Outlier Detection

Detection of critical events in renewable energy production time series

Laurens P. Stoop, Erik Duijm, Ad Feelders, Machteld van den Broek

AALTD'21 – 13 September 2021

Utrecht University

Koninklijk Nederlands Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

TenneT

university of groningen

European Energy System

- One directional system
- Central control & modulation
- Relies on large generators

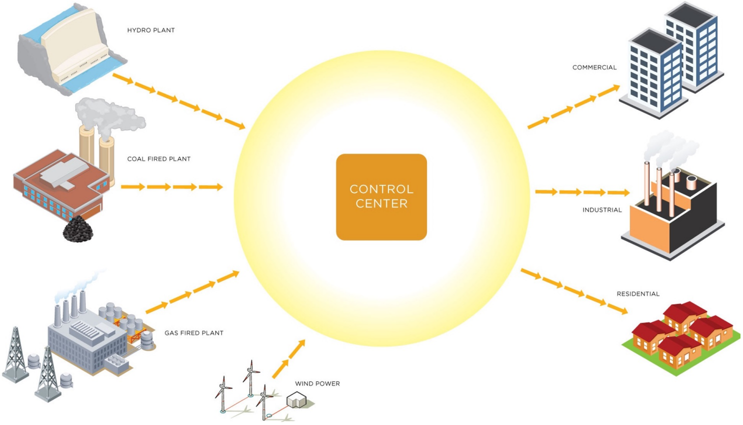


Figure – Historical large scale electrical infrastructure. Image from International Energy Agency.

European Energy System

- Energy Transition to renewable sources
- Increased variability in generation and demand
- Complex bi-directional network

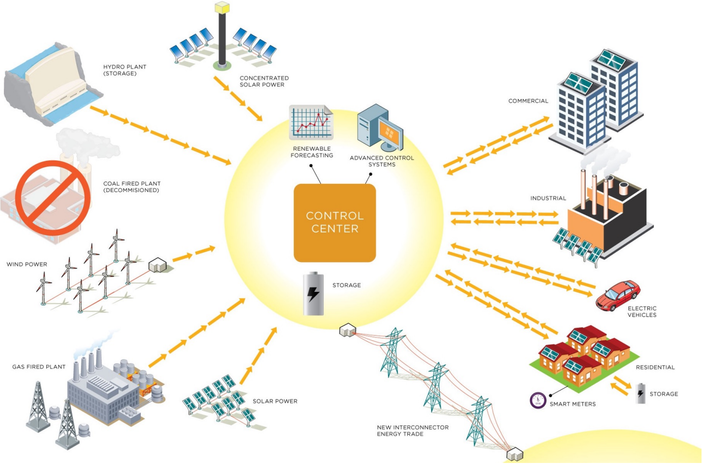


Figure – Likely large scale electrical infrastructure. Image from International Energy Agency.

Preparing for the future

Solve the Unit Commitment problem

How to schedule the generators to produce the demand at minimal cost?

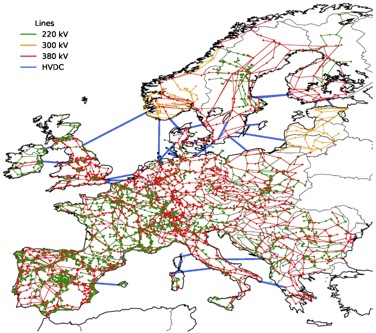


Figure – European transmission network model, includes lines that are planned and under construction. Image from Hörsch et al. [1].

Preparing for the future

Solve the Unit Commitment problem

How to schedule the generators to produce the demand at minimal cost?

- Upgrading and improving the grid
- Many scenarios have to be considered
- Large variability of weather
- Risks; Load mismatch, blackout

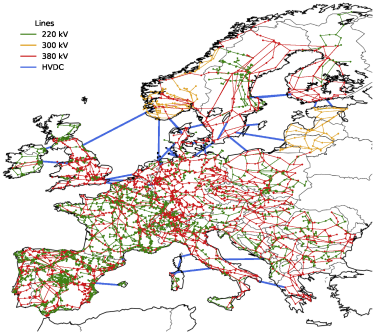


Figure – European transmission network model, includes lines that are planned and under construction. Image from Hörsch et al. [1].

Preparing for the future

Solve the Unit Commitment problem

How to schedule the generators to produce the demand at minimal cost?

- Upgrading and improving the grid
- Many scenarios have to be considered
- Large variability of weather
- Risks; Load mismatch, blackout

NP-Hard optimization problem

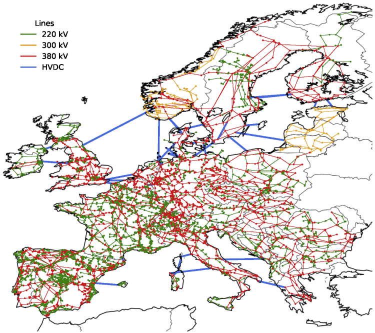


Figure – European transmission network model, includes lines that are planned and under construction. Image from Hörsch et al. [1].

Preparing for the future

Outlier Detection

Outliers represent the most extreme events. Reduces the input, while allowing for detailed modelling of the system.

ERA5 Reanalysis Data

- ERA5 Reanalysis Data

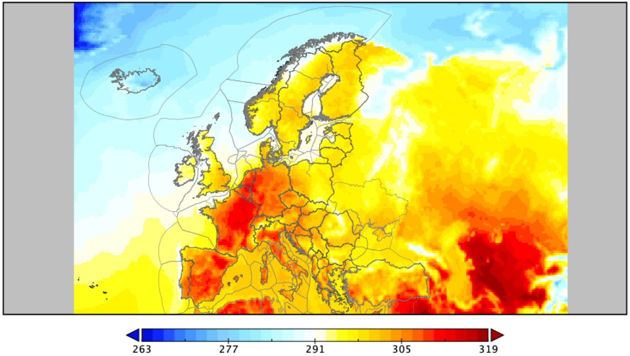


Figure – Example of ERA5 data. T2m during heatwave of 25th july 2019 13:00.

ERA5 Reanalysis Data

- ERA5 Reanalysis Data
- 1950-2019
- Hourly Resolution
- 0.25 degree resolution

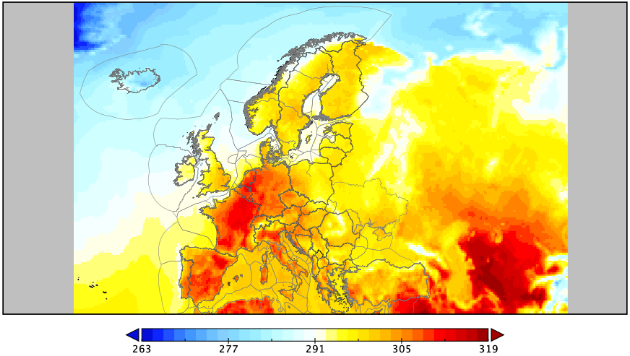


Figure – Example of ERA5 data. T2m during heatwave of 25th july 2019 13:00.

ERA5 Reanalysis Data

- ERA5 Reanalysis Data
- 1950-2019
- Hourly Resolution
- 0.25 degree resolution
- Autocorrelated and Heteroscedastic

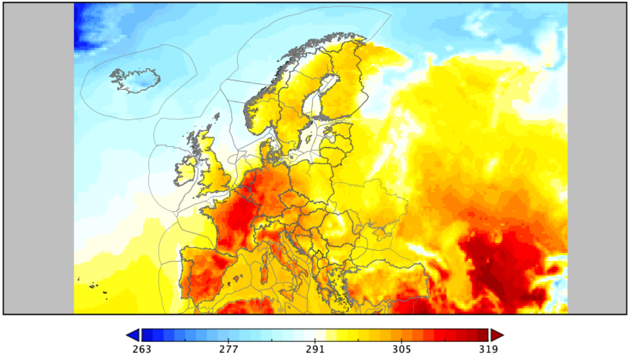


Figure – Example of ERA5 data. T2m during heatwave of 25th july 2019 13:00.

ERA5 Reanalysis Data

- ERA5 Reanalysis Data
- 1950-2019
- Hourly Resolution
- 0.25 degree resolution
- Autocorrelated and Heteroscedastic

Energy Conversion Models

Wind turbines onshore (WON),
Wind turbines offshore (WOF),
Solar Photovoltaic panels (SPV)

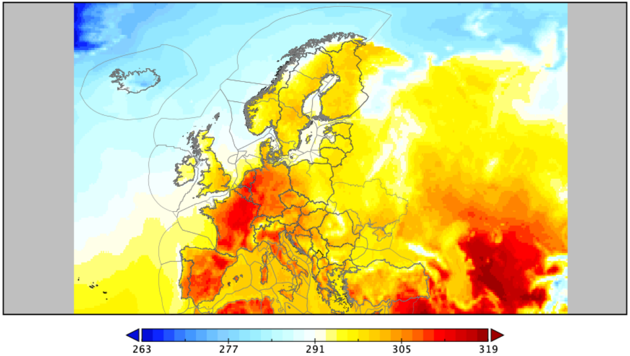


Figure – Example of ERA5 data. T2m during heatwave of 25th july 2019 13:00.

Maximally Divergent Intervals (MDI) Algorithm

- Introduced by Rodner et al. (2015) [2] to detect outliers in temporal data
- Improved and Expanded to work for spatial-temporal data by Barz et al. (2017-2018) [3][4]

Maximally Divergent Intervals (MDI) Algorithm

- Introduced by Rodner et al. (2015) [2] to detect outliers in temporal data
- Improved and Expanded to work for spatial-temporal data by Barz et al. (2017-2018) [3][4]
- Compares Distributions of Interval I with remaining data Ω
- Divergence is the outlier score
- Approach works for multivariate spatial-temporal data

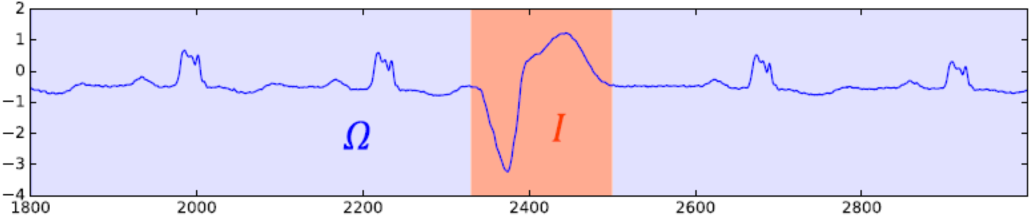


Figure – Principle idea of the MDI algorithm. The distribution of the interval *I* is compared to the remaining data Ω. Image from Barz et al. (2018)[4]

MDI: Context Embedding

- Data is autocorrelated
- Transform data point to phase space that is uncorrelated: Context Embedding

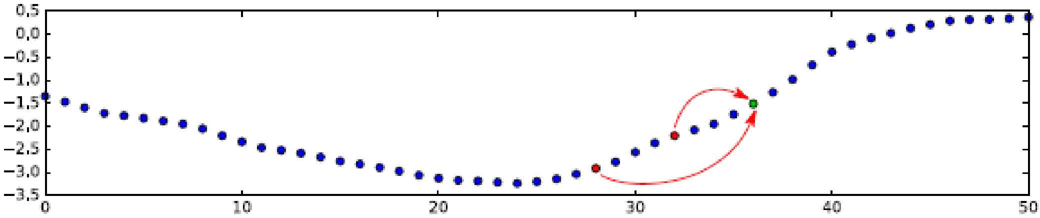


Figure – Examples of spatial and temporal context embedding. Images from Barz et al. (2017) [3].

Experimental setup

- Used to detect temporal outliers with context embedding
- Outlier scores: Cross Entropy and Unbiased Kullback-Leibler
- Multi-variate data (WON/WOF/SPV)

- Experiment 1: Outliers on the entire region $\rightarrow$ *tuning*
- Experiment 2: Climate Change Experiments
- Experiment 3: Temporal Outliers and their Spatial Location $\rightarrow$ *not presented*

Experiment 1: The European Region Cross Entropy

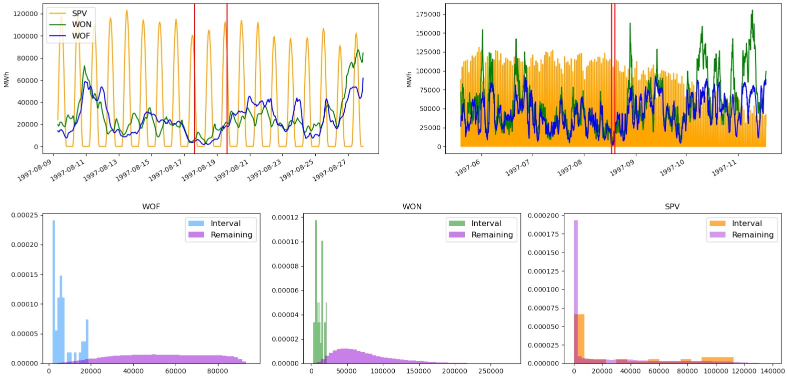


Figure – Top Cross Entropy Result Western Europe

Experiment 1: The European Region Cross Entropy

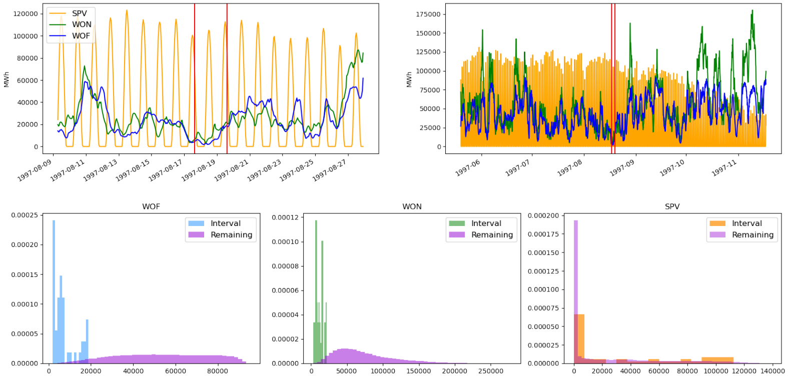


Figure – Top Cross Entropy Result Western Europe

- An adverse weather system for the electricity system of the UK and Europe[5].

Experiment 1: The European Region Unbiased Kullback-Leibler

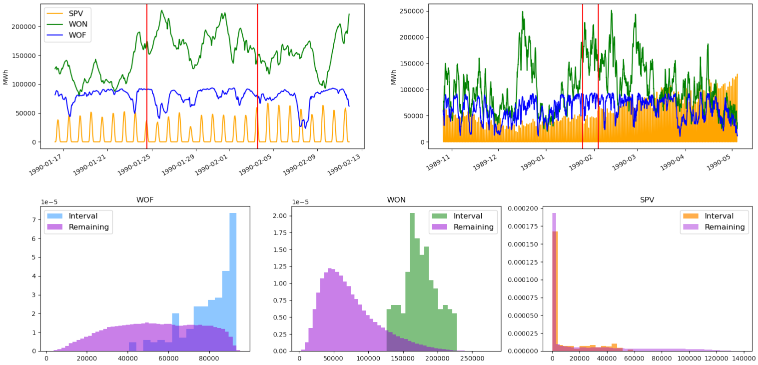


Figure – Top Unbiased Kullback-Leibler Result Western Europe

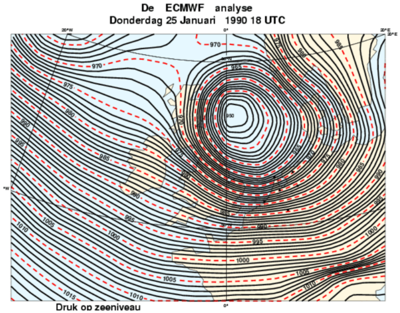


Figure – Hurricane Daria, air pressure map 25-01-1990, Source KNMI

Experiment 1: The European Region Events

Cross Entropy						
Top-k	Month	Length(h)	SPV	WON	WOF	Type
1/6/13/19	*Aug.*	48-72	+	−	−	T
3/5	*June*	48-72	+	−	−	T
07/09/2017	*July*	48-72	+	−	−	T
16	*July*	72-96	+	−	−	T
10/15	*July*	150-175	+	−	−	T
14	*Feb.*	48-72	−	+	−	PT
4	*Apr.*	48-72	0	+	+	P
2/11	*Dec.*	48-72	−	+	+	P
12/18	*Feb.*	48-72	−	+	+	P
20	*Jan.*	48-72	−	+	+	P

Figure – Top 20 Cross Entropy Outliers

- Summer Deficiency
- Winter Surplus
- Interval Length Preferance

Experiment 2: Climate Change Experiment

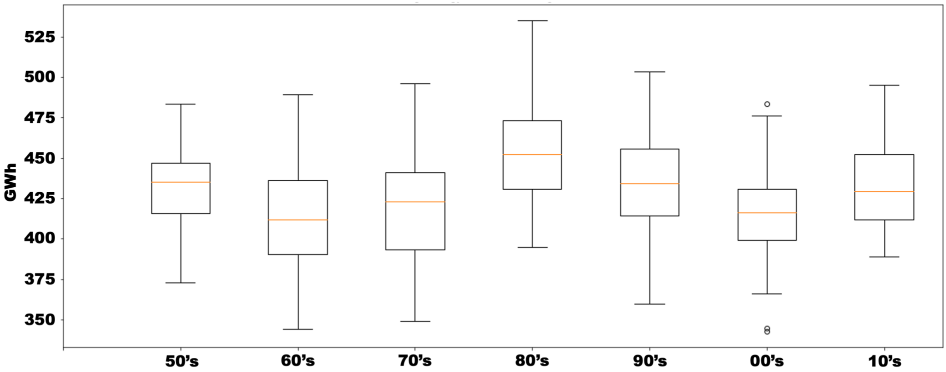


Figure – Average Intensity of top 50 Cross Entropy Outliers. Total Energy generation over Western Europe used

Experiment 2: Climate Change Experiment

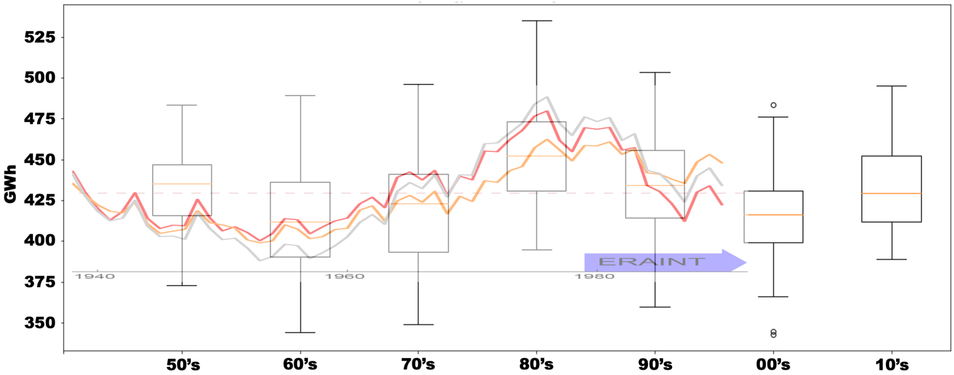


Figure – Average Intensity of top 50 Cross Entropy Outliers. Total Energy generation over Western Europe used

- Similar Multidecadal Variability detected by Wohland et al. [6].

Summary

- Application of tuned MDI on Energy-Climate data
- Highlighting of Extreme Events for the energy sector
- Events detected used as input for Energy System Models (Unit Commitment)
  - Assessment of adequacy and possible changes of risk during extreme events
  - Allows assessment of wider range of scenarios (climate & energy)

Summary

- Application of tuned MDI on Energy-Climate data
- Highlighting of Extreme Events for the energy sector
- Events detected used as input for Energy System Models (Unit Commitment)
  - Assessment of adequacy and possible changes of risk during extreme events
  - Allows assessment of wider range of scenarios (climate & energy)

More information

On the ACDC-ESM project and other works, see: `uu.nl/staff/LPStoop`

References

- Jonas Hörsch et al. "PyPSA-Eur: An open optimisation model of the European transmission system". In: *Energy Strategy Reviews* 22 (2018), pp. 207–215. ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2018.08.012. URL: http://www.sciencedirect.com/science/article/pii/S2211467X18300804.
- Erik Rodner et al. "Maximally divergent intervals for anomaly detection". In: *arXiv preprint arXiv:1610.06761* (2016).
- Björn Barz et al. "Maximally divergent intervals for extreme weather event detection". In: *OCEANS 2017-Aberdeen.* IEEE. 2017, pp. 1–9.
- Björn Barz et al. "Detecting regions of maximal divergence for spatio-temporal anomaly detection". In: *IEEE transactions on pattern analysis and machine intelligence* 41.5 (2018), pp. 1088–1101.
- Laura Dawkins et al. "Characterising Adverse Weather for the UK Electricity System". In: (2020).
- Jan Wohland et al. "Significant multidecadal variability in German wind energy generation". In: *Wind Energy Science* 4.3 (2019), pp. 515–526.

Research part of Algorithmic Computing and Data-mining for Climate integrated Energy System Models (ACDC-ESM) project This research received funding from the Netherlands Organisation for Scientific Research (NWO) under grant number 647.003.005.

The data used in the experiment contains modified Copernicus Climate Change Service information 2020. https://doi.org/10.24381/cds.adbb2d47.

This presentation is part of the IS-ENES3 project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084.

Experiment 3: Spatial Location Experiment

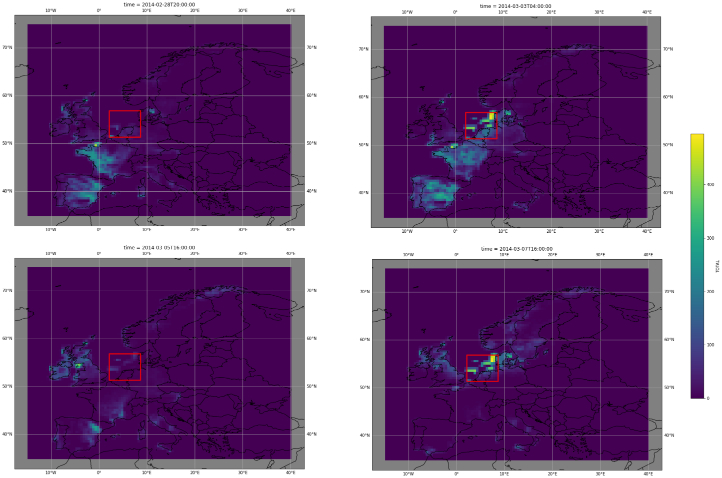
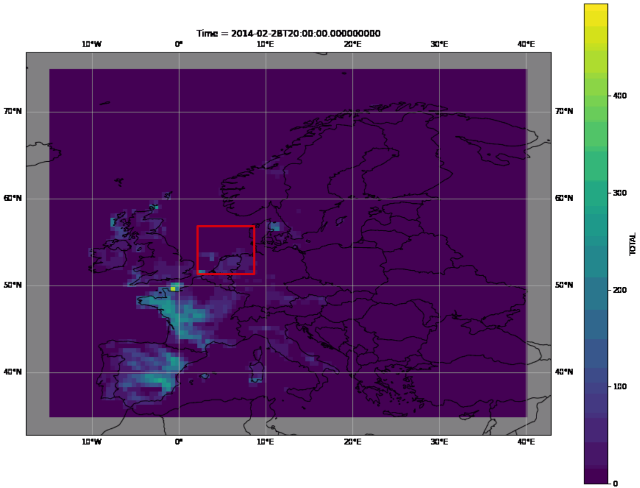
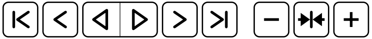


Figure – Four time steps of top Unbiased Kullback-Leibler Outlier 2010-2019 Total Energy Generation. Spatial Location of the Outlier Highlighted

Experiment 3: Spatial Location





Experiment 1: Peak-and-Trough

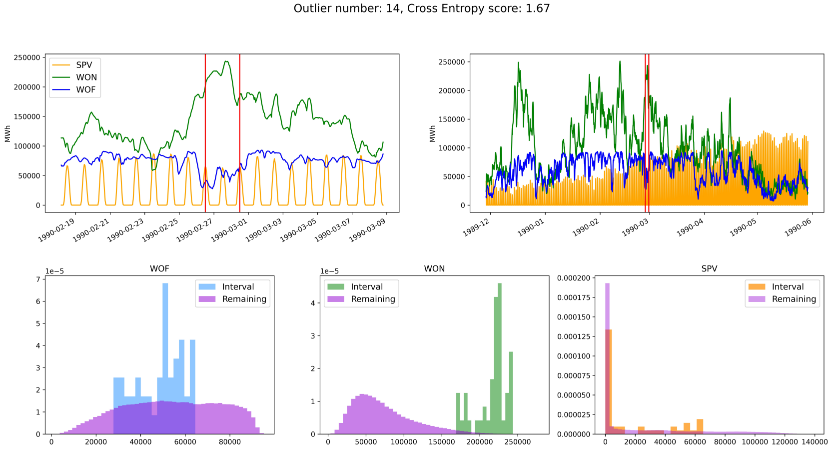


Figure – Peak and Trough. Outlier rank 14 Western Europe Cross Entropy

Experiment 1: The European Region Events

unbiased Kullback-Leibler						
Top-k	Month	Length(h)	SPV	WON	WOF	Type
1/5/7/10	*Jan.*	216+	−	+	+	P
2/8/17	*Dec.*	216+	−	+	+	P
3/4	*Feb.*	216+	−	+	+	P
11/18-20	*Nov.*	216+	−	+	+	P
9	*Jan.*	192 − 216	−	+	+	P
6/13/16	*Feb.*	216+	0	+	+	P
14	*Aug.*	216+	+	−	−	T
15	*July*	216+	+	−	−	T

Figure – Top 20 unbiased Kullback-Leibler

Other experiments: Preference for interval length CE

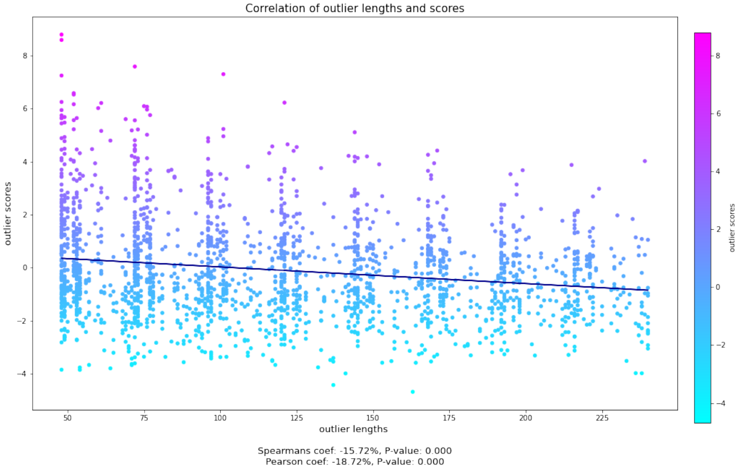


Figure – Correlation between outlier lengths and their scores under Cross Entropy

Other experiments: Preference for interval length uKL

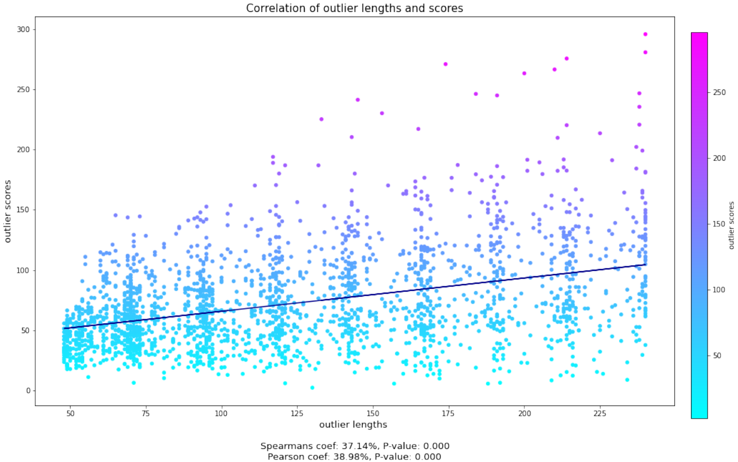
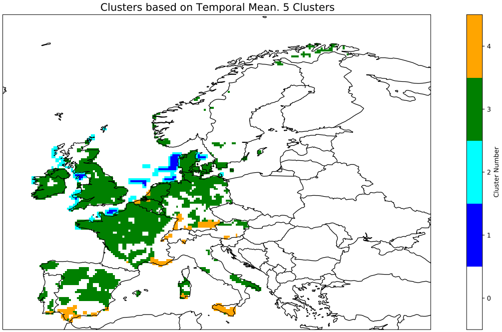
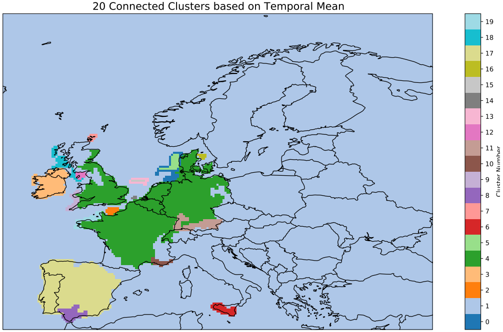


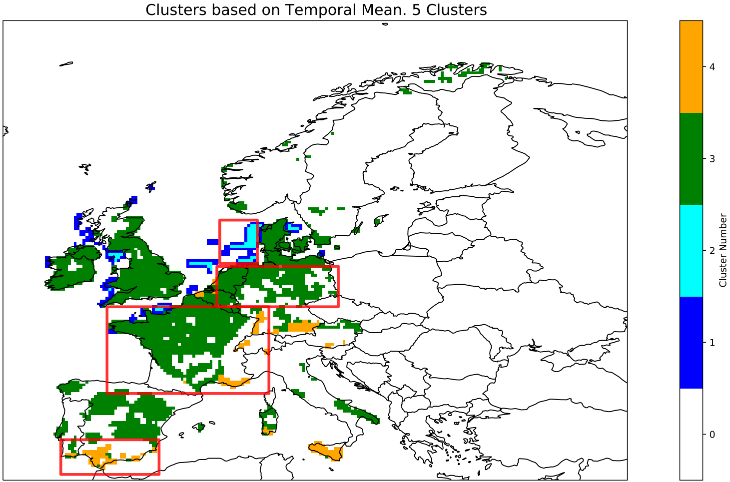
Figure – Correlation between outlier lengths and their scores under unbiased Kullback-Leibler Entropy

Other experiments: Clustering





Other experiments: Regions



MDI: Complexity & spatial component

- 613,594 Hours
- 21.5 billion grid cells
- Naive approach
  $\mathcal{O}(N \cdot L(N + L))$ for the Gaussian model and $\mathcal{O}(N^2 \cdot L^2)$ for KDE using Gaussian kernels.
- Cumulative sums
  $\mathcal{O}(N \cdot L^2)$ for the Gaussian model and $\mathcal{O}(N^2 + N \cdot L^2)$ for the KDE model.
- Closed form solutions for Gaussian model
- $\mathcal{O}(N \cdot L)$
- Hotteling's T^2 Squared heuristic.

MDI: Unbiased Kullback-Leibler

- Kullback-Leibler Divergence bias towards smaller intervals
- Unbiased KL Divergence:

$$\mathcal{D}_{U-KL}(I, \Omega) := 2 \cdot |I| \cdot \mathcal{D}_{KL}(I, \Omega)$$

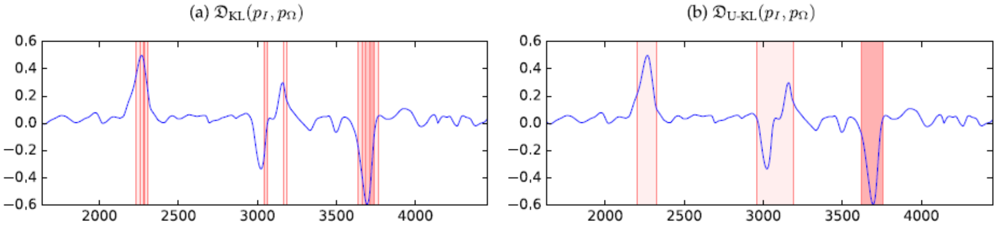


Figure – Top detections on real time data using both regular and unbiased Kullback-Leibler divergence. Image from Barz et al. (2018) [4]

MDI: Divergence Measures

Cross Entropy

- $H(p, q) = \mathbf{E}_p[-log\ q]$
- How surprising is a drawn sample
- Can be estimated empirically:

$$\mathcal{D}_{CE}(I, \Omega) = \frac{1}{|I|} \sum_{i \in I} \log p_{\Omega}(\mathbf{X}_i)$$

Kullback-Leibler Divergence

- $H(p, q) - H(p, p)$
- Can be estimated empirically:

$$\mathcal{D}_{KL}(I, \Omega) = \frac{1}{|I|} \sum_{i \in I} \log p_I(\mathbf{X}_i) - \log p_{\Omega}(\mathbf{X}_i)$$

Unbiased Kullback-Leibler

$$\mathcal{D}_{U-KL}(p_I, p_{\Omega}) := 2 \cdot |I| \cdot \mathcal{D}_{KL}(p_I, p_{\Omega})$$

MDI: Hottelings T^2 Interval Proposal Heuristic

- Multivariate Generalization of Students T test
- Point wise outlier score
- Outlying intervals likely to have high point wise score
- Start and end with high T^2 gradient
- Threshold value of the gradient proposes potentially outlying intervals

MDI: Context Embedding Results

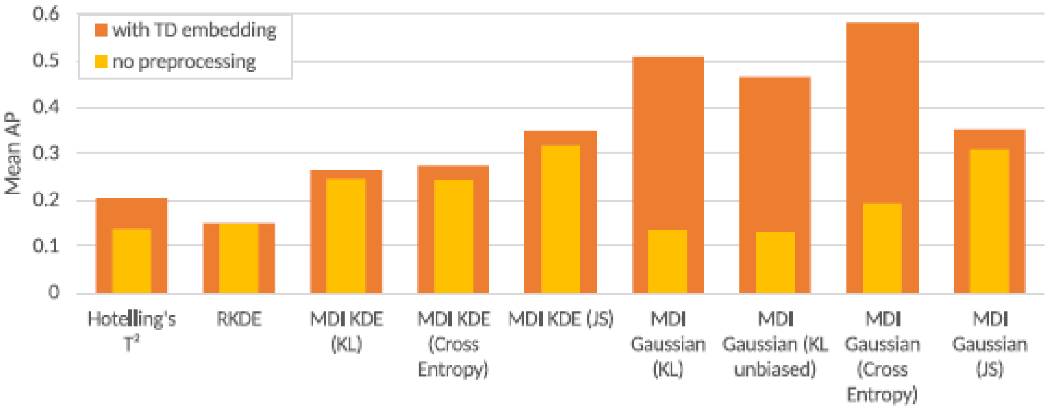


Figure – Comparison of results of MDI on synthetic data with and without time delay embedding, k=6, t = 2. Image from Barz et al. (2018) [4]

MDI: Partial Autocorrelation

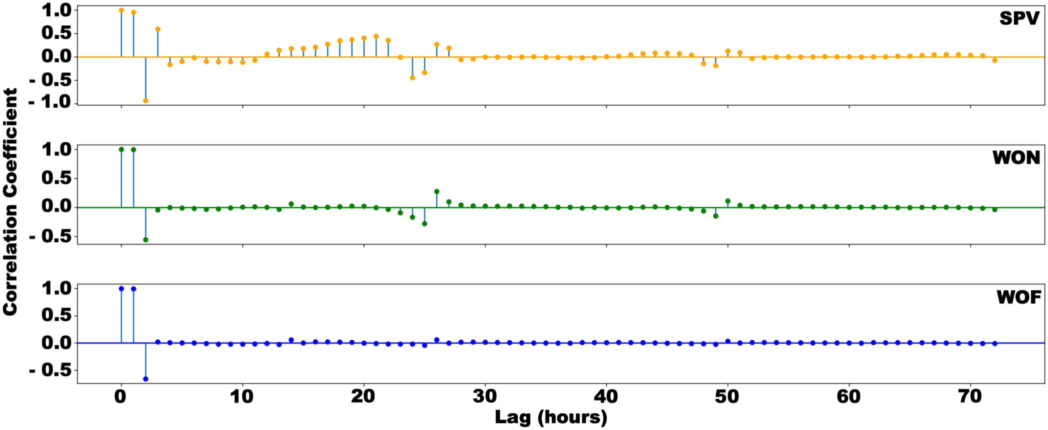
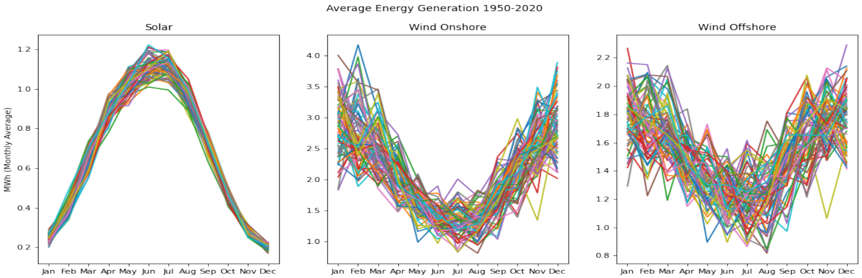


Figure – Partial Autocorrelation plots

MDI: Seasonality

- Seasonal Behaviour



- Hourly Z score

$$Z = \frac{\mathbf{X} - \mu}{\sigma}$$