

# First Insights into Wind Farm Data Analysis

**Tobias Braun**

Group Seminar  
AG Guhr



5. Januar 2017

Motivation

General Facts & Literature Overview

Data Cleansing

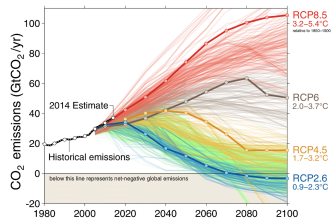
First Glance of Qualitative Behaviour

Ideas and Prospect

# Motivation

# Why should one be interested in wind farm data?

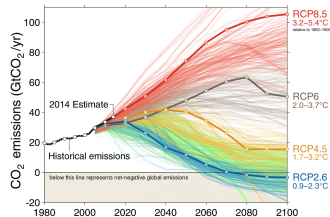
- **Climate Change:**  
to achieve  $< 2^{\circ}\text{C}$ -goal,  
coal demand has to decrease  
by 40% between 2012 and 2040



Source: IPCC report

# Why should one be interested in wind farm data?

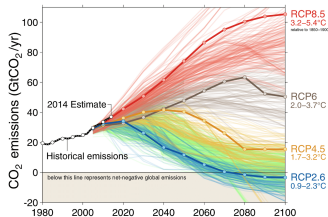
- **Climate Change:**  
to achieve  $< 2^{\circ}\text{C}$ -goal,  
coal demand has to decrease  
by 40% between 2012 and 2040
- **Today:**  
 $\text{CO}_2$ -emissions still **increasing**



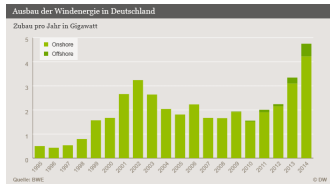
Source: IPCC report

# Why should one be interested in wind farm data?

- Climate Change:**  
 to achieve  $< 2^{\circ}\text{C}$ -goal,  
 coal demand has to decrease  
 by 40% between 2012 and 2040
- Today:**  
 $\text{CO}_2$ -emissions still **increasing**
- Emissions:**  
 $20 \frac{\text{g}}{\text{kWh}}$  vs.  $1000 \frac{\text{g}}{\text{kWh}}$  of  $\text{CO}_2$  by  
 wind energy resp. coal

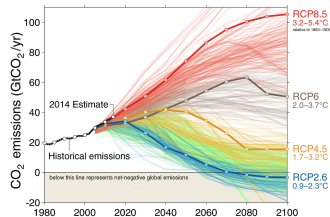


Source: IPCC report



# Why should one be interested in wind farm data?

- Climate Change:**  
 to achieve  $< 2^{\circ}\text{C}$ -goal,  
 coal demand has to decrease  
 by 40% between 2012 and 2040
- Today:**  
 $\text{CO}_2$ -emissions still **increasing**
- Emissions:**  
 $20 \frac{\text{g}}{\text{kWh}}$  vs.  $1000 \frac{\text{g}}{\text{kWh}}$  of  $\text{CO}_2$  by  
 wind energy resp. coal
- Highest **expansion potential** and  
 theoretical **effectiveness** up to 59%

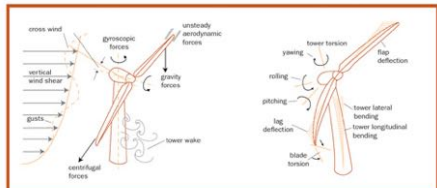


Source: IPCC report



## Why should **we** be interested in wind farm data?

- *Single wind turbine:*  
highly complex system with a lot of analysable properties
- *Wind farm:*  
compound complex system with certain synchronicities

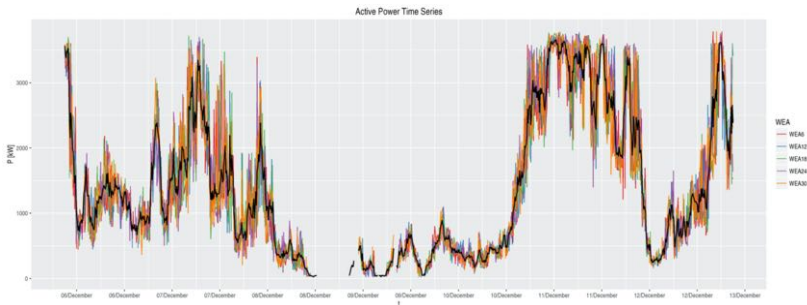


Source: Kießling, F. Modellierung des anelastischen Gesamtsystems einer Windturbin mit Hilfe symbolischer Programmierung. DFVLR-Report, DFVLRFB 84-10, 1984.



## Why should **we** be interested in wind farm data?

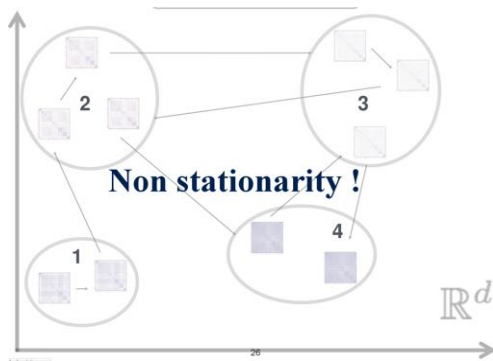
- *Single wind turbine*:  
highly complex system with a lot of analysable properties
- *Wind farm*:  
compound complex system with certain synchronicities
- Time series of several properties have **stochastic**, **non-stationary** and **quasi-periodical** character



→ **Present problems**: volatile power output and grid overload

# What am I planning to do with wind farm data?

Find **quasistationary states** of wind farm with methods analogue to *Market States*!



Y. Stepanov: Meta-Stable Stock Market Collective Dynamics with Applications, presentation 2016

# General Facts & Literature Overview

# General Physics of Wind Turbines

## An easy estimation of Power Production

**Power of wind** results from kinetic energy of moved air:

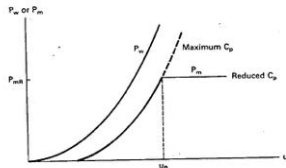
$$E_{\text{kin}} = \frac{m}{2} u^2 \quad \Leftrightarrow \quad P_{\text{W}} = \dot{E}_{\text{kin}} = \frac{\dot{m}}{2} u^2 \quad \text{with} \quad u = \text{const.}$$

The *mass flow*  $\dot{m} = A\rho u$  gives us

$$P_{\text{W}} = \frac{A}{2} \rho u^3$$

**Mechanical power** can be estimated by including a performance constant  $C_P$ :

$$P_{\text{m}}(u) = C_P P_{\text{W}} = C_P \frac{A}{2} \rho u^3$$



# General Physics of Wind Turbines

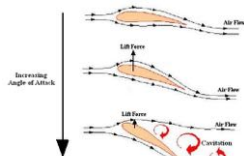
## Betz's law

Rotor blades are driven by *lift force*

→ **deceleration** of wind

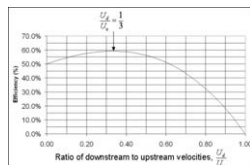
- Very weak deceleration: weak lift force
- Very strong deceleration:

$$\dot{m} \rightarrow 0 \quad \Rightarrow \quad P_m \rightarrow 0$$



*Conclusion:* There is a best value for  $\frac{u_{11}}{u_d}$  and a **limit** for aerodynamical efficiency!

$$C_P \leq 0.59$$



# Control of Wind Farms

## SCADA and power control

### SCADA

**S**upervisory **C**ontrol and **D**ata **A**cquisition

Collects real-time data to control the wind farm (mostly fully automatic)

# Control of Wind Farms

## SCADA and power control

### SCADA

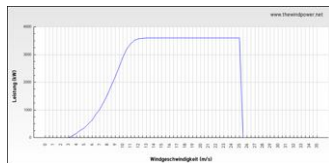
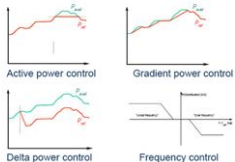
Supervisory **C**ontrol and **D**ata **A**cquisition

Collects real-time data to control the wind farm (mostly fully automatic)

### Active Power Control

Data is used to limit power within **operating range**:

SCADA Control Functions  
For Improved Grid Operations



# Control of Wind Farms

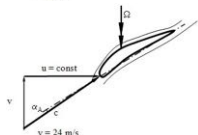
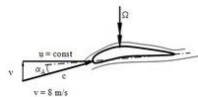
How it is realized

## Pitch Control

$$u < 3 \frac{\text{m}}{\text{s}} \quad \text{or} \quad u > 25 \frac{\text{m}}{\text{s}} : \quad \phi \approx 90^\circ$$

$$3 \frac{\text{m}}{\text{s}} \leq u \leq 12 \frac{\text{m}}{\text{s}} : \quad \phi = 0^\circ$$

$$12 \frac{\text{m}}{\text{s}} < u \leq 25 \frac{\text{m}}{\text{s}} : \quad \phi \in [0^\circ, 30^\circ]$$





# Control of Wind Farms

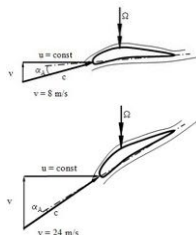
How it is realized

## Pitch Control

$$u < 3 \frac{\text{m}}{\text{s}} \quad \text{or} \quad u > 25 \frac{\text{m}}{\text{s}} : \quad \phi \approx 90^\circ$$

$$3 \frac{\text{m}}{\text{s}} \leq u \leq 12 \frac{\text{m}}{\text{s}} : \quad \phi = 0^\circ$$

$$12 \frac{\text{m}}{\text{s}} < u \leq 25 \frac{\text{m}}{\text{s}} : \quad \phi \in [0^\circ, 30^\circ]$$



## Autotracking and Trundeling

**Autotracking:** within operating range, nacelle rotates so that  $\vec{u} \parallel \vec{A}$

**Trundeling:** beyond operating range, turbine is not shut down immediately but *trundles*

# What are the main problems and approaches in wind energy research?

Brief Overview by Keywords



# Popular Approaches

What already has been done

- **Aggregated windfarm models:**
  - Reducing complexity & computation times by determining multi machine representations (since  $\approx 2000$ )
  - *Methods:*  
Clustering of wind velocities and further data mining methods

# Popular Approaches

What already has been done

- **Aggregated windfarm models:**
  - Reducing complexity & computation times by determining multi machine representations (since  $\approx 2000$ )
  - *Methods:*  
Clustering of wind velocities and further data mining methods
- **Power forecast:**
  - Several approaches to predict wind power from empirical or simulated data to e.g. minimize number of shutdowns (well advanced)
  - *Methods:*  
mostly sophisticated big data and machine learning methods

# Popular Approaches

What already has been done

- **Reliability Assessment:**
  - Comprehensive analysis of wind energy reliability by including multiple wind farms
  - *Methods:*  
Correlations → wind velocity simulations

# Popular Approaches

What already has been done

- **Reliability Assessment:**

- Comprehensive analysis of wind energy reliability by including multiple wind farms
- *Methods:*  
Correlations → wind velocity simulations

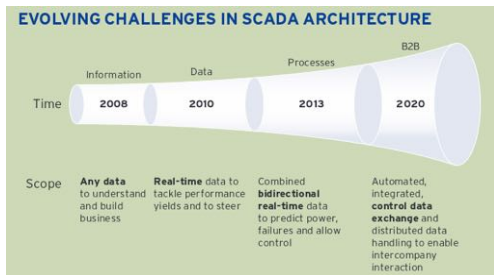
⋮

- **Structural Health Monitoring**
- **Wind Farm Integration into Power Grid**
- **Models for Power Fluctuations**
- **Operational Risk of Wind Farms**

# What are the main problems and approaches in wind energy research?

..and how can I/we contribute?

Improve **grid integration** by **balancing** multiple intermittent power outputs of WTs, especially for **offshore** wind farms.

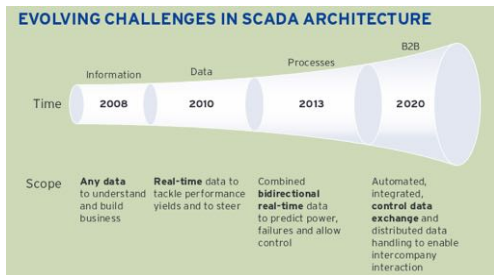


Expert Report: **SCADA 2014 systems for wind**

# What are the main problems and approaches in wind energy research?

..and how can I/we contribute?

Improve **grid integration** by **balancing** multiple intermittent power outputs of WTs, especially for **offshore** wind farms.



Expert Report: **SCADA 2014 systems for wind**

→ Learn more about (instationary) **correlation** structure between single WTs!



# ForWind

Center for Wind Energy Research

- since 2003
- Oldenburg, Bremen and Hannover
- **Research:** many interdisciplinary topics, including turbulence, power forecast and stochastic modeling
  
- **Previous collaborations:** Yuriy and Philip Rinn about *Market States*
- Further collaboration with AG Peinke is intended



# RIFFGAT

## Offshore wind farm near Borkum

- 30 wind turbines with active power of 3.6MW  $\hat{=}$  120.000 provided homes
- Operating range:  $3 - 25 \frac{\text{m}}{\text{s}}$  and  $u < 70 \frac{\text{m}}{\text{s}}$



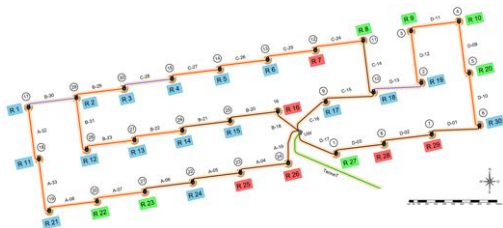
# RIFFGAT

## Offshore wind farm near Borkum

- 30 wind turbines with active power of 3.6MW  $\hat{=}$  120.000 provided homes
- Operating range:  $3 - 25 \frac{m}{s}$  and  $u < 70 \frac{m}{s}$



- Distances: 554m in rows and 600m between rows
- Diameter: 120m
- Height: 90m



# Data Cleansing

# My Dataset

## WTs Measured Quantities as Time Series

---

$$30 \times \left( \begin{array}{c} P \\ u \\ \alpha_{\text{nac}} \\ \phi_{\text{rot}} \\ \Omega_{\text{T}} \\ \vdots \end{array} \right)_{t_0} \left. \vphantom{\begin{array}{c} P \\ u \\ \alpha_{\text{nac}} \\ \phi_{\text{rot}} \\ \Omega_{\text{T}} \\ \vdots \end{array}} \right\} n \approx 20 \rightarrow \left( \begin{array}{c} P \\ u \\ \alpha_{\text{nac}} \\ \phi_{\text{rot}} \\ \Omega_{\text{T}} \\ \vdots \end{array} \right)_{t_1} \rightarrow \dots$$

T = 1 year

including values for mean, maximum, minimum and standard deviation  
 always calculated over consecutive *10-minute intervals*  
 (56.593 values per WT).

# My Dataset

## WTs Measured Quantities as Time Series

---

$$30 \times \left( \begin{array}{c} P \\ u \\ \alpha_{\text{nac}} \\ \phi_{\text{rot}} \\ \Omega_{\text{T}} \\ \vdots \end{array} \right)_{t_0} \left. \vphantom{\begin{array}{c} P \\ u \\ \alpha_{\text{nac}} \\ \phi_{\text{rot}} \\ \Omega_{\text{T}} \\ \vdots \end{array}} \right\} n \approx 20 \rightarrow \left( \begin{array}{c} P \\ u \\ \alpha_{\text{nac}} \\ \phi_{\text{rot}} \\ \Omega_{\text{T}} \\ \vdots \end{array} \right)_{t_1} \rightarrow \dots$$

$T = 1 \text{ year}$

including values for **mean**, maximum, minimum and standard deviation always calculated over consecutive *10-minute intervals* (56.593 values per WT).

# Limiting to Operating Range

Operating Range:

$$\begin{aligned} 3 \frac{\text{m}}{\text{s}} &\leq u \leq 25 \frac{\text{m}}{\text{s}} \\ 20 \text{kW} &\leq P \leq 3800 \text{kW} \end{aligned}$$

# Limiting to Operating Range

## Operating Range:

$$\begin{aligned} 3 \frac{\text{m}}{\text{s}} &\leq u \leq 25 \frac{\text{m}}{\text{s}} \\ 20\text{kW} &\leq P \leq 3800\text{kW} \end{aligned}$$

## Raw Data:

- 16% of  $P_{\text{raw}} < 0\text{kW}$ !  
But only 2% of  $P_{\text{raw}} < -20\text{kW}$
- 10% of  $u_{\text{raw}} < 3 \frac{\text{m}}{\text{s}}$ !  
But only 0.1% of  $u_{\text{raw}} > 25 \frac{\text{m}}{\text{s}}$



# Limiting to Operating Range

## Operating Range:

$$3 \frac{\text{m}}{\text{s}} \leq u \leq 25 \frac{\text{m}}{\text{s}}$$

$$20\text{kW} \leq P \leq 3800\text{kW}$$

## Raw Data:

- 16% of  $P_{\text{raw}} < 0\text{kW}$ !  
But only 2% of  $P_{\text{raw}} < -20\text{kW}$
- 10% of  $u_{\text{raw}} < 3 \frac{\text{m}}{\text{s}}$ !  
But only 0.1% of  $u_{\text{raw}} > 25 \frac{\text{m}}{\text{s}}$

Further discard redundant values and such with  $\text{stddev} = 0$ :

⇒ 32.42% NAs

# Non-physical Relative Changes of Power Output

## Insight into questionable values

Abs. Differenz/3600kW	Relat. Aenderung	P1	P2	Min1	Max1	Min2	Max2	Stddev1	Stddev2
0.9919554	117.7760137	30.3206000	3601.3600000	-718.0000000	2438.0000000	3508.0000000	3698.0000000	283.7120000	26.7370000
0.2458506	38.8493541	22.7819000	907.8440000	-95.0000000	1538.0000000	608.0000000	1396.0000000	194.7060000	149.3420000
0.1830456	24.0896190	27.3547000	686.3190000	-40.0000000	161.0000000	576.0000000	745.0000000	38.0380000	31.5924000
0.2426734	37.3410698	23.3958000	897.0200000	-60.0000000	330.0000000	796.0000000	1029.0000000	101.2240000	45.0744000
0.9326092	24.4695460	137.2070000	3494.6000000	-132.0000000	367.0000000	2103.0000000	3697.0000000	163.8960000	291.3220000
0.2235168	39.1688133	20.5434000	825.2040000	-58.0000000	178.0000000	746.0000000	938.0000000	57.0338000	42.0686000
0.2603715	32.5061400	28.8357000	966.1730000	-91.0000000	368.0000000	-97.0000000	2695.0000000	118.2420000	912.9050000

## Possible Explanation & Rejection Criterion

# Non-physical Relative Changes of Power Output

## Insight into questionable values

Abs. Differenz/3600kW	Relat. Aenderung	P1	P2	Min1	Max1	Min2	Max2	Stddev1	Stddev2
0.9919554	117.7760137	30.3206000	3601.3600000	-718.0000000	2438.0000000	3509.0000000	3698.0000000	283.7120000	26.7370000
0.2458506	38.8493541	22.7819000	907.8440000	-95.0000000	1538.0000000	608.0000000	1396.0000000	194.7060000	149.3420000
0.1830456	24.0896190	27.3547000	686.3190000	-40.0000000	161.0000000	576.0000000	745.0000000	38.0380000	31.5924000
0.2426734	37.3410698	23.3958000	897.0200000	-60.0000000	330.0000000	796.0000000	1029.0000000	101.2240000	45.0744000
0.9326092	24.4695460	137.2070000	3494.6000000	-132.0000000	367.0000000	2103.0000000	3697.0000000	163.8960000	291.3220000
0.2235168	39.1688133	20.5434000	825.2040000	-58.0000000	178.0000000	746.0000000	938.0000000	57.0338000	42.0686000
0.2603715	32.5061400	28.8357000	966.1730000	-91.0000000	368.0000000	-97.0000000	2695.0000000	118.2420000	912.9050000

## Possible Explanation & Rejection Criterion

# Non-physical Relative Changes of Power Output

## Insight into questionable values

Abs. Differenz/3600kW	Relat. Aenderung	P1	P2	Min1	Max1	Min2	Max2	Stddev1	Stddev2
0.9919554	117.7760137	30.3206000	3601.3600000	-718.0000000	2438.0000000	3509.0000000	3698.0000000	283.7120000	26.7370000
0.2458506	38.8493541	22.7819000	907.8440000	-95.0000000	1538.0000000	608.0000000	1396.0000000	194.7060000	149.3420000
0.1830456	24.0896190	27.3547000	686.3190000	-40.0000000	161.0000000	576.0000000	745.0000000	38.0380000	31.5924000
0.2426734	37.3410698	23.3958000	897.0200000	-60.0000000	330.0000000	796.0000000	1029.0000000	101.2240000	45.0744000
0.9326092	24.4695460	137.2070000	3494.6000000	-132.0000000	367.0000000	2103.0000000	3697.0000000	163.8960000	291.3220000
0.2235168	39.1688133	20.5434000	825.2040000	-58.0000000	178.0000000	746.0000000	938.0000000	57.0338000	42.0686000
0.2603715	32.5061400	28.8357000	966.1730000	-91.0000000	368.0000000	-97.0000000	2695.0000000	118.2420000	912.9050000

## Possible Explanation & Rejection Criterion

Already discarded values could cause large changes → **partially verified**

# Non-physical Relative Changes of Power Output

## Insight into questionable values

Abs. Differenz/3600kW	Relat. Aenderung	P1	P2	Min1	Max1	Min2	Max2	Stddev1	Stddev2
0.9919554	117.7760137	30.3206000	3601.3600000	-718.0000000	2438.0000000	3509.0000000	3698.0000000	283.7120000	26.7370000
0.2458506	38.8493541	22.7819000	907.8440000	-95.0000000	1538.0000000	608.0000000	1396.0000000	194.7060000	149.3420000
0.1830456	24.0896190	27.3547000	686.3190000	-40.0000000	161.0000000	576.0000000	745.0000000	38.0380000	31.5924000
0.2426734	37.3410698	23.3958000	897.0200000	-60.0000000	330.0000000	796.0000000	1029.0000000	101.2240000	45.0744000
0.9326092	24.4695460	137.2070000	3494.6000000	-132.0000000	367.0000000	2103.0000000	3697.0000000	163.8960000	291.3220000
0.2235168	39.1688133	20.5434000	825.2040000	-58.0000000	178.0000000	746.0000000	938.0000000	57.0338000	42.0686000
0.2603715	32.5061400	28.8357000	966.1730000	-91.0000000	368.0000000	-97.0000000	2695.0000000	118.2420000	912.9050000

## Possible Explanation & Rejection Criterion

**Criterion:** discard all values with  $|\rho| > 10$  that show jump

$$P_{\min}(t_2) - P_{\max}(t_1) > 360\text{kW} \hat{=} 0.1P_{\text{WT}}$$

⇒ 32.45% NAs

but only

0.6% of  $|\Delta P_{\text{WF}}| > 0.2P_{\text{WF}}$

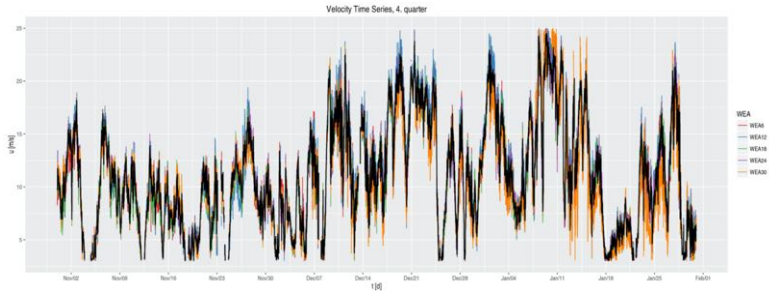
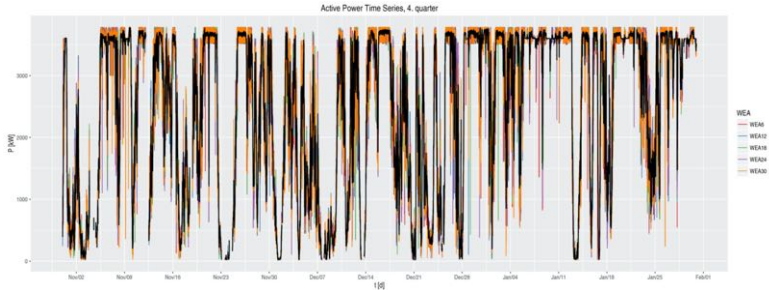
## Remaining Data

We arrive at ⇒ 32.45% missing values  $\hat{=}$  38.246 values per WT.

- Did we sort out *relevant* data?
  - How *reliable* is the remaining data set?
- **How can we deal with all this missing data?**

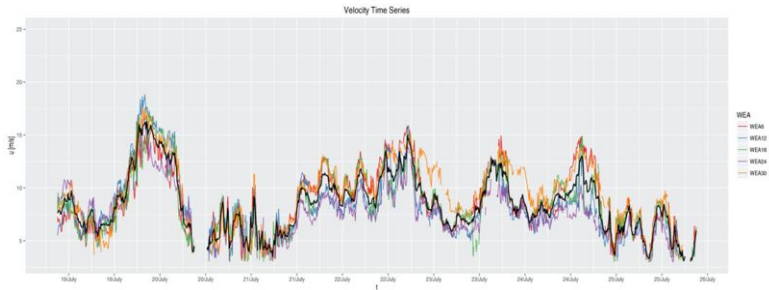
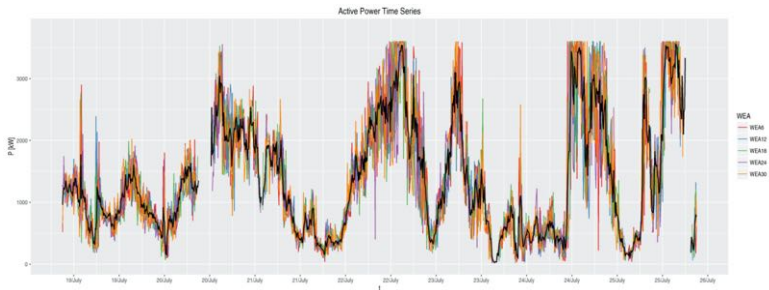
# First Glance of Qualitative Behaviour

# Time Series

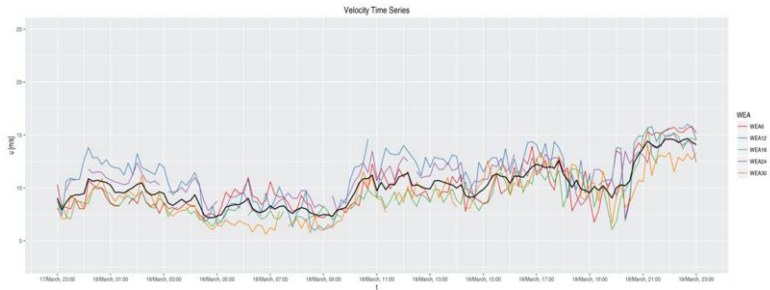




# Time Series



# Time Series

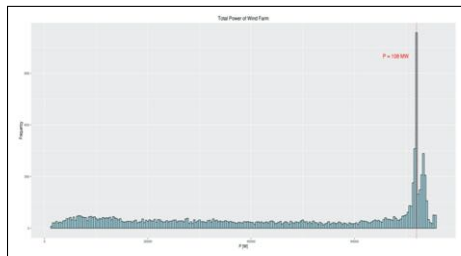


# Distributions

## Power and Velocity

### Power

- sharply massed around actual **rated output**
- lower output approx. equally distributed

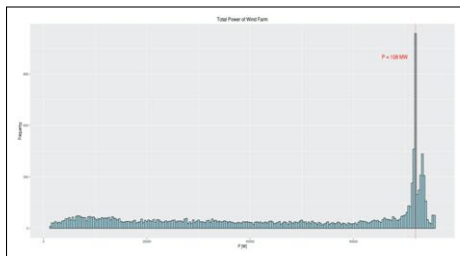


# Distributions

## Power and Velocity

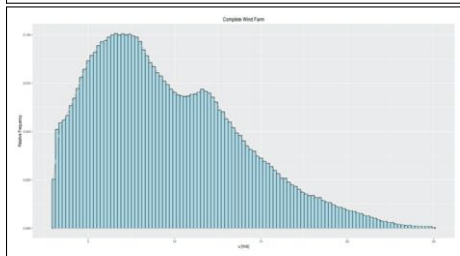
### Power

- sharply massed around actual **rated output**
- lower output approx. equally distributed



### Velocity

- not Weibull-distributed, **bimodal!**
- second maximum because WT's mostly run at nominal power at around  $\approx 12 \frac{m}{s}$

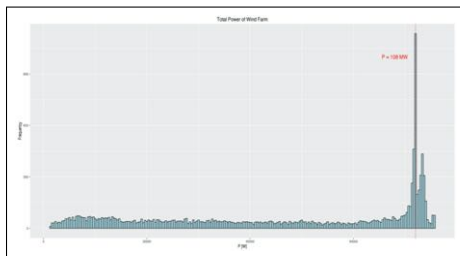


# Distributions

## Power and Velocity

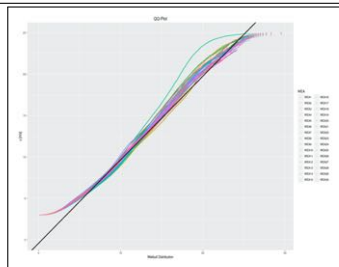
### Power

- sharply massed around actual **rated output**
- lower output approx. equally distributed



### Velocity

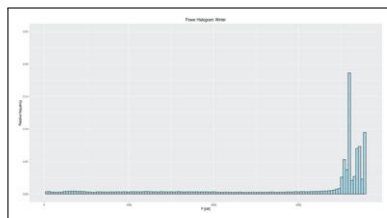
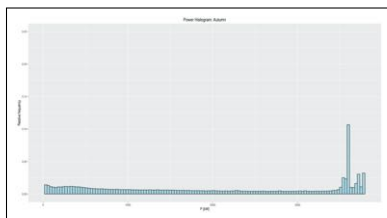
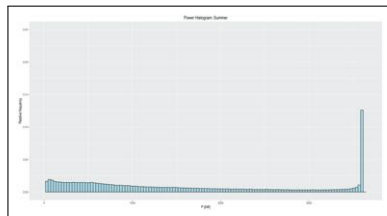
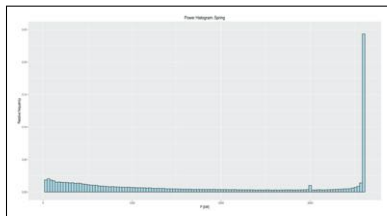
- not Weibull-distributed, **bimodal!**
- second maximum because WTs mostly run at nominal power at around  $\approx 12 \frac{m}{s}$



# Distributions

## Power and Velocity – Seasons of the Year

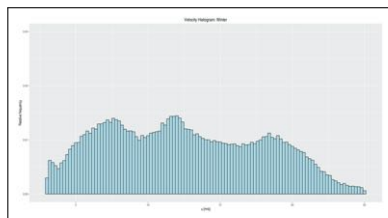
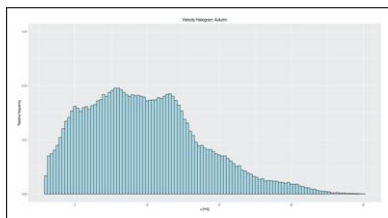
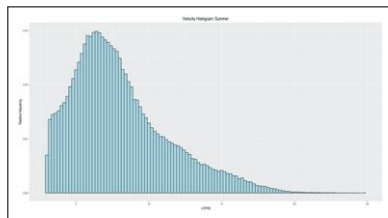
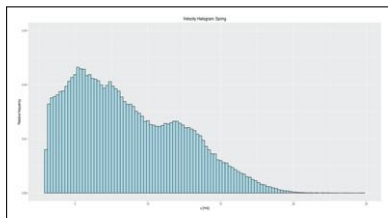
### Power



# Distributions

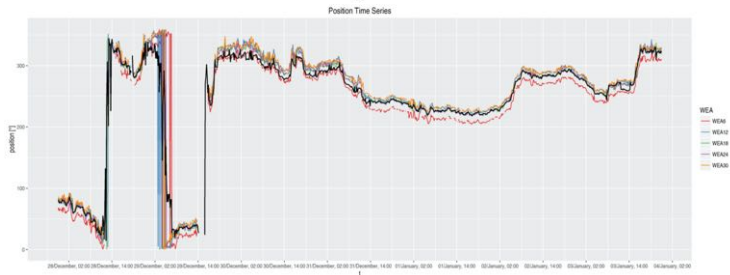
Power and Velocity – Seasons of the Year

## Velocity



# Time Series

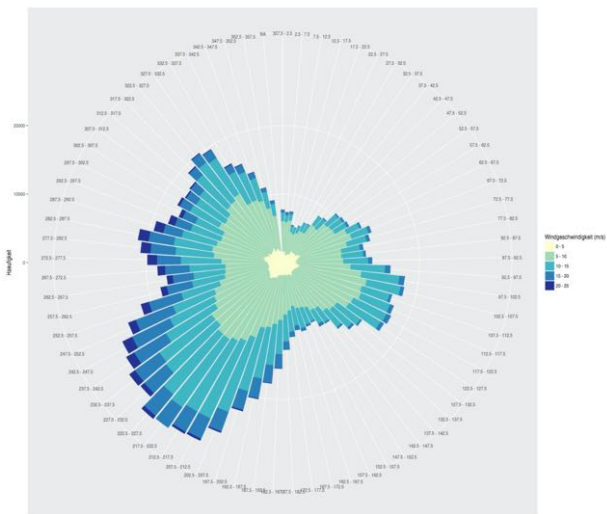
## Nacelle direction





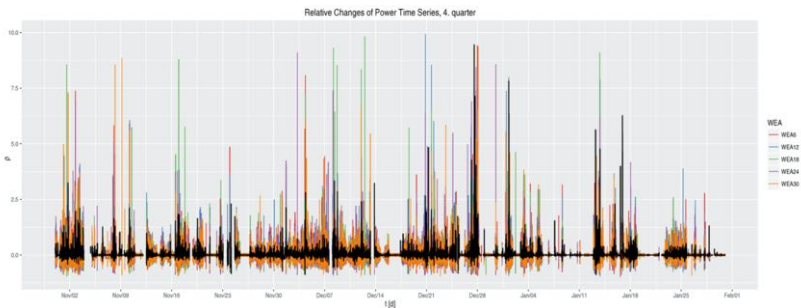
# Distributions

## Nacelle direction



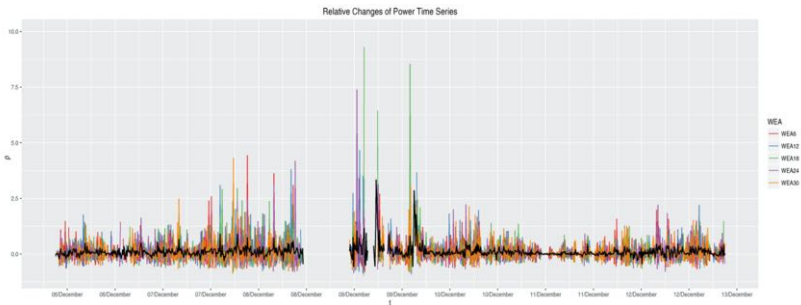
# Time Series

## Relative Changes of Power



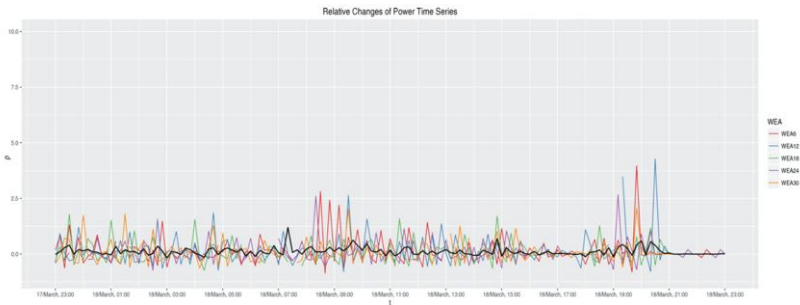
# Time Series

## Relative Changes of Power



# Time Series

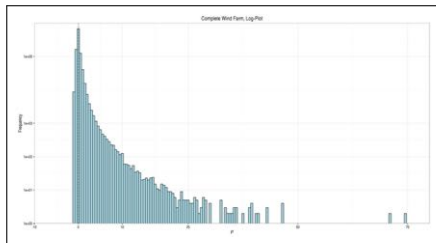
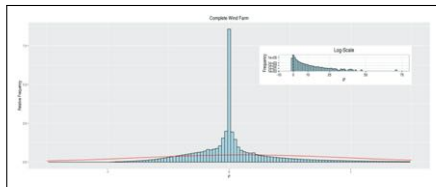
## Relative Changes of Power



# Distributions

## Relative Changes of Power

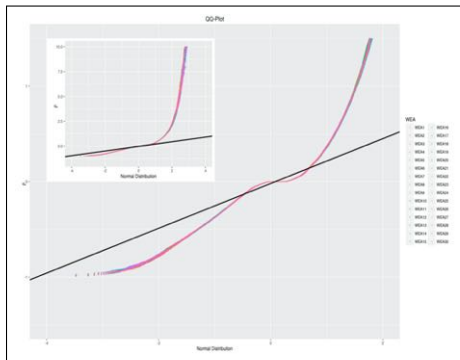
- **Non-Gaussian** but **not heavy tailed!**
- Roughly **symmetric** around zero
- $\approx 13\%$  of  $\rho \in [-0.0125, 0.0125]$
- QQ-Plot: really strong deviations for large values



# Distributions

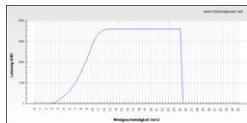
## Relative Changes of Power

- **Non-Gaussian** but **not heavy tailed!**
- Roughly **symmetric** around zero
- $\approx 13\%$  of  $\rho \in [-0.0125, 0.0125]$
- QQ-Plot: really strong deviations for large values

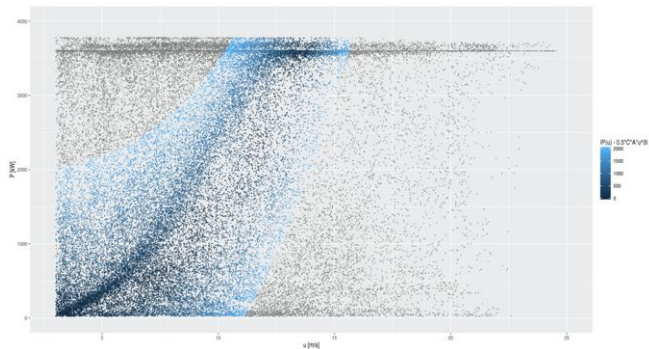


# Power Curve

$$P_m(u) = \frac{C_P A \rho}{2} u^3$$

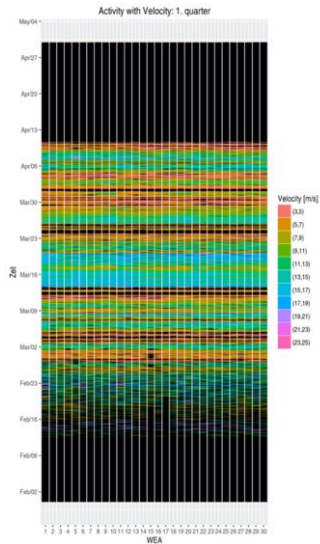
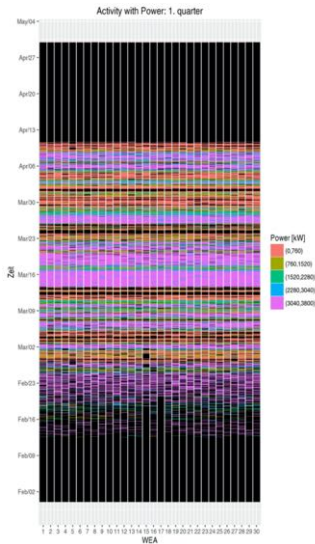


## Averaged over all Wind Turbines



# Activity

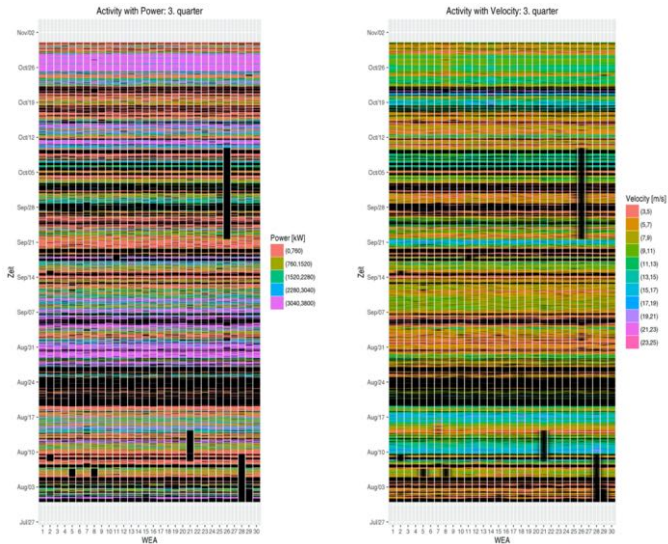
## 1. Quarter





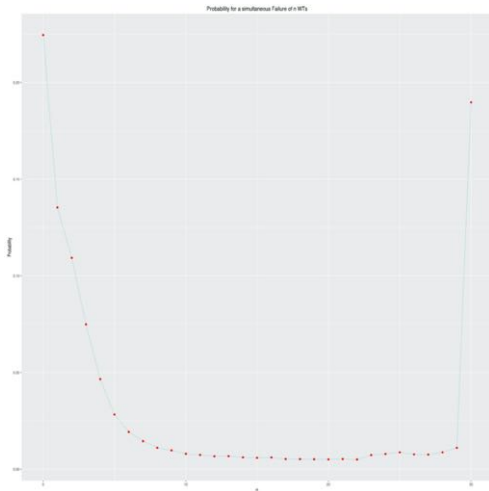
# Activity

## 3. Quarter



## Activity

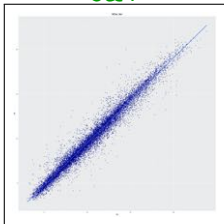
How high is the probability for  $n$  WTs being out of operating range or failing from another reason?



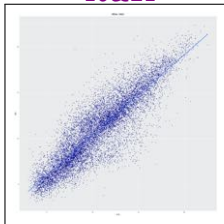
# Scatter Plots

## Velocity and Power

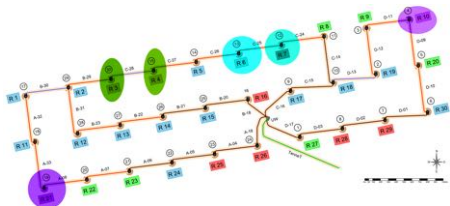
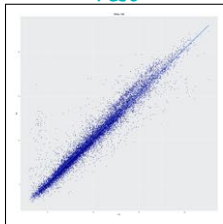
3&4



10&21



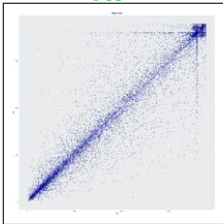
7&6



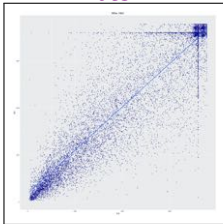
# Scatter Plots

Velocity and Power

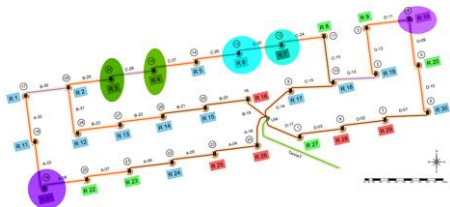
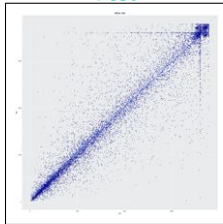
3&4



10&21

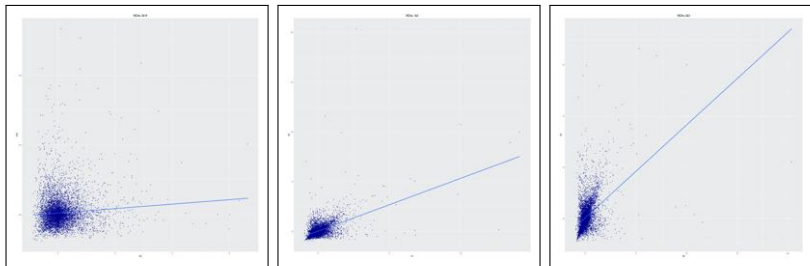


7&6



# Scatter Plots

## Relative Changes of Power



# Ideas and Prospect

## Next Steps

1. Find a way to treat **missing data**  
(without introducing any statistical bias!)

## Next Steps

1. Find a way to treat **missing data**  
(without introducing any statistical bias!)
2. **Finish** first insight into data (failures, “volatility“, relat. changes of velocities, autocorrelations, ...)



## Next Steps

1. Find a way to treat **missing data**  
(without introducing any statistical bias!)
2. **Finish** first insight into data (failures, “volatility“, relat. changes of velocities, autocorrelations, ...)
3. Calculate **cross-correlation-matrices** and get first insights again

## Next Steps

1. Find a way to treat **missing data**  
(without introducing any statistical bias!)
2. **Finish** first insight into data (failures, “volatility“, relat. changes of velocities, autocorrelations, ...)
3. Calculate **cross-correlation-matrices** and get first insights again
4. Start with **cluster-computation** etc.

# Idea of my Master Thesis

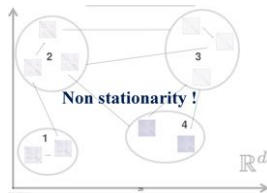
## Summary

### • First Approach:

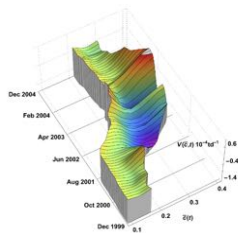
- Calculate  $\underline{\underline{C}}$  for *locally normalized* values of  $\rho(t) = \frac{P(t+\Delta t) - P(t)}{P(t)}$
- Build clusters of **instationary**  $\underline{\underline{C}}(t)$  by *bisecting k-means* ("Wind Farm States")
- Extend analysis by stochastic description of dynamics via *potential functions*

$$V(\bar{c}, t) = - \int^{\bar{c}} f(x, t) dx$$

with empirically estimated drift function  $f(x, t)$



Y. Stepanov: Meta-Stable Stock Market Collective Dynamics with Applications, presentation 2016



Y. Stepanov et al: Stability and Hierarchy of Quasi-Stationary States: Financial Markets as an Example, 2015

# Idea of my Master Thesis

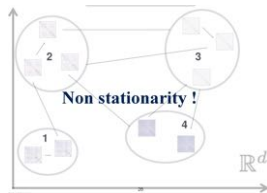
## Summary

- **Advanced Approach:**

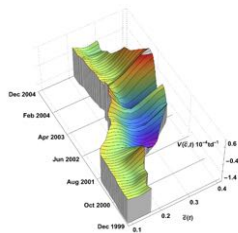
- Use all measured data in

$$30 \times \begin{pmatrix} P \\ u \\ \alpha_{\text{nac}} \\ \phi_{\text{rot}} \\ \Omega_T \\ \vdots \end{pmatrix}_t$$

- Derive only state of **one** WT first with the same methods
- Compare structure and stability of all single-WT-states

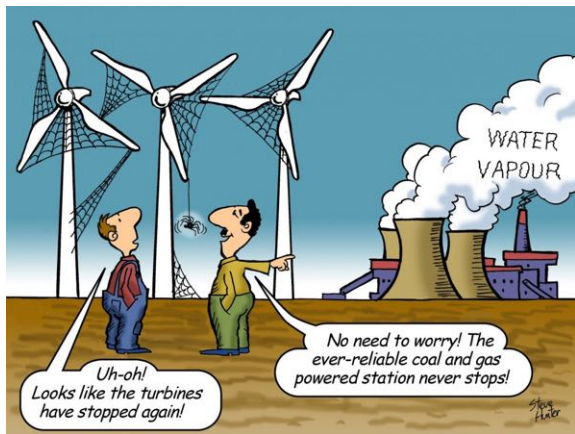


Y. Stepanov: Meta-Stable Stock Market Collective Dynamics with Applications, presentation 2016



Y. Stepanov et al: Stability and Hierarchy of Quasi-Stationary States: Financial Markets as an Example, 2015

# Thank you for your attention!



References: Feel free to ask me!