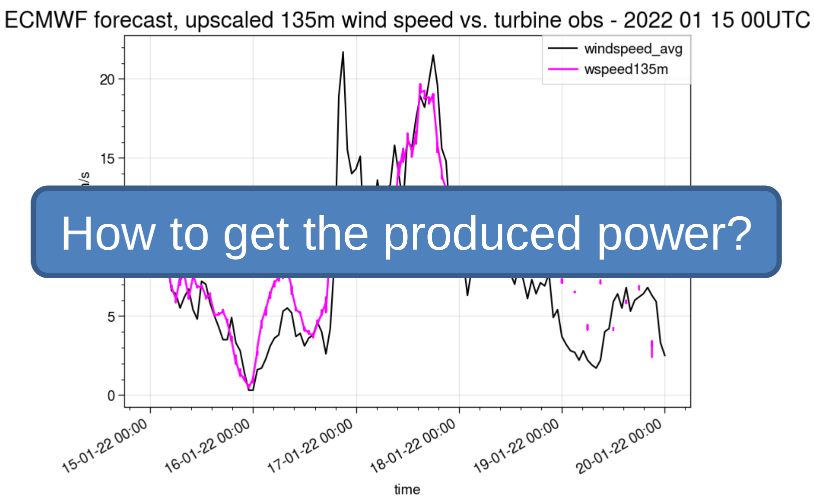


power curve manufacturer



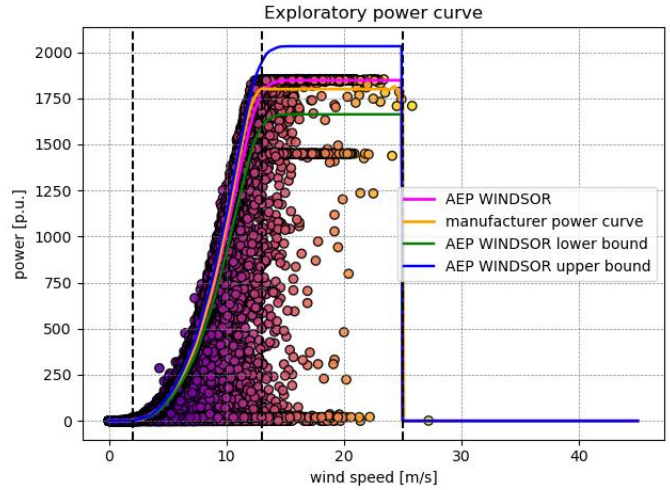


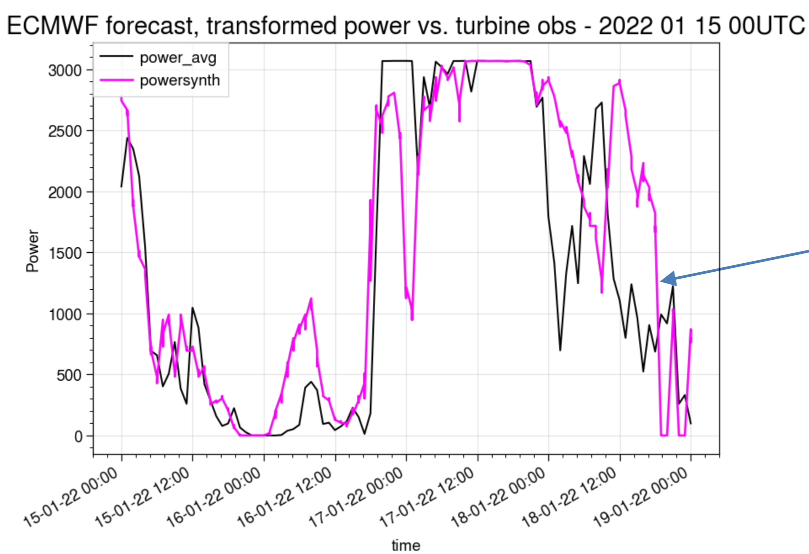
Post-processing for wind energy applications / forecasting – simple approach



How to get the produced power?

Ideally, you get historic turbine data usable for diff. purposes





Look-up table

ZAMG

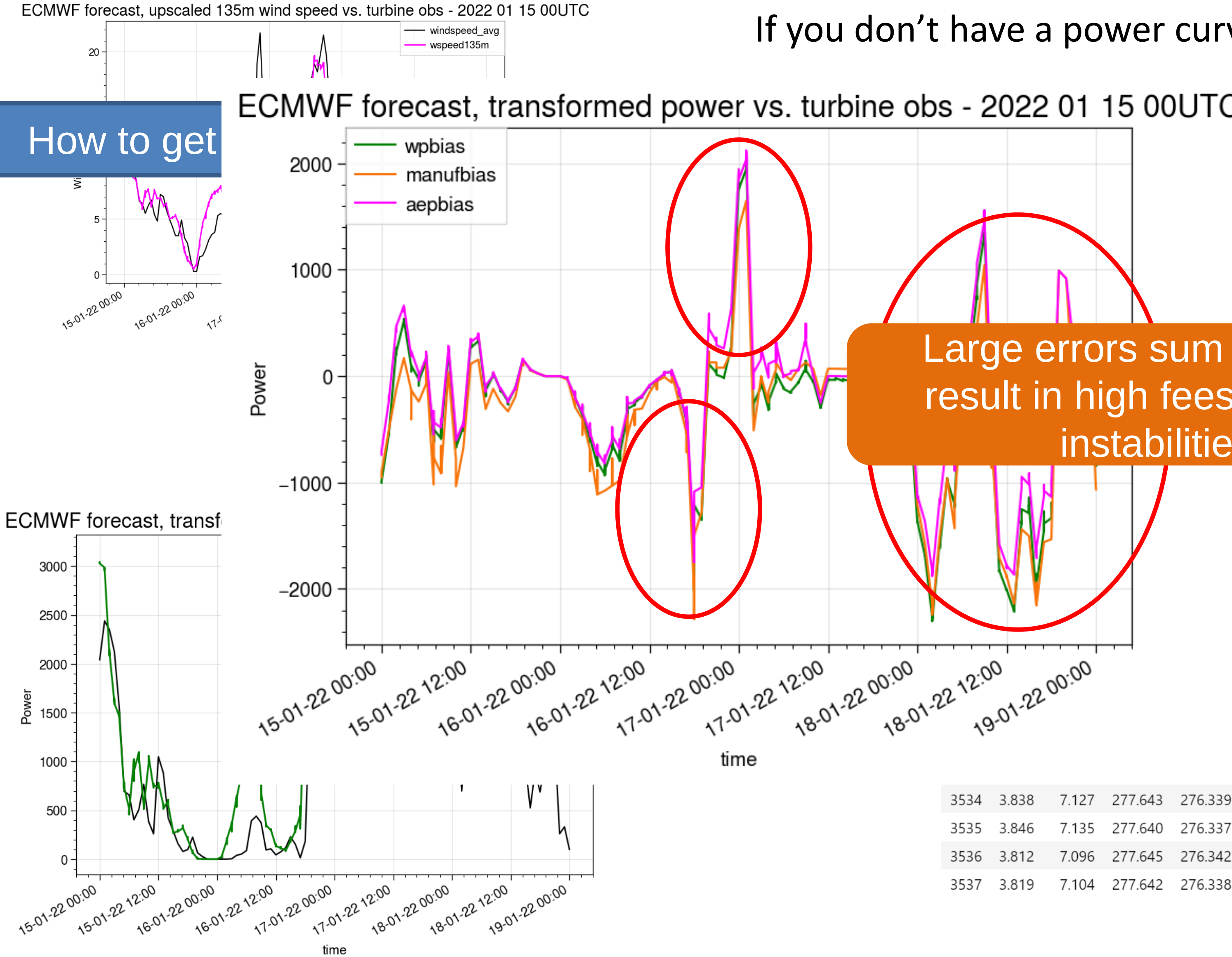
Post-processing for wind energy applications / forecasting – simple approach

If you don't have a power curve…

How to get

some nifty

teorological
eal or



Large errors sum up & can result in high fees and grid instabilities

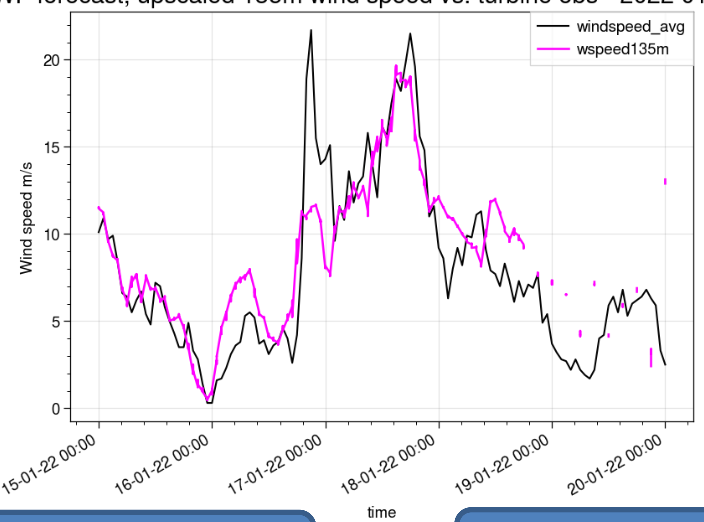
non

					ressure
					01397.302
					01390.691
					...
					02021.643
3534	3.838	7.127	277.643	276.339	102027.564
3535	3.846	7.135	277.640	276.337	102031.968
3536	3.812	7.096	277.645	276.342	102026.422
3537	3.819	7.104	277.642	276.338	102032.449

ZAMG

Post-processing for wind energy applications / forecasting – simple approach

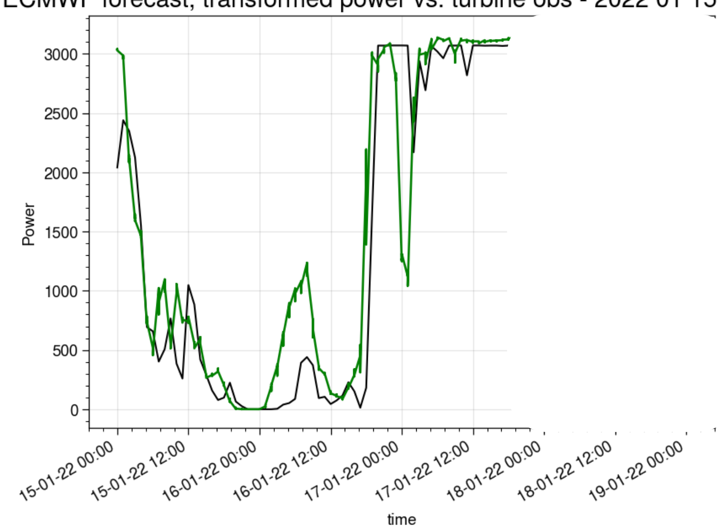
ECMWF forecast, upscaled 135m wind speed vs. turbine obs - 2022 01 15 00UTC



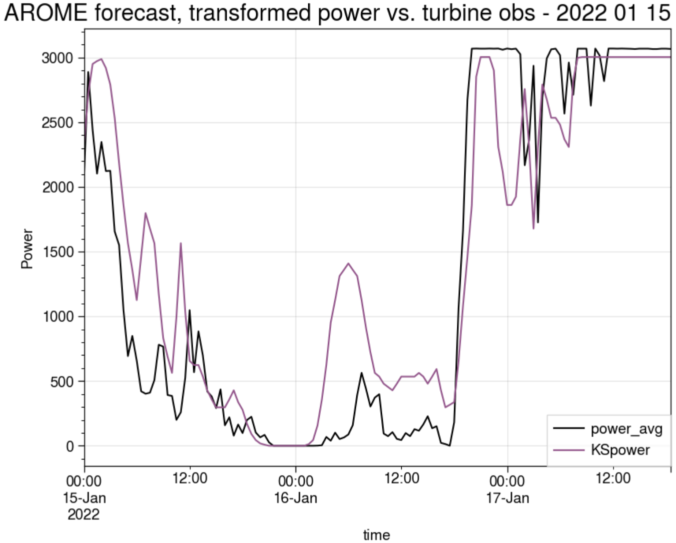
9 km model – wind 2 power

2.5 km model – wind 2 power

ECMWF forecast, transformed power vs. turbine obs - 2022 01 15 00UTC



AROME forecast, transformed power vs. turbine obs - 2022 01 15 00UTC



Machine Learning and statistics – not-so-simple PP

statistical probabilistic post-processing/forecasting EMOS/gEMOS/SAMOS @ ZAMG

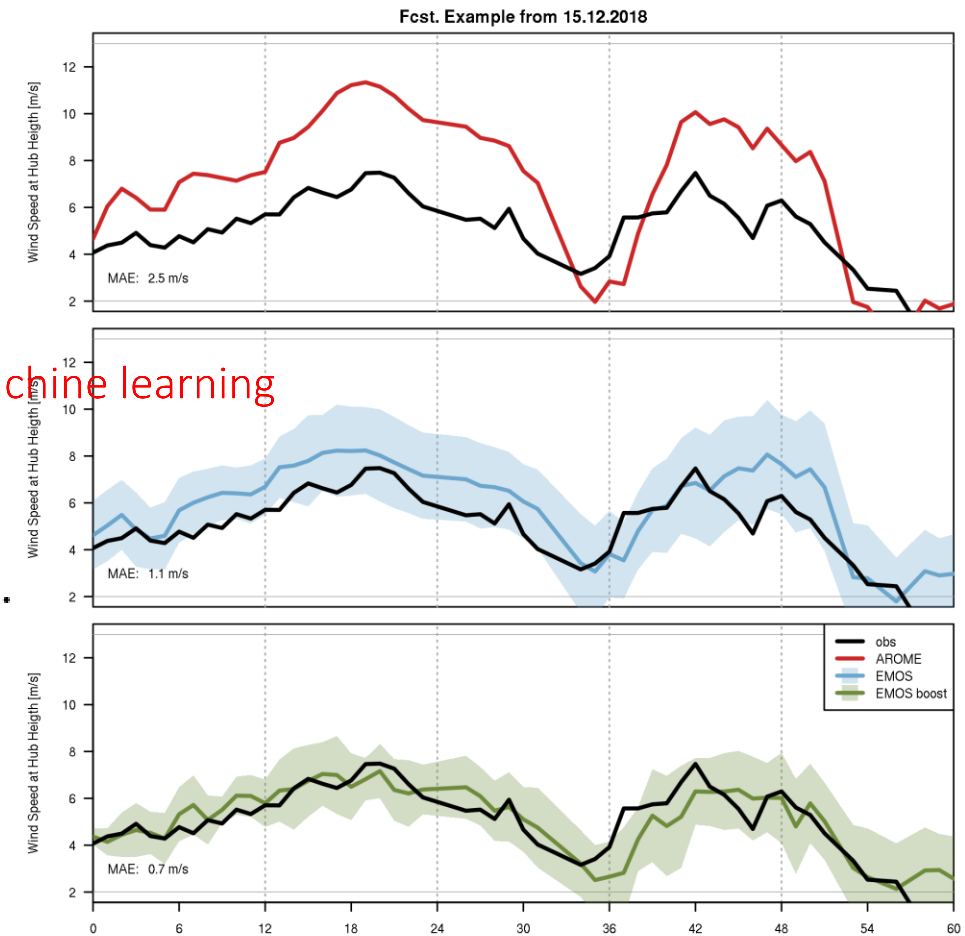
EMOS:

$$y \sim N(\mu, \sigma)$$
$$\mu = b_0 + b_1 \mathrm{m}$$
$$\sigma = c_0 + c_1 \mathrm{s}$$

EMOS boost:

Boosting: adds some machine learning

$$y \sim N(\mu, \sigma)$$
$$\mu = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots$$
$$\sigma = c_0 + c_1 z_1 + c_2 z_2 + c_3 z_3 + \ldots$$

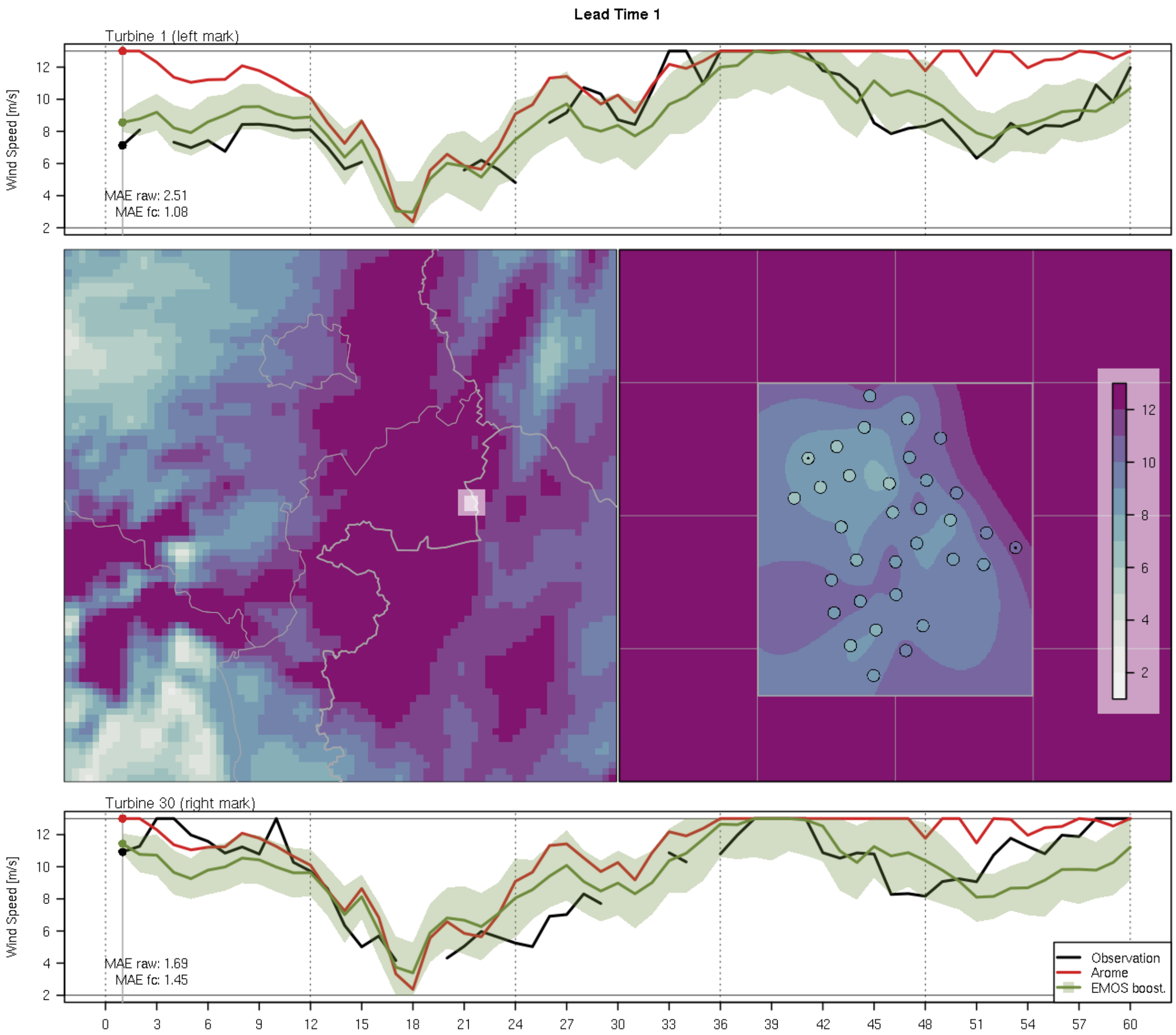


Advantage: we get an uncertainty estimation on-the-fly with the statistical-based method

Machine Learning and statistics – not-so-simple PP

statistical probabilistic post-processing/forecasting EMOS/gEMOS/SAMOS @ ZAMG

Advantage: could be Extended to +90 „easily“

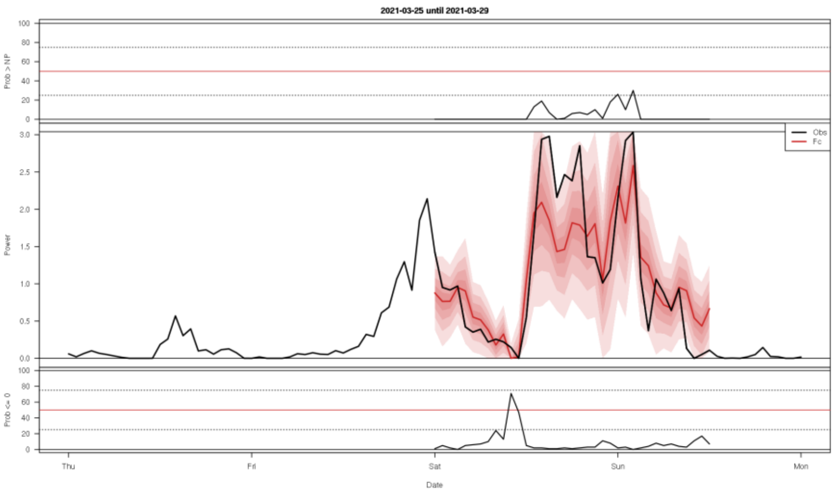


ZAMG

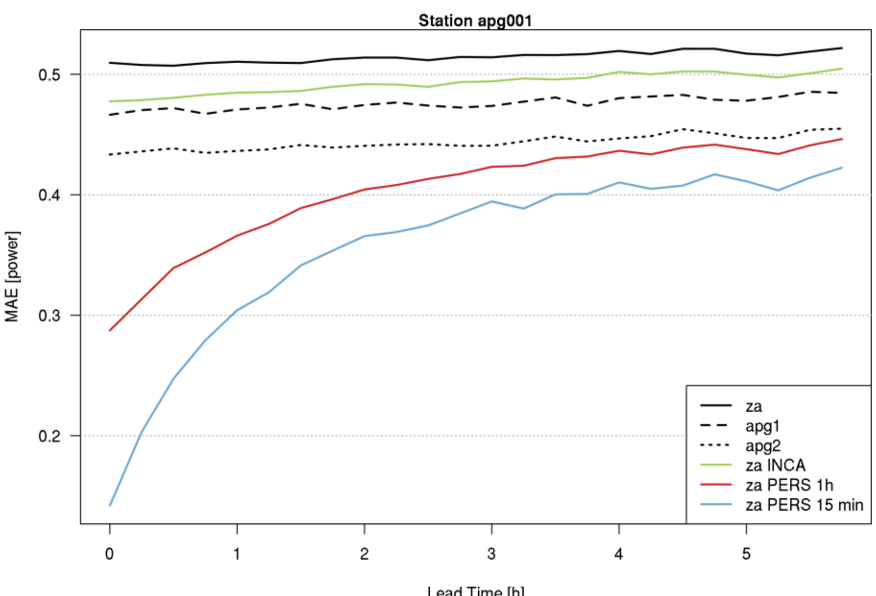
Machine Learning and statistics – not-so-simple PP

statistical probabilistic post-processing/forecasting EMOS/gEMOS/SAMOS @ ZAMG

Ramping event:

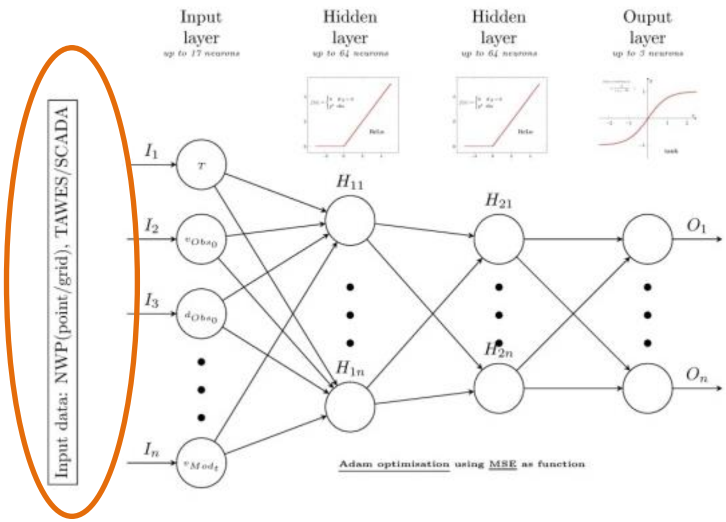


Importance of observations



ZAMG

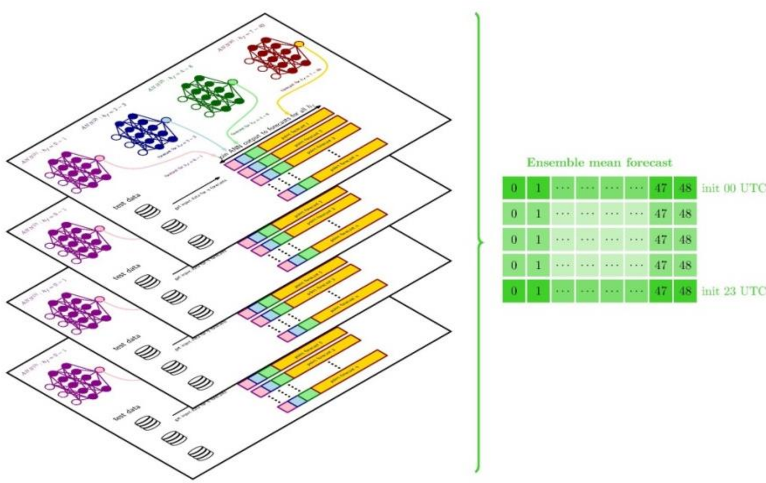
Machine Learning and statistics - machine learning methods
semi-operational model ZiANN (ZAMG interval Artificial Neural Network)



- Hourly forecasts for the next 48 hours ahead
- Uses a neural network in “ensemble mode” (deterministic forecast) and a random forest
- Forecast right now deterministic, probabilistic on the way

Skills:

- Direct access to “online” SCADA data
- In-built QC (AEP and range control)
- Adjustable forecast intervals, neurons, layers, etc.
- Adjustable training length depending on data availability



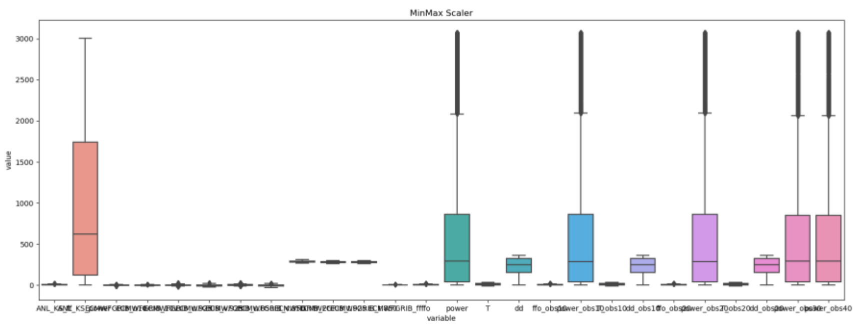
Challenges:

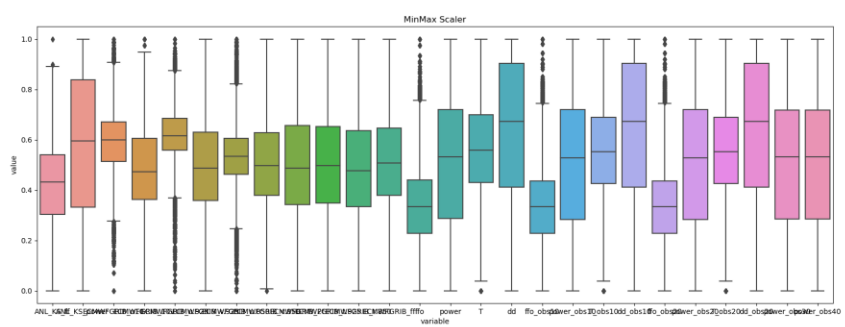
- Meteorological obs data available every 10-minutes with only a tiny delay → SCADA data can have a delay of up to 2 hours or even more
- NWP data so far with a (semi-)large delay of up to 7 hours
- Non-convection permitting models are easy to learn of, don‘t need long time series of data – convection permitting models not, need lots of data
- Changes in the NWP model – how to deal with them? After 3 – 4 years a model changes nearly completely

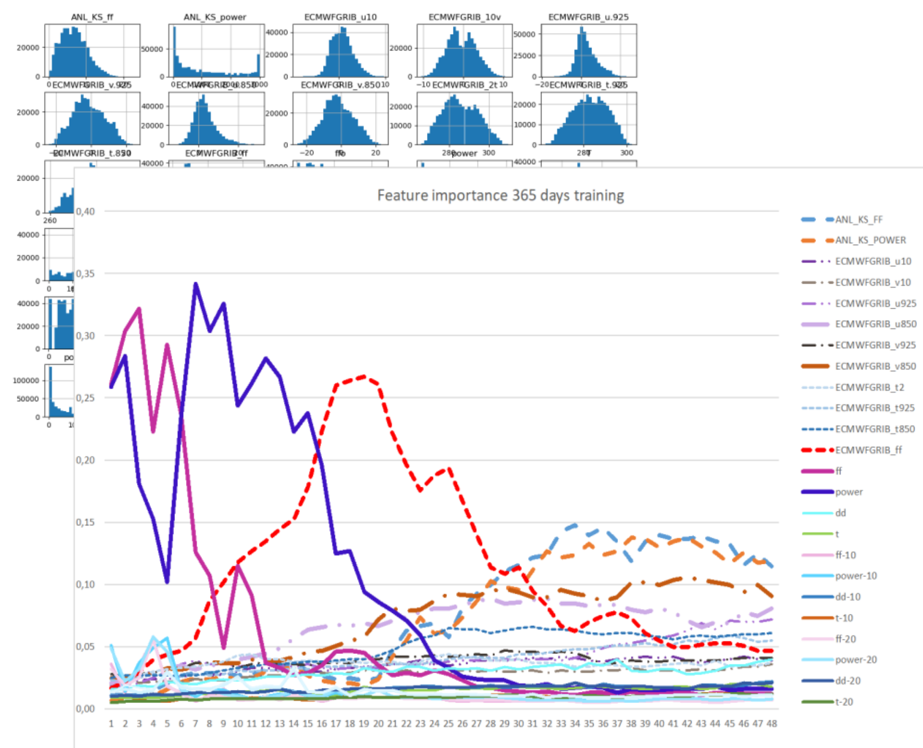
ZAMG

Machine Learning and statistics - – **not-so-simple PP**
machine learning methods

- Input data:
  - can not use data „as is“ → units of parameters, need scaling
  - Power transformation → ML methods like gaussian / norm data. Need to transform skewed parameters
  - (basic) quality control algorithms
  - Periodically updating the annual energy prodction curves

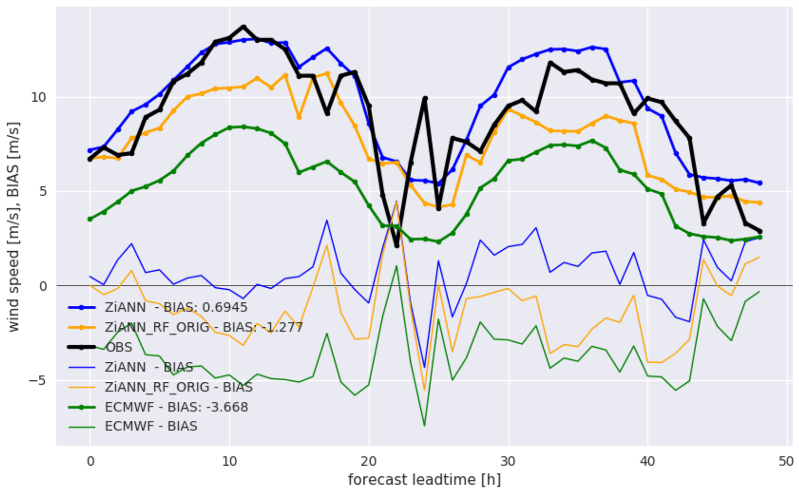




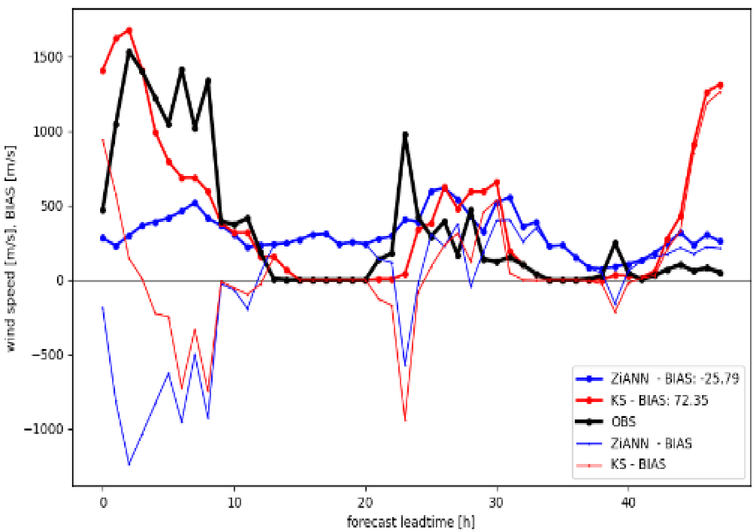


Machine Learning and statistics - machine learning methods
semi-operational model ZiANN (ZAMG interval Artificial Neural Network)

Example wind turbine wind speed non-Austrian, two different machine learning algorithms



Example wind turbine wind power



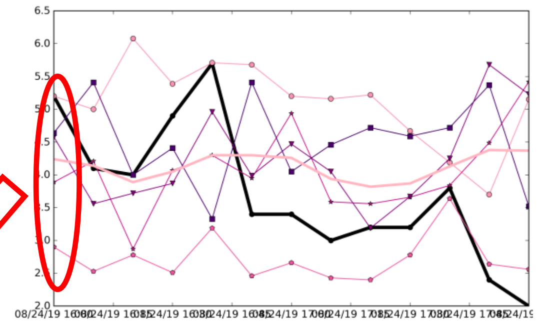
ZAMG

Machine Learning and statistics - machine learning methods
experimental nowcasting model

Data

Observations only

(changes in future when AROME-RUC available)



- Sub-hourly forecasts for the next ~ 6 hours
- Ensemble of methods (in ensemble mode
- Forecast

Perturbing observations for ensemble

Feature selection (LASSO, XGBoost, Random Forest)

Ensemble nowcasting methods (single/multiple selections possible):

- Multilinear regression
- SVR (grid searched)
- **Random Forest**
- XGBoost
- **FF ANN**
- Complex NN
- **Monte Carlo**
- Stochastic Noise Forecast
- **LightGBM**
- Gradient Boosting
- .....

Forecasts: up to 3 hours, 5 – 15-min. frequency

Skills:

- Direct access to "online" SCADA data
- In-built QC (AEP and range control)
- Adjustable
- So far no NWP model (AROME-RUC next step)

Challenges:

- Data availability
- Needs a large amount of observation data
- Could synthetic data be a solution?

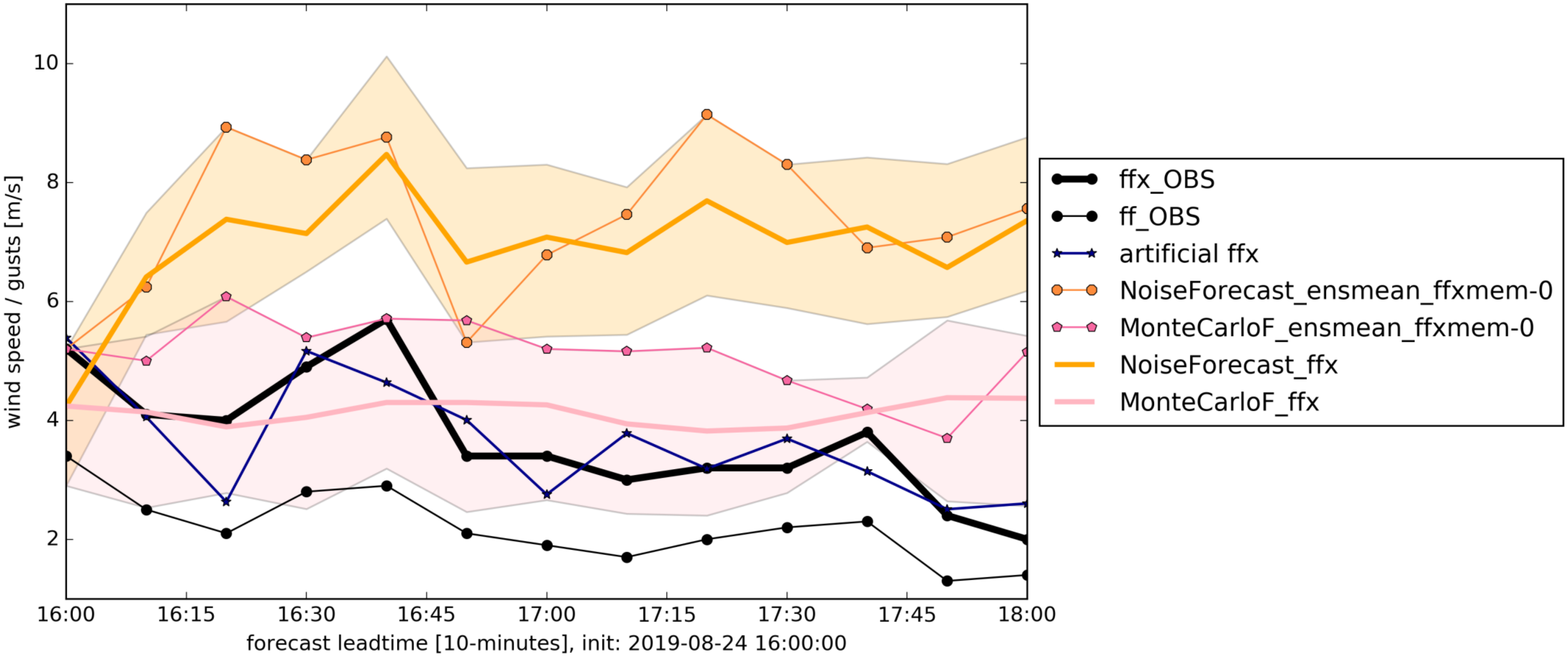
ZAMG

Machine Learning and statistics - machine learning methods
experimental nowcasting model

meteorological observation site Wien Hohe Warte, forecast of 24.08.2019, init at 16 UTC

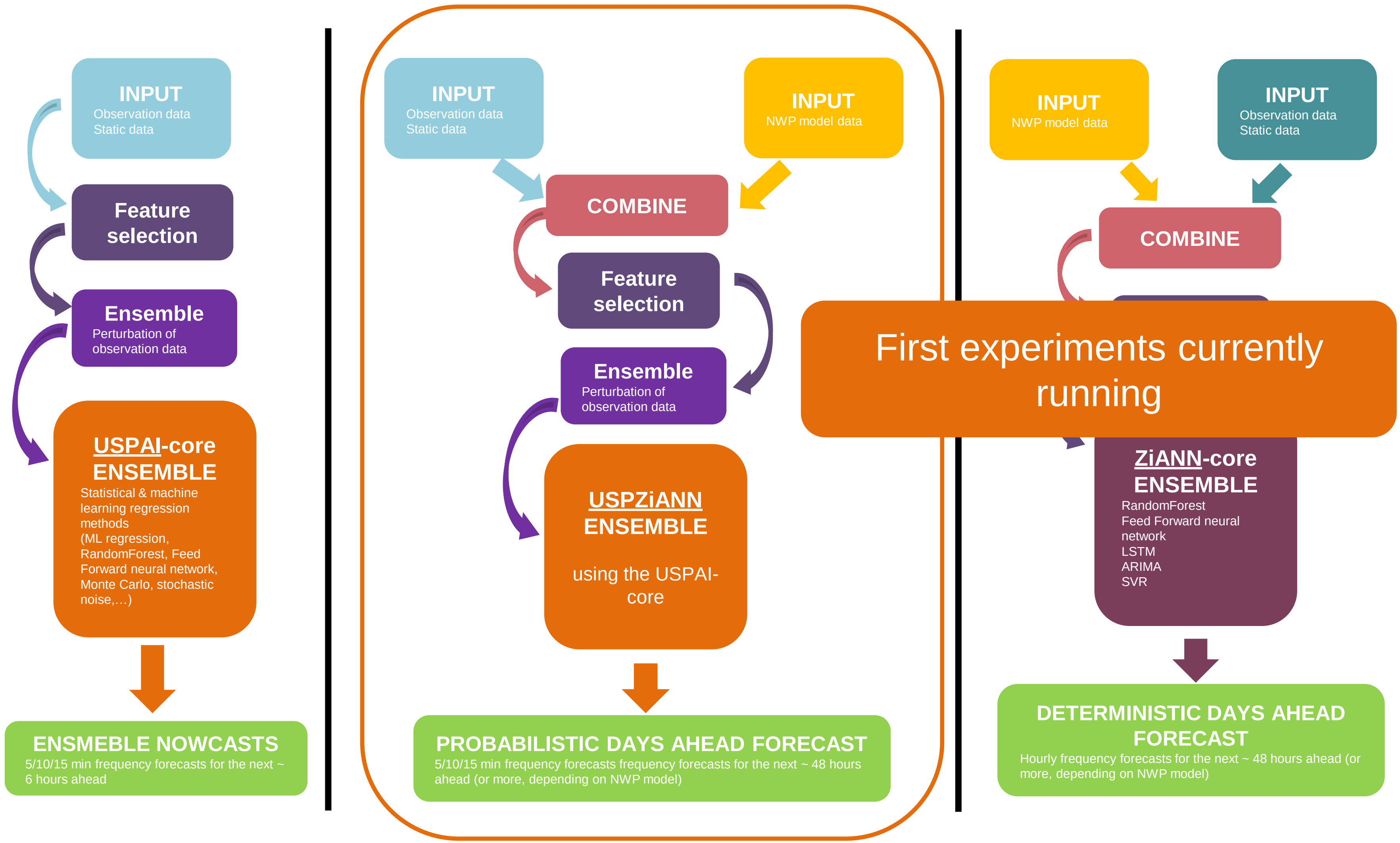
artificial gusts used in training&forcast, measured plotted

Newly developed for wind turbine gust estimation



ZAMG

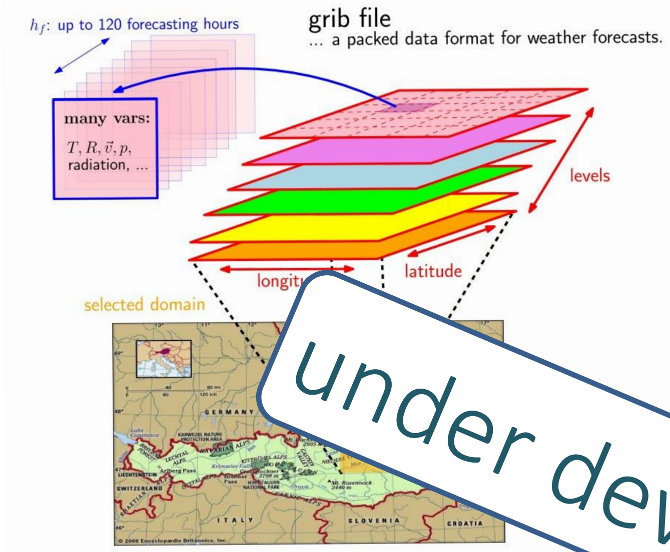
Machine Learning and statistics - machine learning methods
experimental sub-hourly medium range predictions

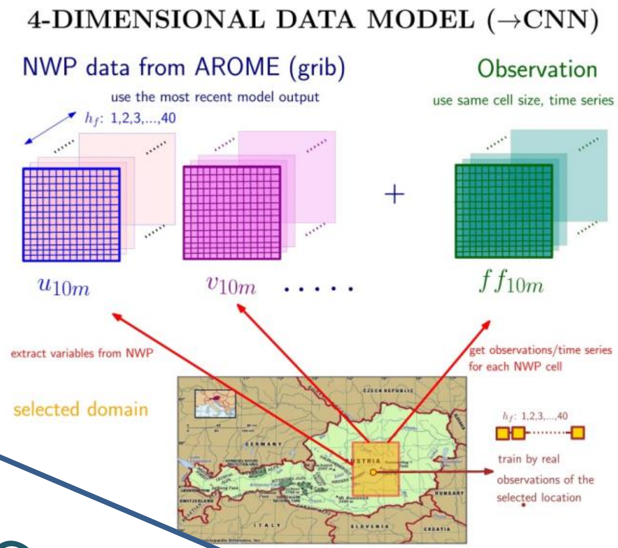


ZAMG

Machine Learning and statistics - machine learning methods
Future: (sub-) km-scale spatial predictions of wind speed for improved power

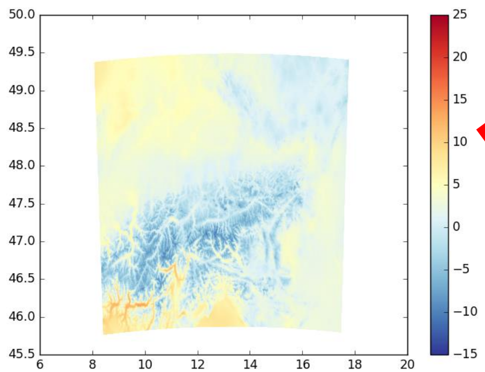
Point forecast using complex neural network setup and multiple data sources (PhD project, AWAkE):

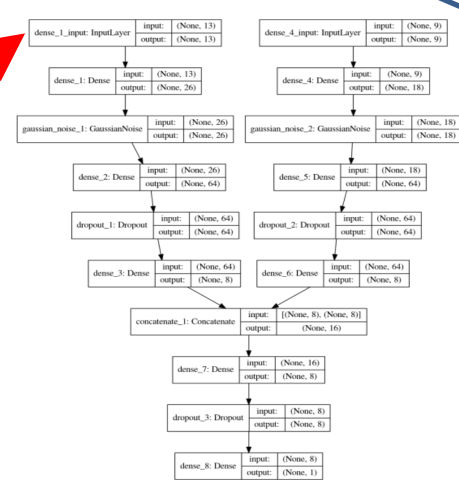


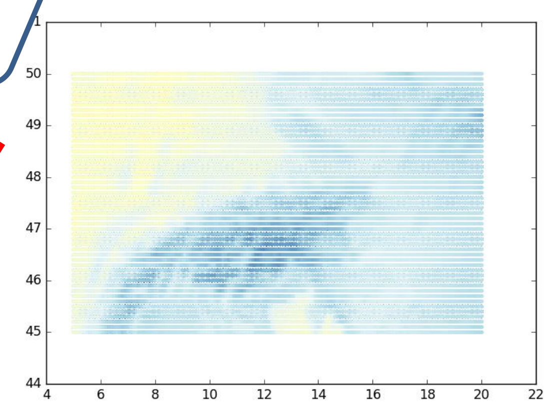


under development

Gridded forecast using complex neural network se... sources:







ZAMG

Methods, data, tools – link selection

Data wind power:

- https://openenergy-platform.org/dataedit/view/supply/wind_turbine_library - contains tons of information on wind turbine types and power curves
- https://www.thewindpower.net/ - country and turbine data, includes maps of known locations per country and, if known, also turbine type and installed total capacity per wind farm → not free but check the site source code

Data meteorology:

- renewables.ninja → wind and solar power, via API weather, pv and wind production data downloadable based on MERRA2 and for some dates SARAH, up to 2019
- https://power.larc.nasa.gov/data-access-viewer/ → Nasa weather data, similar to renewables.ninja, nearly real-time. More parameters accessible
- ECMWF, GFS, ICON, … NWP forecasts
- For synthetic data: ERA5, MERRA2, COSMO RE, …

Usefule python libs and notebook:

- Windpowerlib
- pyWAKE
- Github e.g. windtools, ..
- Data exploitation SCADA: https://www.kaggle.com/winternguyen/wind-power-curve-modeling

→ Check units!

ZAMG

Thank you for your attention!

irene.schicker@zamg.ac.at