TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Max Domagk

Faculty of Electrical and Computer Engineering
Institute of Electrical Power Systems and High Voltage Engineering

Analysis and Visualization of Large-Scale Power Quality Monitoring Campaigns

ICHQP 2024 Panel
Turning data into information: Recent developments and future needs in Power Quality data analysis
October 17, Chengdu, China

Agenda

Power Quality data analysis

Example use case: Correlation analysis

- Methodology and dataset
- Correlations between PQ parameters
- Correlations between measurement sites

Conclusions

Power Quality data analysis

Data amounts

- Measurement of PQ parameters (current and voltage) for one measurement site

Measured	every 10 min	per day	per week	per year
up to 2.5 kHz	**324**	**46.656**	**326.592**	**ca. 17 Mill.**
up to 150 kHz	**4767**	**686.016**	**ca. 5 Mill.**	**ca. 250 Mill.**

- Increasing number of large PQ monitoring campaigns in all voltage levels
- Resulting data amounts can often no longer be evaluated manually

Power Quality data analysis

Use cases / analysis goals

- Compliance with limits
- Anomaly detection
- Emission profiles
- Correlation and propagation
- Seasonal variations
- Trend identification
- Trend forecasting
- …

General analysis steps

1. Define use case / analysis goal
2. Data acquisition
3. Data pre-processing
4. Analysis methods
5. Visualization and knowledge extraction

Example use case: Correlation analysis

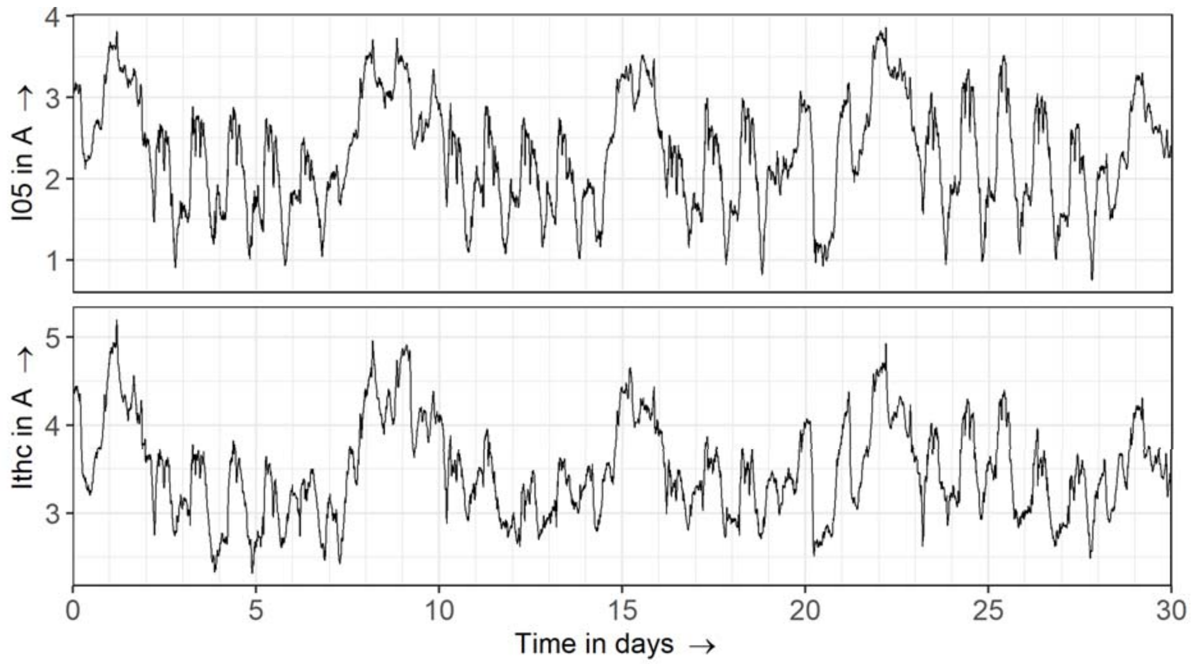
Correlation analysis

- Key for analysing relationships and dependencies in measurement data
- Can help to identify similarities in observed behaviour or asses propagation within the network

Types of correlations

- Auto-Correlation
  between different time intervals (e.g. days) within one time series
- Cross-Correlation
  between PQ parameters of one site or between sites for one PQ parameter

PQ parameters of one measurement site



Example use case: Correlation analysis
Methodology

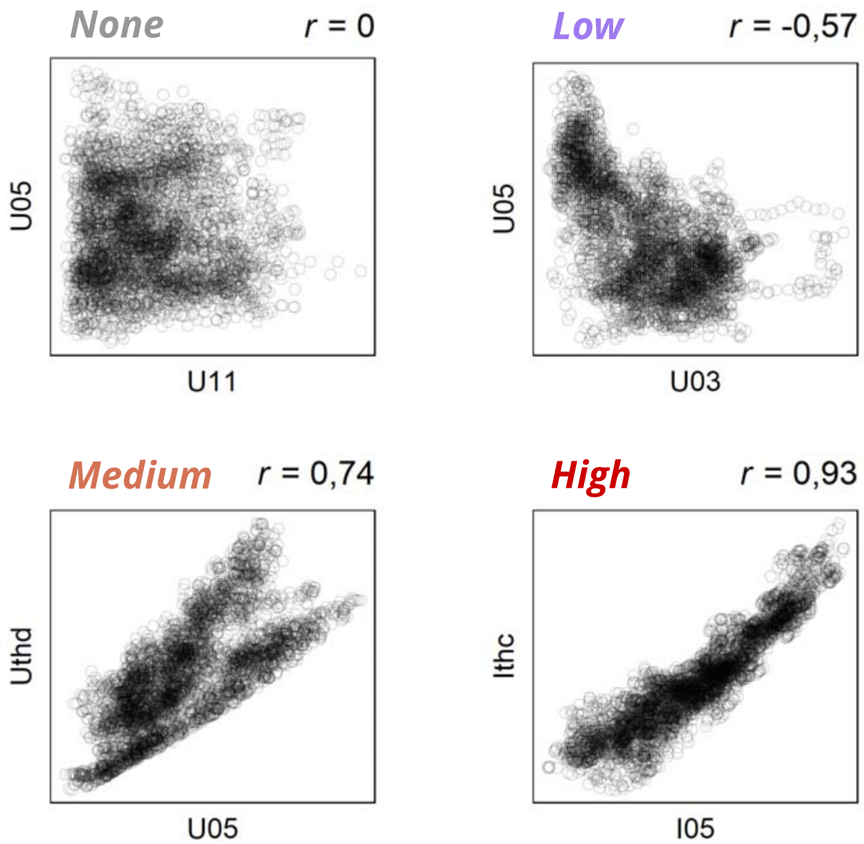
Spearman's correlation coefficient

- Measure of monotonic behaviour based on ranks
- Not limited to linear relationships

$$r = \frac{\sum_i \left(\mathrm{R}(x_i) - \bar{\mathrm{R}}_x\right)\left(\mathrm{R}(y_i) - \bar{\mathrm{R}}_y\right)}{\sqrt{\sum_i \left(\mathrm{R}(x_i) - \bar{\mathrm{R}}_x\right)^2}\sqrt{\sum_i \left(\mathrm{R}(y_i) - \bar{\mathrm{R}}_y\right)^2}}$$

Levels of correlation

- None $0 \leq |r| < 0.5$
- Low $0.5 \leq |r| < 0.7$
- Medium $0.7 \leq |r| < 0.9$
- High $0.9 \leq |r| \leq 1$



Example use case: Correlation analysis
Dataset

Measurement data

- Monitoring of German transmission system
- **85 measurement sites** in HV and EHV
  - 110 kV with 38 sites
  - 220 kV with 21 sites
  - 380 kV with 26 sites
- **20 PQ parameters**
  - RMS of voltage (Urms) and current (Irms)
  - Voltage unbalance (UNB)
  - Short-term flicker (Upst)
  - Distortion of voltage (Uthd) and current (Ithc)
  - Harmonics of voltage (U03-U15) and current (I03-I15)
  - In total 58 parameter-phase combinations
- 10 min average values for a duration of 30 days

Correlations between PQ parameters

- 58 parameters = 1.653 unique correlations
- 1.653 x 85 sites = 140.505 correlations

Correlations between measurement sites

- 85 sites = 3.570 unique correlations
- 3.570 x 58 parameters = 207.060 correlations

Analysis of a total of 347.565 correlations

Correlations between PQ parameters
Results aggregation (1)

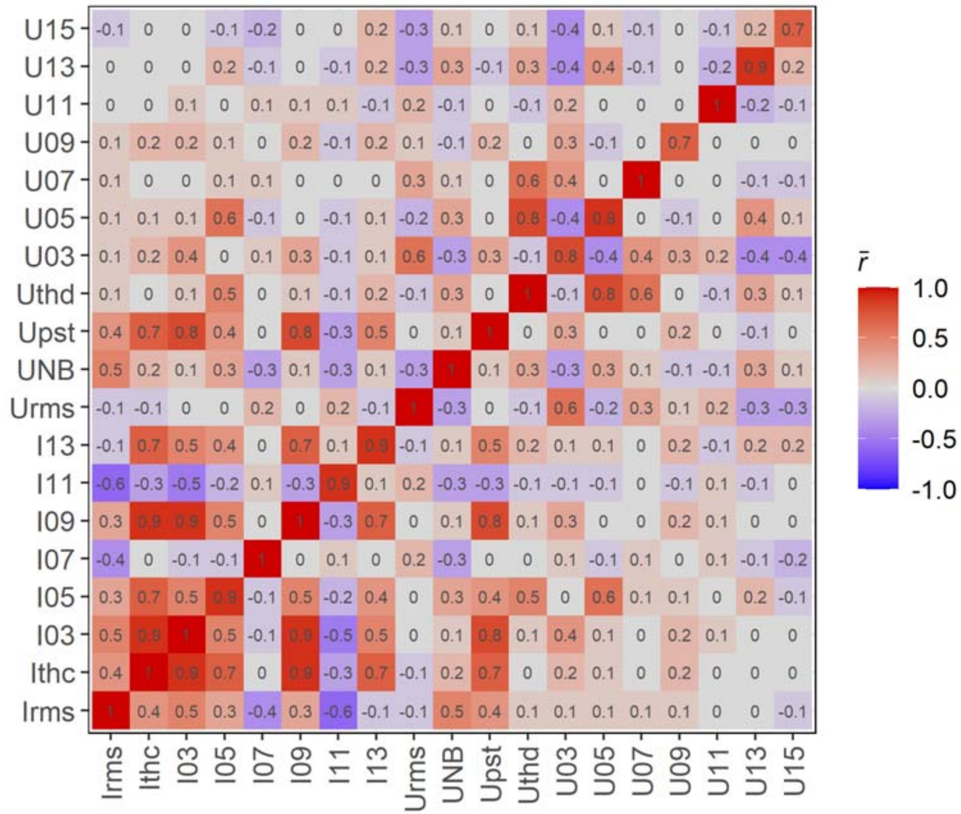
Averaging of correlation coefficients

- Up to 9 phase combinations between two parameters
  (e.g. Ithc L1/L2/L3 ~ I05 L1/L2/L3)
- Averaging reduces matrix by nearly a factor of 9
  (from 58^2 = 3.364 to 20^2 = 400 elements)
- Averaging using Fisher Z-transformation
  (otherwise too low averages due to value range [-1, 1])

- Analysing multiple matrices remains challenging
  (e.g. 85 matrices for all measurement sites)

→ Further aggregation of results necessary

Correlations between PQ parameters for one measurement site in 380 kV



TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Correlations between PQ parameters
Results aggregation (2)

Aggregation of significant correlations

- Filtering of significant correlations (e.g. $|r| \geq 0.7$)
- Calculation of share across all matrices

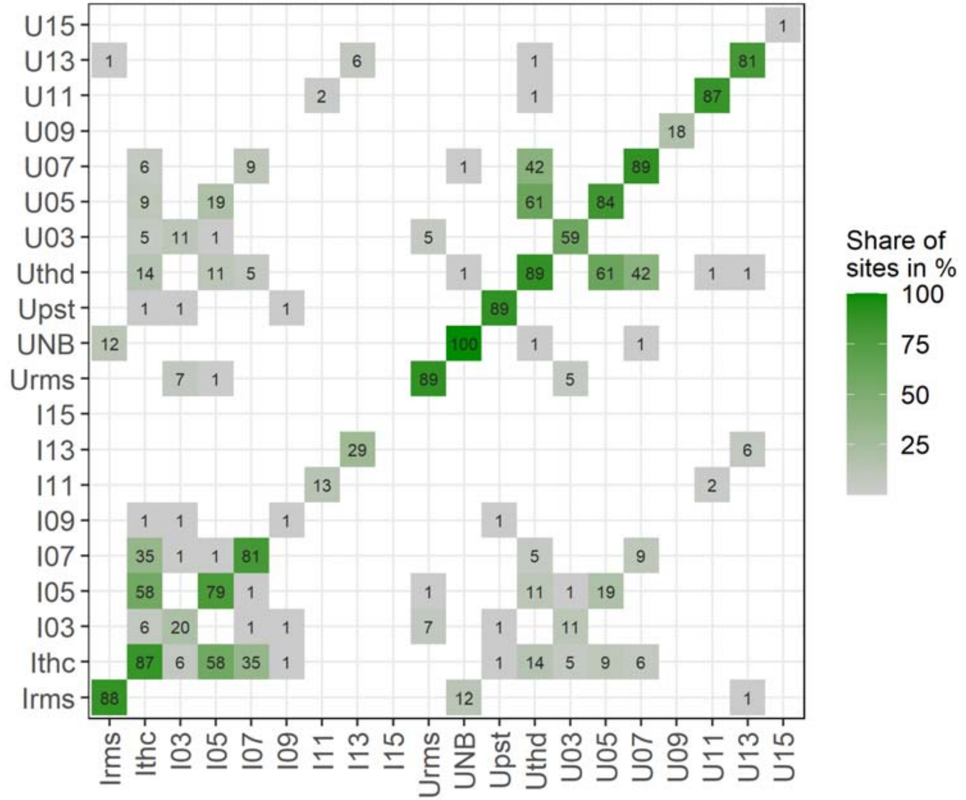
Visualization of matrices

- Very large matrices may be difficult to interpret
- Various visualizations available for distance matrices

→ Calculation of distance matrix from the matrix of significant correlation shares:

$$D_{ij} = \begin{cases} 100 - S_{ij}, & \text{for } i \neq j \\ 0, & \text{for } i = j \end{cases}$$

Share of significant correlations (|r|>0.7) between PQ parameters across all sites

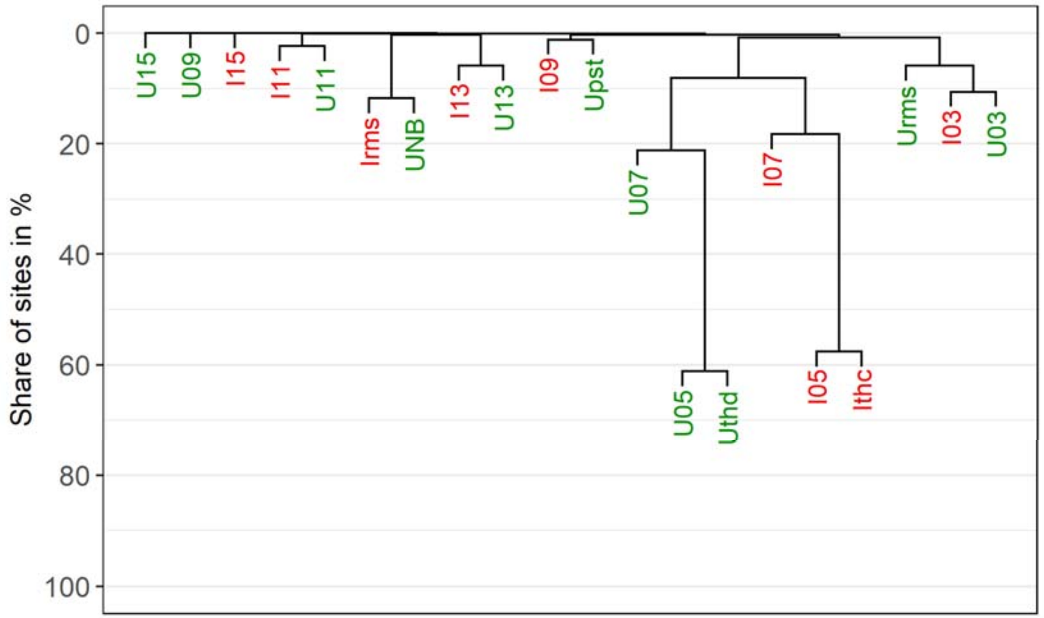


Correlations between PQ parameters
Visualization (1)

Dendrogram of clustering

- Clusters merged step-by-step based on similarity/distance
- Tree-like diagram showing hierarchical relationships between elements
- Height of branch indicates the level of similarity

Dendrogram of clustering based on the share of significant correlations



Correlations between PQ parameters
Visualization (2)

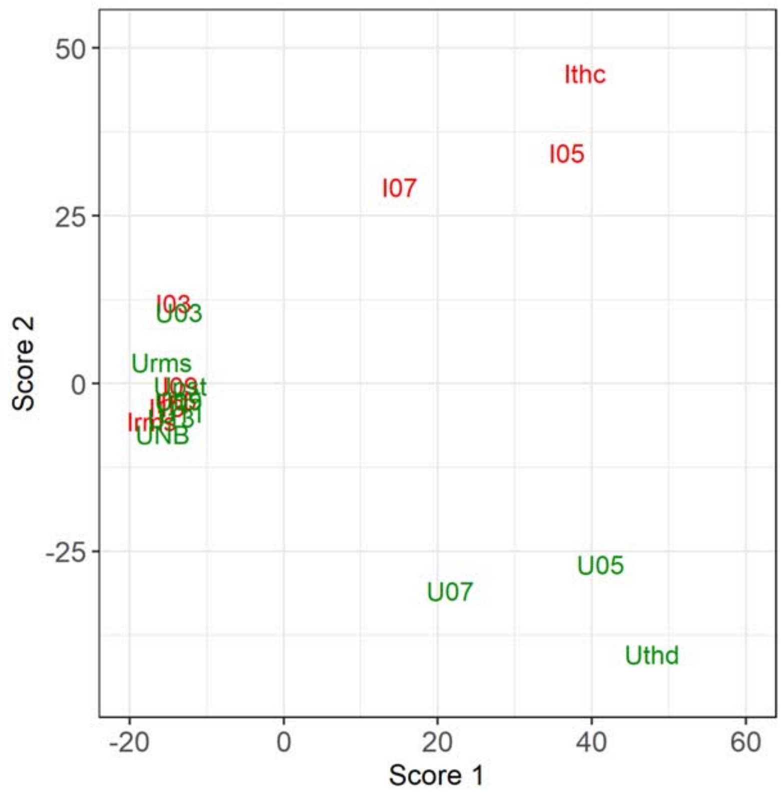
Dendrogram of clustering

- Clusters merged step-by-step based on similarity/distance
- Tree-like diagram showing hierarchical relationships between elements
- Height of branch indicates the level of similarity

Multidimensional scaling (MDS)

- Technique for dimension reduction
- Visualizes distances between elements in lower-dimensional space
- Preserves relative distances to reflect similarities

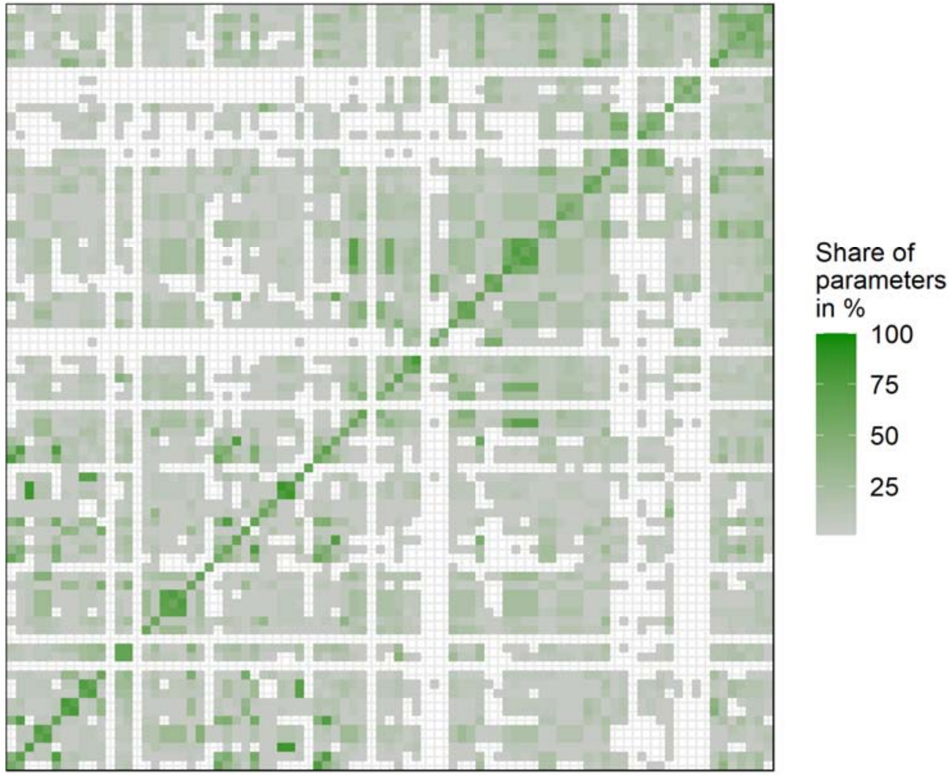
2D scatterplot using MDS based on the share of significant correlations



Correlations between measurement sites
Results aggregation

- Difficult to interpret holistically due to large number of individual elements
- Still, some observations:
  - Some sites with no correlations
  - Many sites with few correlations (ca.10 %)
  - Few sites with many correlations (up to 85 %)

Share of significant correlations (|r|>0.7) between sites across all parameters

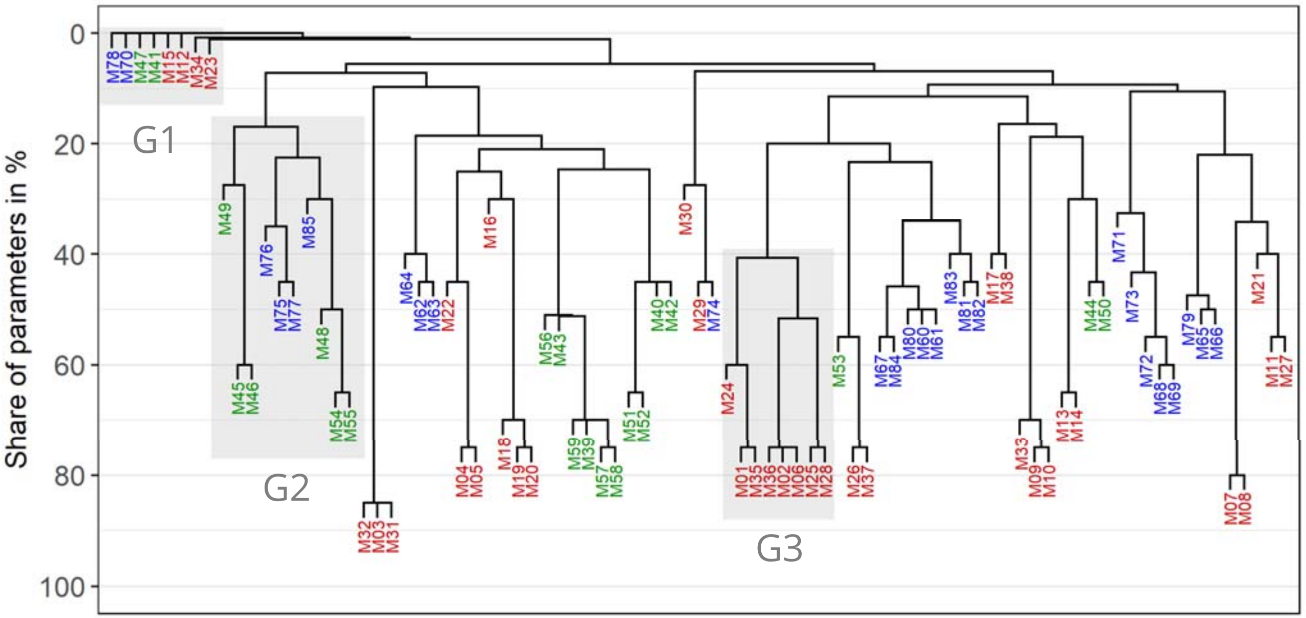


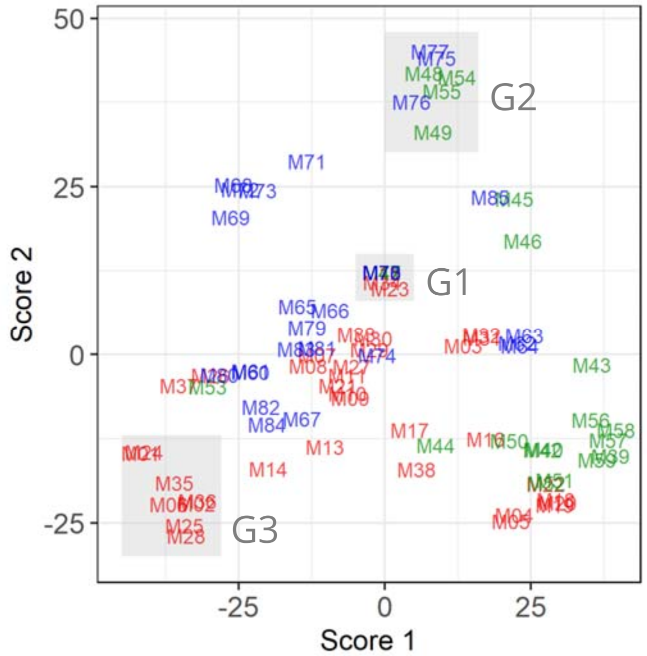
Correlations between measurement sites

Visualization (1)

- Combination of dendrogram and 2D scatterplot allows to identify different groups

- G1 = Sites that do not or barely correlate
- G2 = EHV sites partially correlated (ca. 20% - 60%)
- G3 = HV sites highly correlated (ca. 40% – 80%)



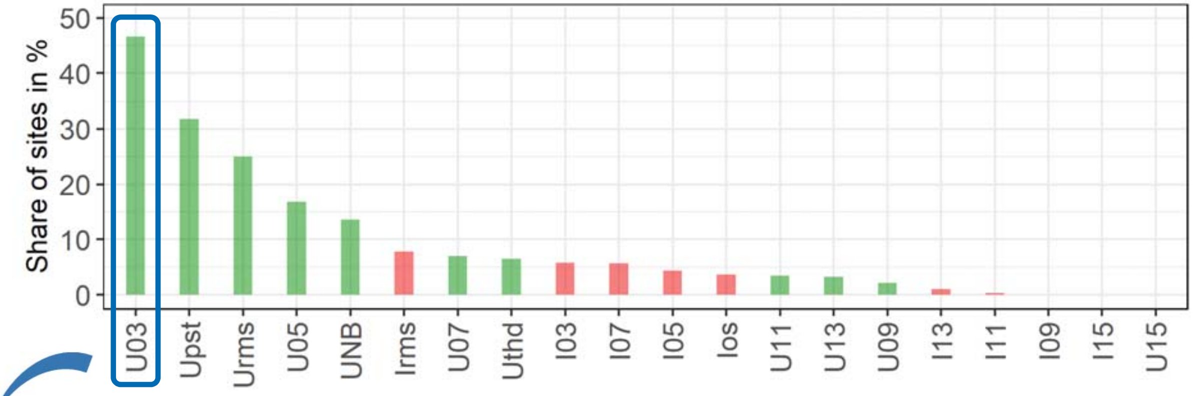


Voltage level: 110 kV, 220 kV, 380 kV

Correlations between measurement sites
Visualization (2)

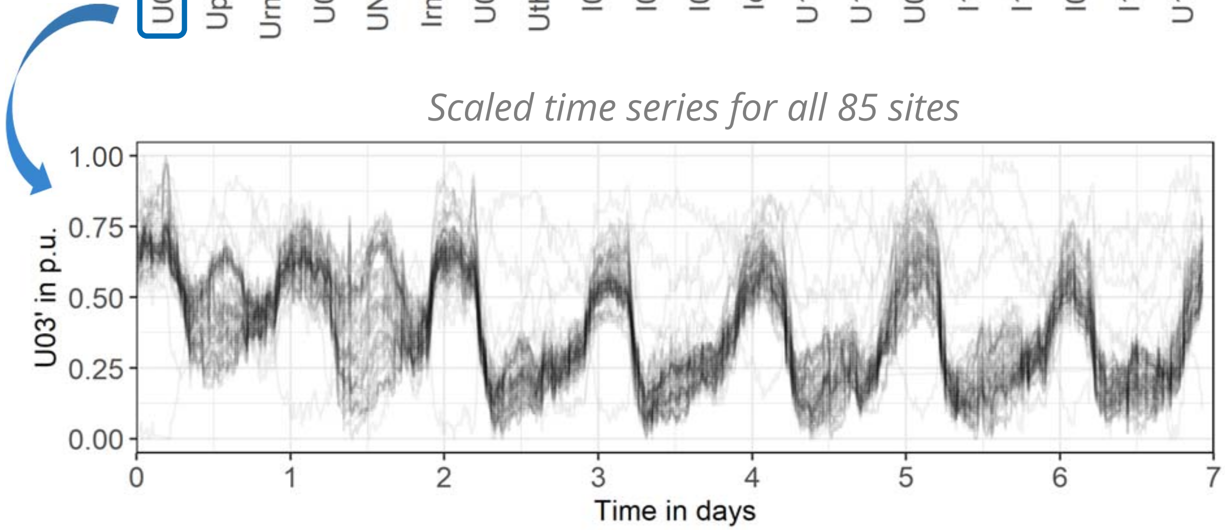
- Most correlations between sites for 1 or 2 parameters
- Highest share of correlations between all 85 sites for:
  - 3rd voltage harmonic (U03) and
  - Short-term flicker (Upst)

Share of correlations with |r|>0.7 between sites



- 3rd voltage harmonic with distinctive daily pattern across all voltage levels
- Possible harmonic sources:
  - Converters
  - Transformer magnetization
  - Corona discharges

Scaled time series for all 85 sites



Conclusions

Power Quality data analysis

- Robust and reliable methods are essential for an automated application
- Results aggregation and visualizations are key for:
  - Extracting actionable insights and
  - Making data accessible to non-experts

Future research needs

- Advanced multivariate analysis methods for large-scale PQ monitoring campaigns (e.g. clustering approaches and dimension reduction)
- Automated data pre-processing:
  - Data validation (e.g. measurement errors, measurement uncertainty)
  - Data imputation (e.g. handling of missing values)

Thank you for your attention!

max.domagk@tu-dresden.de
+49 351 463 35223
maxdomagk.de