Power System Reliability with Deep Learning

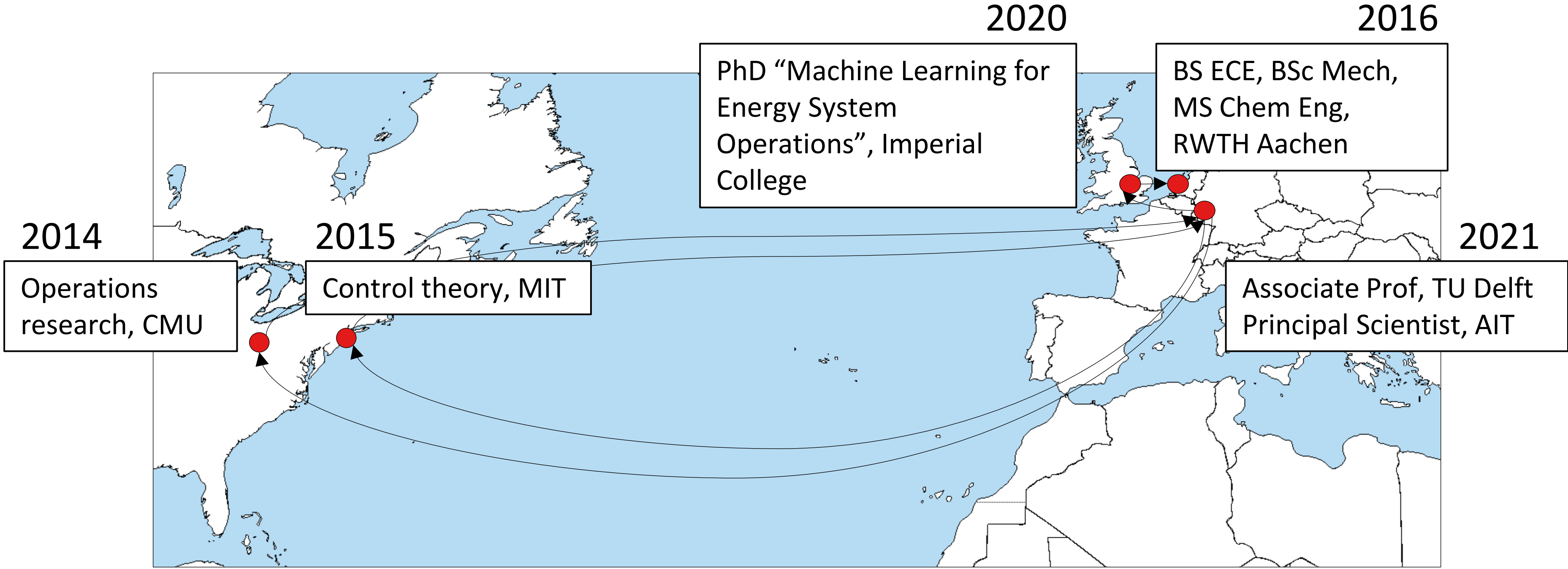
DTU PES Summer School, 20-05-2025

Prof. Dr. Jochen L. Cremer,

Associate Professor

TU Delft

Introduction



TU Delft

Delft AI Energy Lab

Mission & objective

- combine groundbreaking ML with the reliable theory of the physical energy system
- make energy systems sustainable, reliable, effective

Education

- EE4C12 ML for Electrical Engineering
- SET 3125 Machine Learning Workflows for Digital Energy Systems
- SC42150 Statistical Signal Processing
- SC42110 Dynamic Programming and Stochastic Control
- MOOC Digitalization of Intelligent and Integrated Energy Systems
- Crash course of “Data-science”

Research

- Supervised learning for real-time grid assessment
- Distributed learning for power system congestion management
- Data-driven grid models for electricity load and weather forecasts
- Characterizing healthy/normal trajectories of complex dynamical systems using dictionary learning
- From fast Fourier transform to fast reinforcement learning

Key innovations

- AI-based algorithms for grid operation
- Real-time security assessment and anomaly detection
- Real-time learning algorithms for control and security of complex dynamical systems

Team


Jochen Cremer


Peyman Mohajerin Esfahani


Haiwei Xie


Olayiwola Arowolo


Shabnam Khodakaramzadeh


Mert Karaçelebi


Basel Morsy


Mohammad Boveiri


Ali Rajaei


Jochen Stiasny


Viktor Zobernig


Demetris Chrysostomou


Luca Hofstadler


Paul Bannmüller


Betül Mamudi


Runyao Yu


Périne Cunat

TU Delft | AI Initiative

https://www.tudelft.nl/ai/delft-ai-energy-lab

DAIEnergyLab

DAI-Energy-Lab@tudelft.nl

Outline

Reliability management and data in control rooms

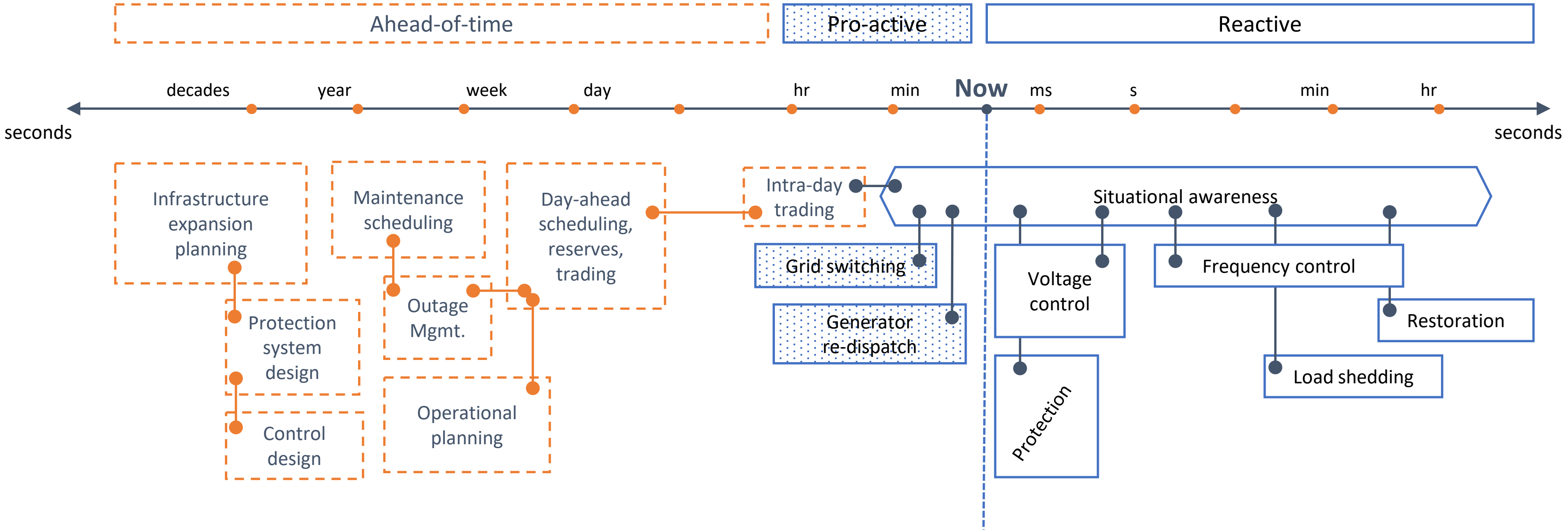
1. Introduction to reliability management
2. Machine learning approaches
3. Security assessment with cost-sensitive supervised learning

Learning models for secure system operation

4. Learning with domain knowledge
5. State estimation with graph neural networks
6. Weakly-supervised learning for secure operation
7. Challenges applying ML to reliability

Experts are in charge to manually operate the power system based on **experience** and with **the support of tools**

A complex process

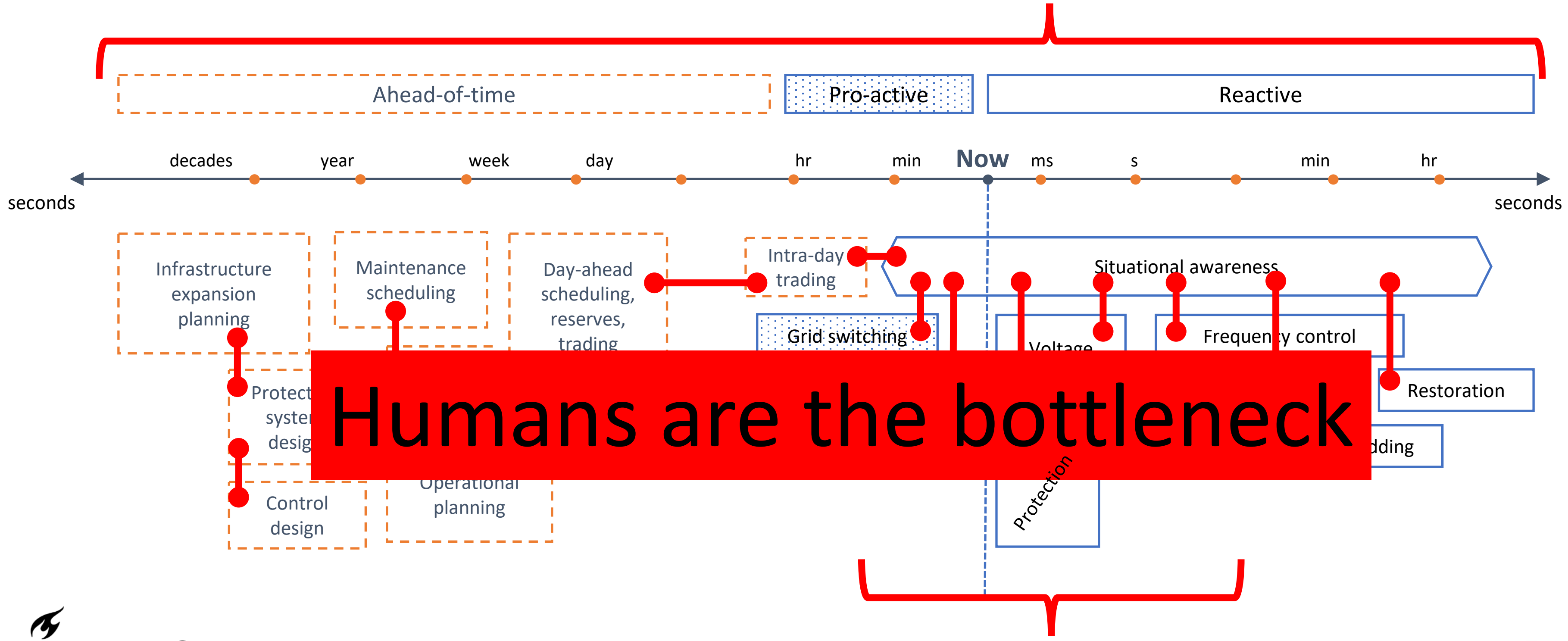


TU Delft

Hitachi Energy

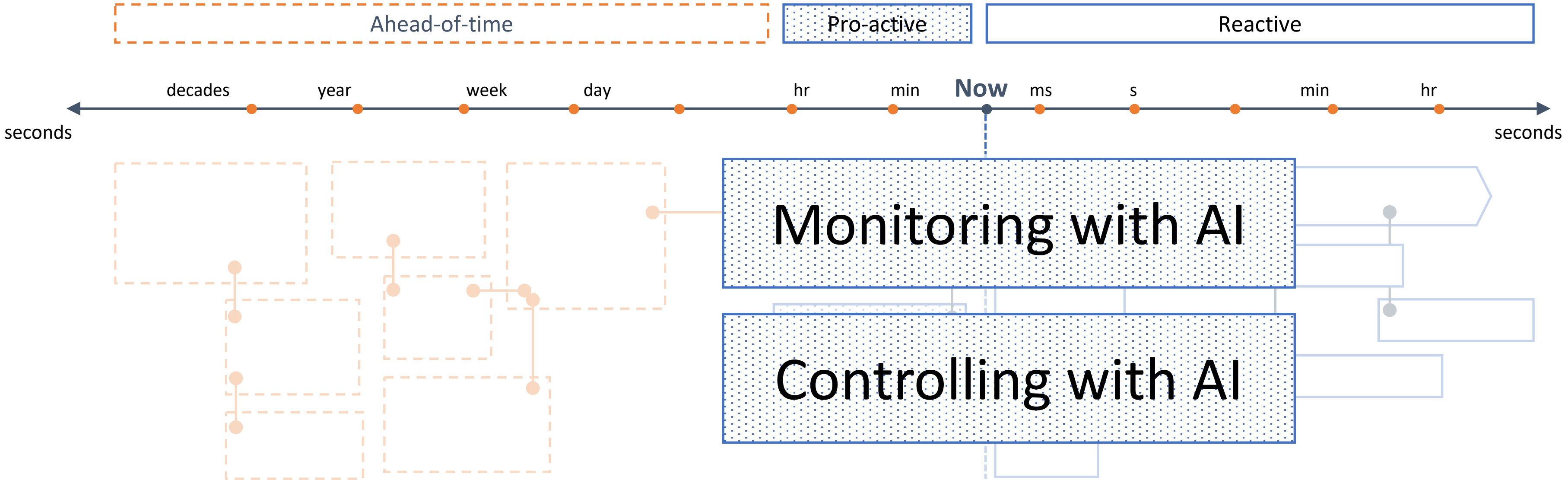
What’s the issue?

Interdependencies
challenge manual rules

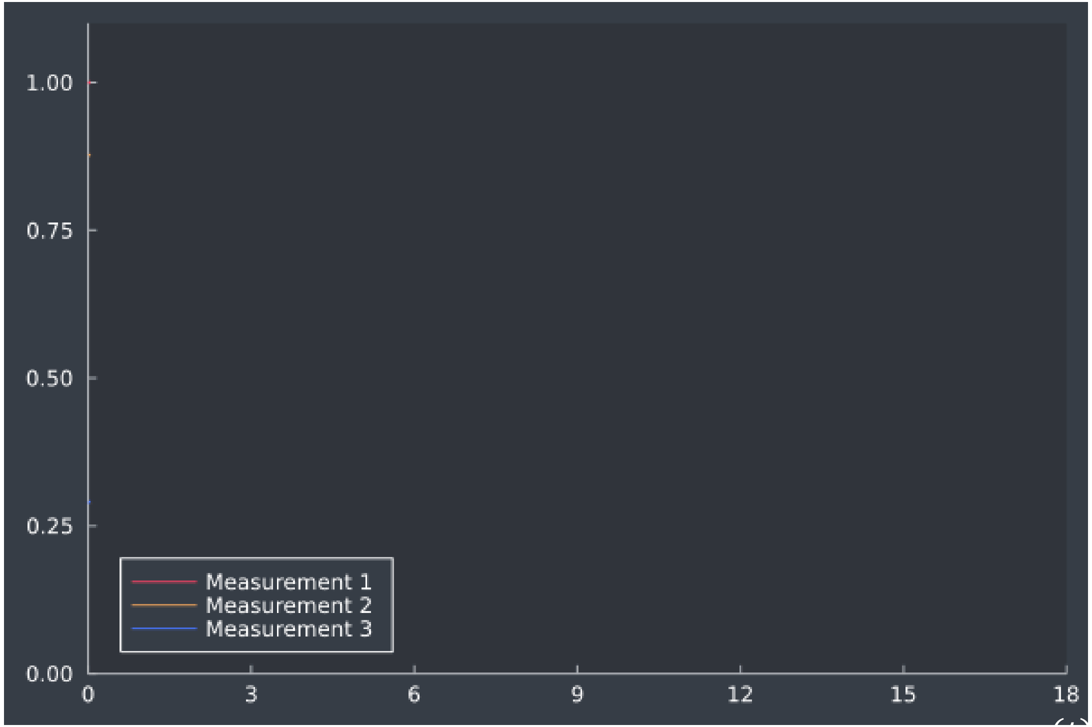


Human decisions are too slow

Automation first realised where urgently needed



Real-time security assessment of disturbances

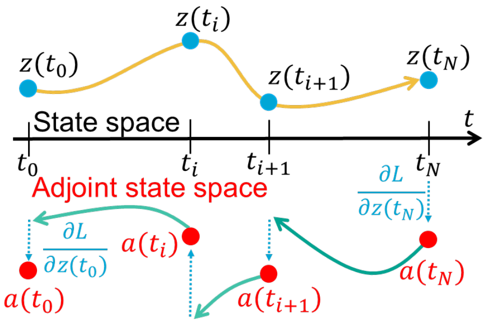


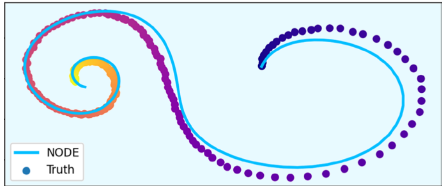
Phase angles [norm]

Time [s]

Mert Karaçelebi

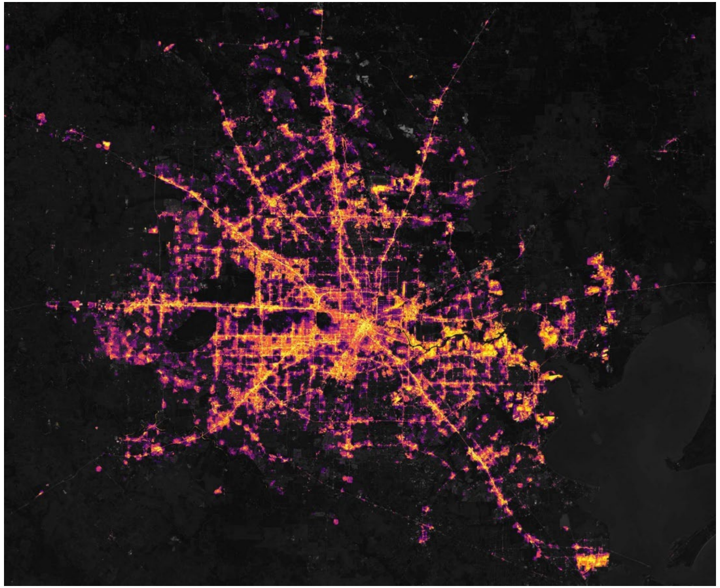
Neural ordinary differential equations





[1] Mert Karaçelebi, Jochen L. Cremer "Online Neural Dynamics Forecasting for Power System Security", *International Journal of Electrical Power & Energy Systems 2025*
[2] Mert Karaçelebi, Jochen L. Cremer, "Predicting Power System Frequency with Neural Ordinary Differential Equations", *12th Bulk Power System Dynamics and Control Symposium and Sustainable Energy, Grids and Networks Journal, 2025*

Why do system operators require reliability monitoring?



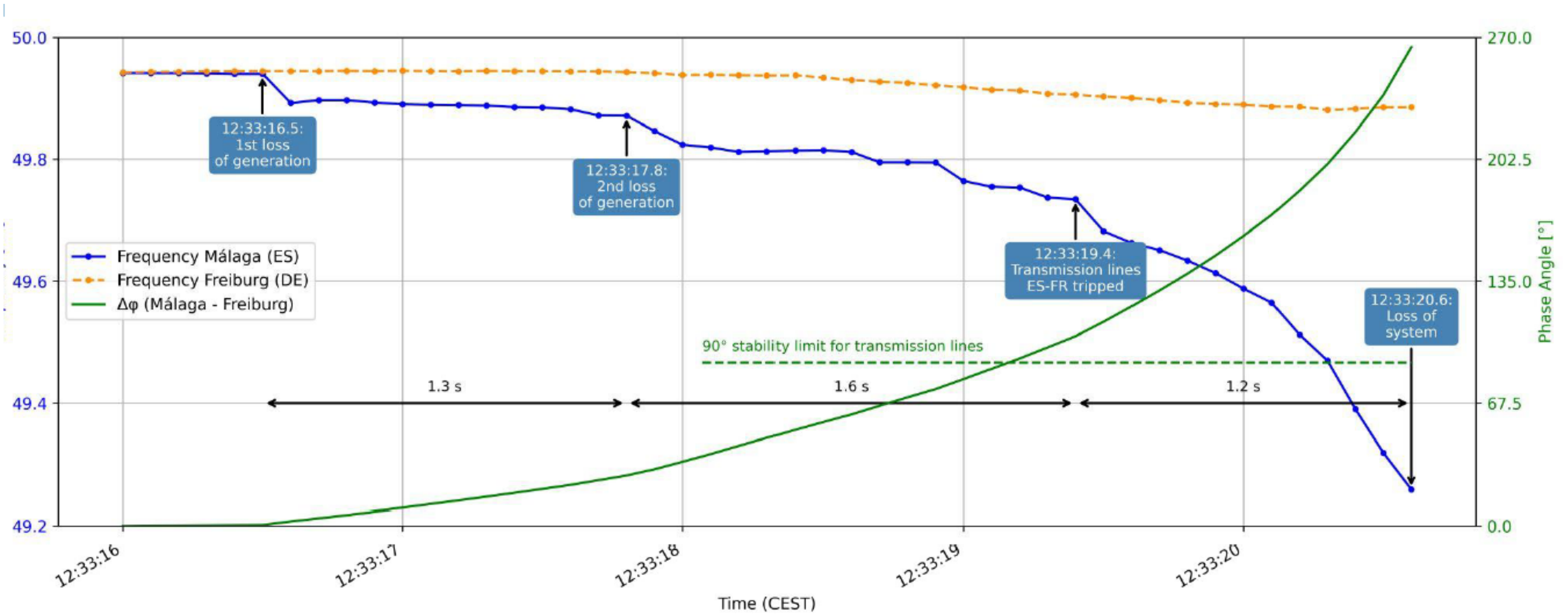
Houston, Texas 07 Feb 2021

Houston, Texas 16 Feb 2021

- Damages from the blackouts were estimated at **$195 billion [3]**
- Seconds away from a total power blackout in Texas

[3] 2021 Winter Storm Uri After-Action Review: Findings Report (PDF) (Report). City of Austin & Travis County. November 4, 2021. Archived (PDF) from the original on November 5, 2021. Retrieved November 5, 2021.

Power blackout 28 April 2025 Spain/Portugal

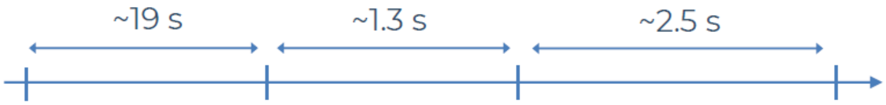


Source

Luis Badesa

N-2?

28th of April 2025 (CET)



Full Blackout

~12:20pm

FACT

Inter-area **oscillations** in Europe

(possibly related to later generation losses)

12:32:57

FACT

Generation loss

(size unknown)

12:33:16

FACT

Generation loss in the Southwest of Spain

(size unknown)

12:33:18

FACT

Generation loss in the Southwest of Spain

(size unknown)

12:33:20

FACT

Iberia becomes an **electric island**, and a "**massive disconnection of generation**" happens in Spain

https://badber.github.io/

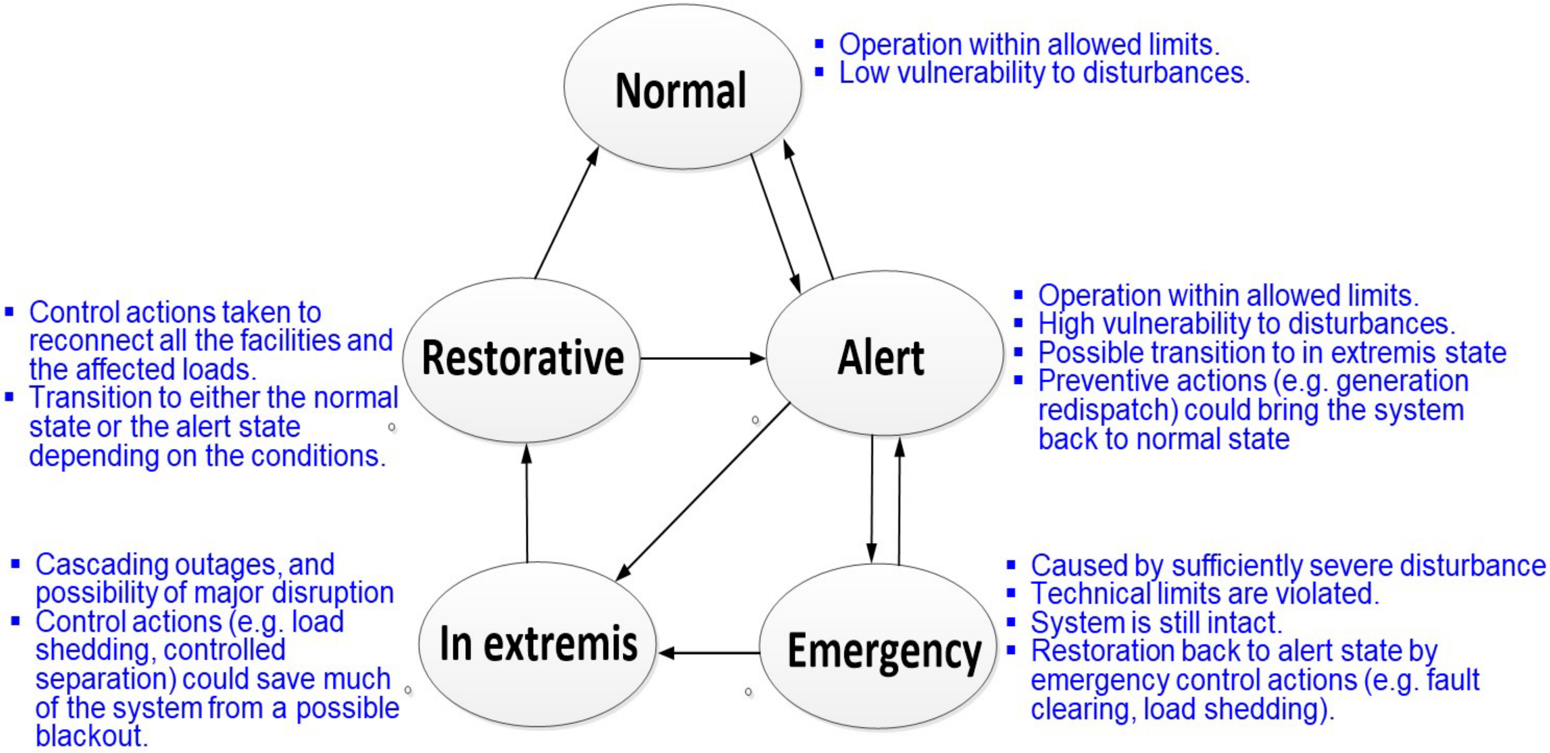
TU Delft

Power system reliability

"…is the probability that an electrical power system can perform a required function under given conditions for a given time interval."

[4] International Electrotechnical Commission, Electropedia: The worlds online electrotechnical vocabulary, http://www.electropedia.org/; 2016 [accessed 16 August 2018].

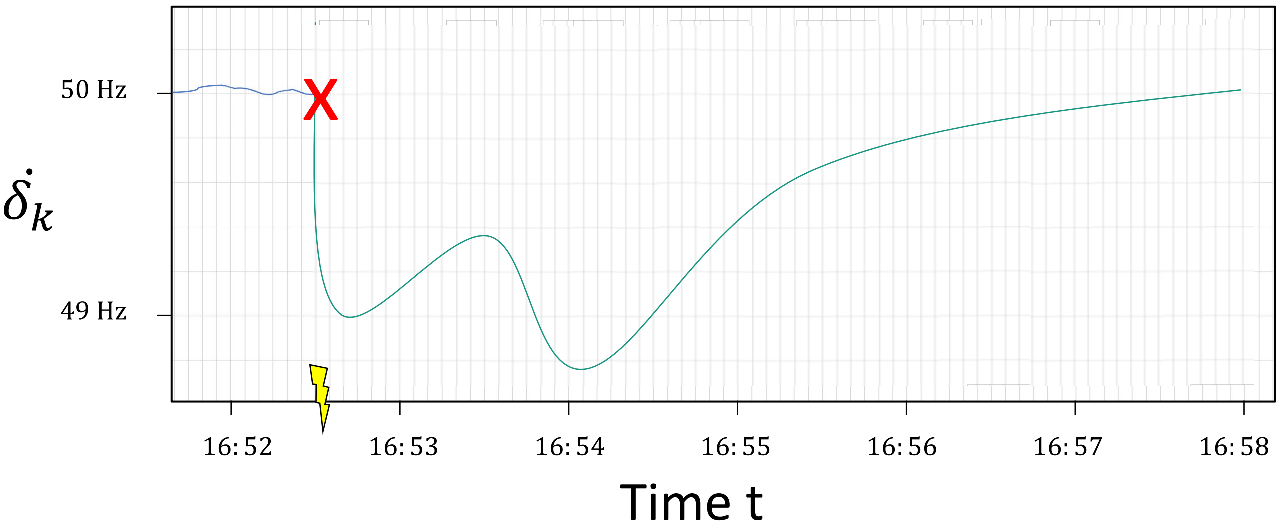
Operating states of power systems



[5] Power System Stability and Control (Kundur)

Conventional (offline) dynamic security assessment

Simulating time-response



Numerical integration

ODE system $\begin{cases} \dot{x} = f(x, t, x_0) \\ x_0 = (P_k^{16h}, Q_k^{16h}) \end{cases}$

Forward Euler

$$x_{k+1} = x_k + hf(x_k, t_k)$$

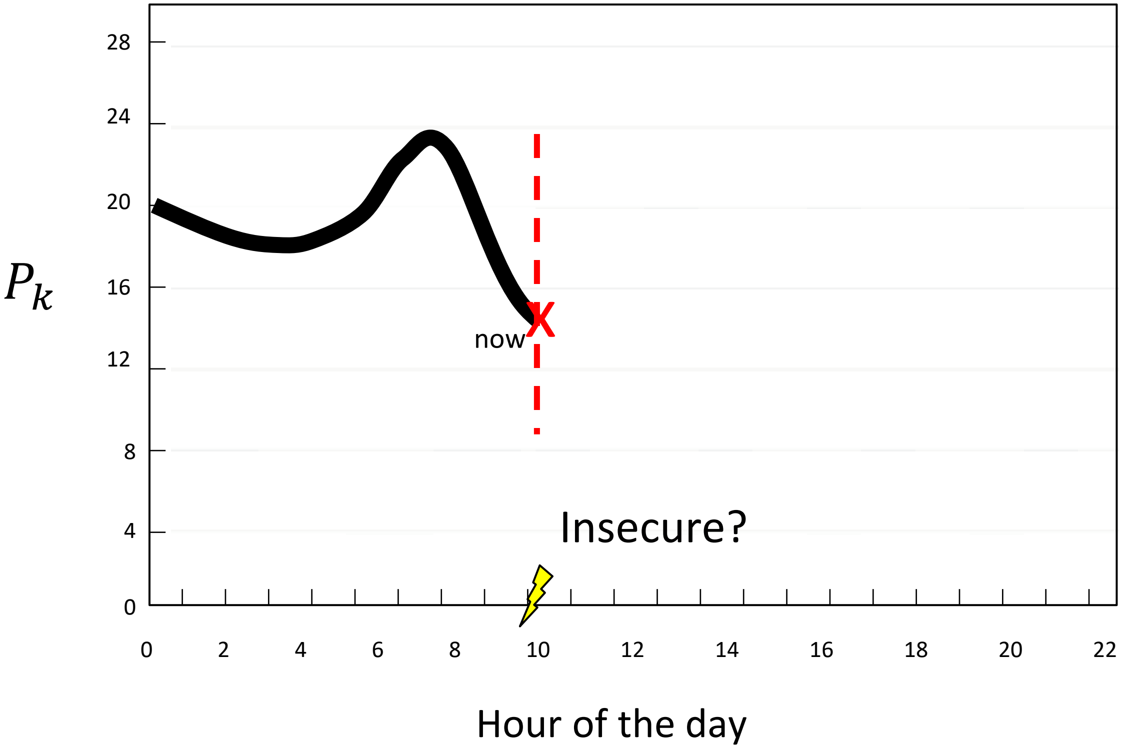
slow for large systems

TU Delft

Real-time dynamic security assessment

Objective: predicting security in real-time

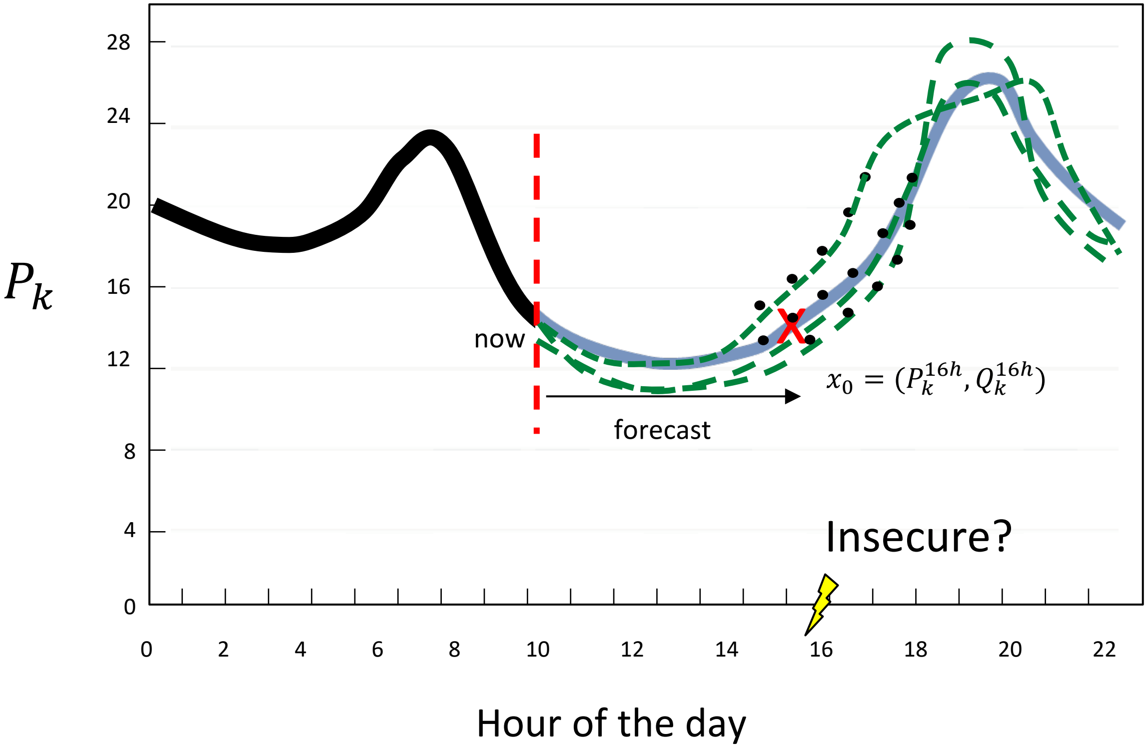
In response, use corrective actions in (near) real-time

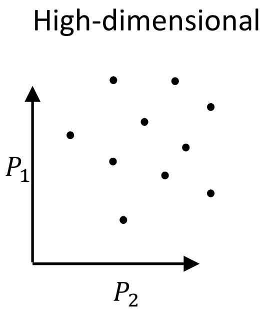


(Preventive) real-time dynamic security assessment

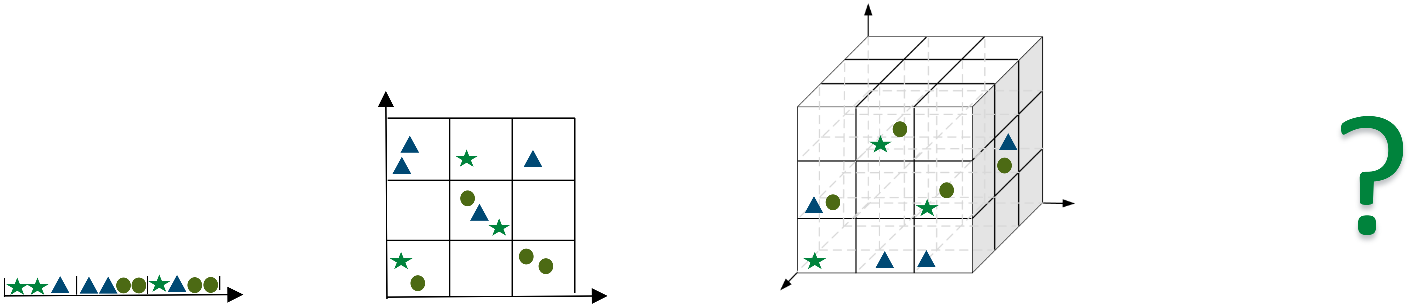
For N-1 security

Preventive actions





Curse of Dimensionality



1d: 3 regions

2d: 3^2 regions

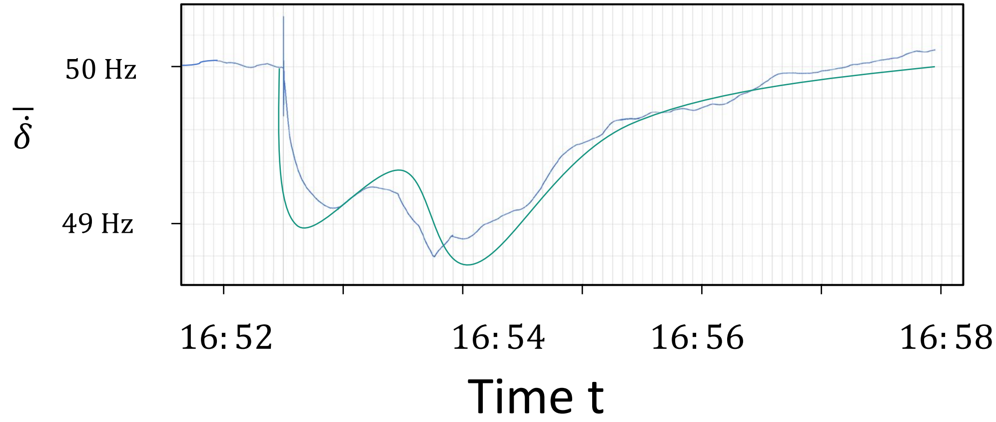
3d: 3^3 regions

1000d: hopeless

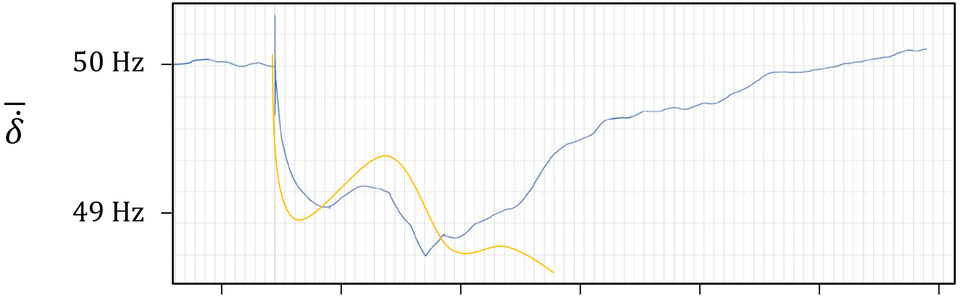
As dimensionality grows: fewer samples per region.

[6] James, Gareth, et al. *An introduction to statistical learning*. Vol. 112. No. 1. New York: springer, 2013.

Security of power systems

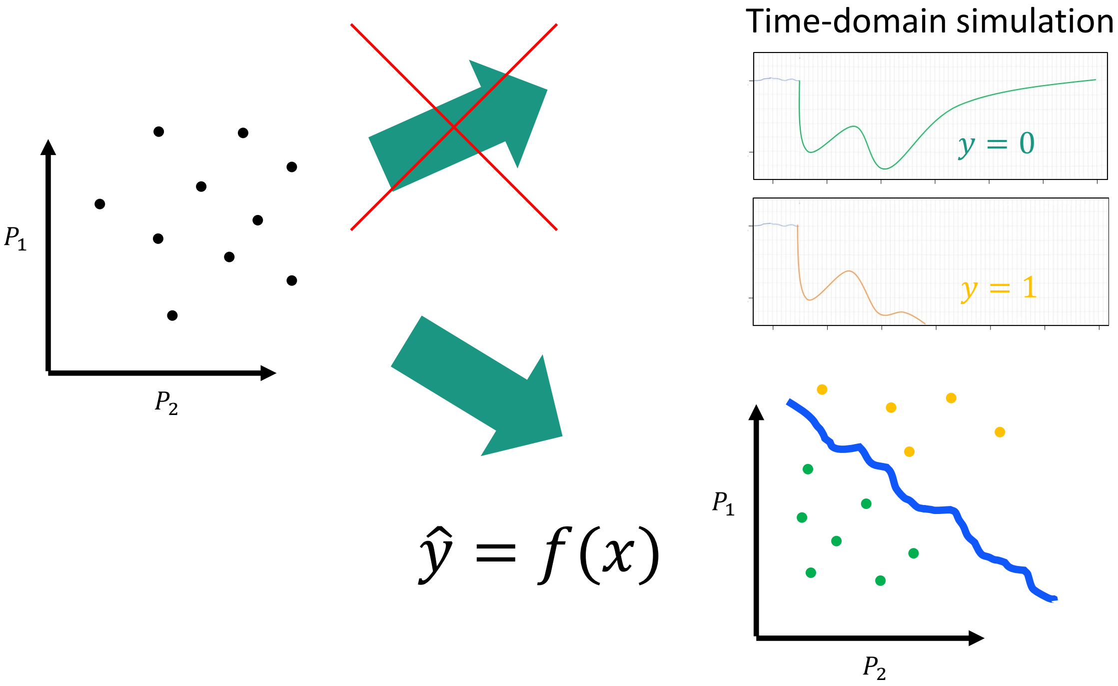


secure $y = 0$



insecure $y = 1$

Machine learning model to predict security



$$\hat{y} = f(x)$$

How to train and use f?

[7] Duchesne, L., Karangelos, E., & Wehenkel, L. (2020). Recent developments in machine learning for energy systems reliability management. *Proceedings of the IEEE*, *108*(9), 1656-1676.a

Challenges for reliability management

- More extreme weather events
- Higher grid load in the system
- Higher uncertainty
- Highly complex problem

Opportunities for reliability management with AI

- Availability of better models and data (weather, grid data, etc)
- New AI techniques
- Once trained, models are quick in ‘predicting’, but challenges also exist

Outline

Reliability management and data in control rooms

1. Introduction to reliability management
2. Machine learning approaches
3. Security assessment with cost-sensitive supervised learning

Learning models for secure system operation

4. Learning with domain knowledge
5. State estimation with graph neural networks
6. Weakly-supervised learning for secure operation
7. Challenges applying ML to reliability

TU Delft

Supervised Learning for Surrogate Models


Federica Bellizio

Olayiwola Arowolo

Notation: Power system s, model m, parameter x

Objective: assess $m(x) \to y$ **very** fast and **often**

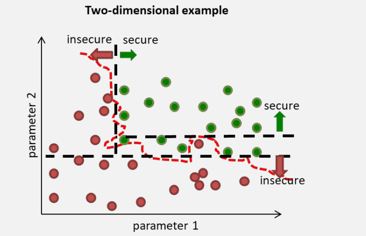
Surrogate approach

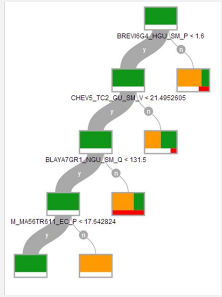
1. Generate a training dataset $\Omega^T = \{(x_i, y_i)\}_{i=1}^N$ where $y_i = m(x_i)$ from the full simulator
2. Train surrogate $f(x) \to \hat{y}$ with supervised loss $\sum_{i\in\Omega^T} \|y_i - \hat{y}_i\|$
3. Use $f(x_j)$ for new j $\notin \Omega^T$

Benefit: speed at inference

Applications

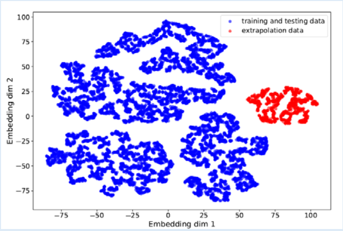
- Real-time dynamic security assessment ([8,9] and many others)





Challenges

- What if s and m changes? e.g., topology changes
- What if the model is inaccurate $s \neq m$? e.g., inverter-based controls
- Need large, representative training data



[8] Wehenkel, Louis, and Mania Pavella. "Decision trees and transient stability of electric power systems." *Automatica* 27.1 (1991): 115-134.
[9] Cremer, Jochen L., Ioannis Konstantelos, and Goran Strbac. "From optimization-based machine learning to interpretable security rules for operation." *IEEE Transactions on Power Systems* 34.5 (2019): 3826-3836.
[10] Olayiwola Arowolo, Jochen Stiasny, Jochen L. Cremer, "Exploring Extrapolation of Machine Learning Models for Power System Time Domain Simulation", *Bulk Power System Dynamics and Control Symposium, 2025*

Physics-Informed Learning

Objective: surrogate learning enhanced with physics knowledge from model m

Idea: Incorporate physics residual (e.g. from a PDE or simulator) to guide learning and improve generalization

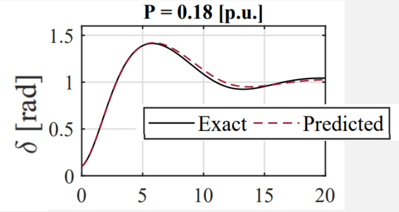
Physics-informed approach

1. Generate offline training dataset $\Omega^T = \{(x_i, y_i)\}_{i=1}^{N}$ with $y_i = m(x_i)$
2. Train surrogate $f(x) \rightarrow \hat{y}$ on composite loss $\sum_{i \in \Omega^T} \|y_i - \hat{y}_i\| + \mathcal{L}_{phys}\,(f(x_i), m)$
3. Use $f(x_j)$ for new j $\notin \Omega^T$

Benefits: Better generalisation performance with **fewer training samples**

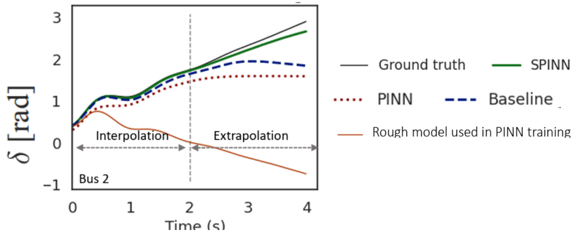
Applications

- Extrapolation in time-domain for dynamic analysis in power systems



Challenges

- Model inaccuracy $s \neq m$
- Changes in s or m
- Data sparsity
- Multi-loss scaling causes training instability
- Scaling issues to many physical loss terms in power systems



[11] Misyris, George S., Andreas Venzke, and Spyros Chatzivasileiadis. "Physics-informed neural networks for power systems." In *2020 IEEE power & energy society general meeting (PESGM)*, pp. 1-5. IEEE, 2020.
[12] Xie, Haiwei, Federica Bellizio, Jochen L. Cremer, and Goran Strbac. "Regularised Learning with Selected Physics for Power System Dynamics." In *2023 IEEE Belgrade PowerTech*, pp. 1-7. IEEE, 2023.

Weakly-Supervised (E2E) Learning

Objective: learn models $f(x)$ for downstream task even when exact labels $y_i = m(x_i)$ from the simulator m are unavailable, uncertain, or only indirectly defined.

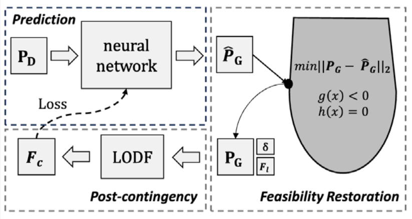
Approach

1. Generate many inputs $\Omega^T = \{(x_i)\}_{i=1}^N$
2. Model task loss $\sum_{i \in \Omega^T} \mathcal{L}(m(f(x_i))$
3. Use $f(x_j)$ for new j $\notin \Omega^T$

Benefits: learning for computationally expensive or ill-defined problems

Applications

- Learn to predict effective inputs to OPF[13]
- Replace conventional solvers with NN [14]
- Distribution system state estimation [15]
- N-k security-constrained OPF [16]



Challenges

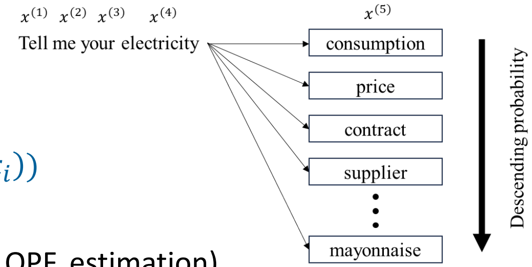
- Inexact supervision $s \neq m$ not so important as success defined by task-loss
- System shift in s or m
- Data coverage. Diverse samples are needed for generalization

[13] Donti, P., Amos, B., & Kolter, J. Z. (2017). Task-based end-to-end model learning in stochastic optimization. *Advances in neural information processing systems*, *30*.
[14] Donon, B., Liu, Z., Liu, W., Guyon, I., Marot, A., & Schoenauer, M. (2020). Deep statistical solvers. *Advances in Neural Information Processing Systems*, *33*, 7910-7921.
[15] Habib, B., Isufi, E., van Breda, W., Jongepier, A., & Cremer, J. L. (2023). Deep statistical solver for distribution system state estimation. *IEEE Transactions on Power Systems*, *39*(2), 4039-4050.
[16] Giraud, B. N., Rajaei, A., & Cremer, J. L. (2024). Constraint-driven deep learning for n-k security constrained optimal power flow. *Electric Power Systems Research*, *235*, 110692.

Self-Supervised Learning

Objective: Learn a **useful internal representation** from unlabeled data by solving a **pretext task** — no human-labeled or simulator-labeled outputs required.

Idea: instead of training on (x_i, y_i) train on auto-generated pseudo-labels or tasks constructed from structure x_i



Approach

1. Generate many inputs $\Omega^T = \{(x_i\ )\}_{i=1}^N$
2. Define self-supervised pretext loss $\mathcal{L}_{pretext}(f(x_i))$
3. Train encoder $\sum_{i \in \Omega^T} \mathcal{L}_{pretext}(f(x_i))$
4. Use $f(x)$ for downstream *task* (e.g. forecasting, OPF, estimation)

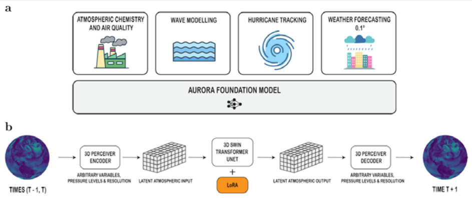
Benefits: Good initialization when little data, good transfer to downstream tasks

Challenges

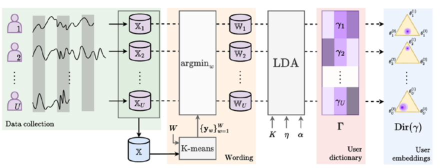
- Design pretext loss and model architectures with broad set of tasks, grid conditions, topologies
- Generate large data sets
- …

Applications

- Natural Language Processing
- Weather foundational models
- Earth system foundational models [17]



Load forecasting of users [18]



Grid foundation models (GFM) [19]

[17] Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Vaughan, A., Brandstetter, J., ... & Perdikaris, P. (2024). A foundation model for the earth system. *arXiv preprint arXiv:2405.13063*.
[18] Bölat, Kutay, and Simon Tindemans. "GUIDE-VAE: Advancing Data Generation with User Information and Pattern Dictionaries." *arXiv preprint arXiv:2411.03936* (2024).
[19] Hamann, H. F., Gjorgiev, B., Brunschwiler, T., Martins, L. S., Puech, A., Varbella, A., ... & Sobolevsky, S. (2024). Foundation models for the electric power grid. Joule, 8(12), 3245-3258.

Graph Neural Networks

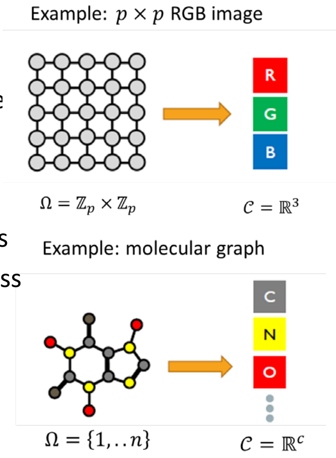
Objective: Improve generalization performance in learning tasks on network-structured systems (like power grids)

Idea: embedding graph topology directly into the model architecture as bias

Approach

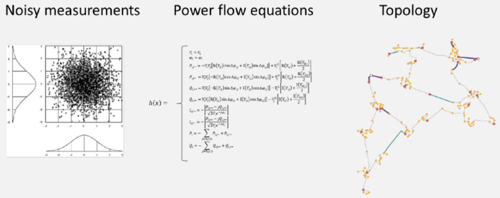
1. Construct graph $G = (V, \mathcal{E})$ with features on nodes and edges
2. Define f_{GNN} and learn with message passing on supervised loss $\sum_{i \in \Omega^T} \|y_i - \hat{y}_i\|$
3. Use $f(x_j)$ for new j $\notin \Omega^T$ or on unseen graphs G'

Benefits: Data efficient, generalisation to changes in topologies



Applications

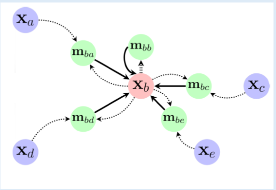
- Graph neural solvers [20] for ACOPF [21]
- Distribution system state estimation [22]



Challenges

- Model inaccuracy $s \neq m$
- Long-range dependencies are difficult to learn. *Power system topology is sparse*
- Challenging to learn for *global* problems (e.g. ACOPF)

Good to learn local relationships



TU Delft

[20] Donon, B., Donnot, B., Guyon, I., & Marot, A. (2019, July). Graph neural solver for power systems. In *2019 international joint conference on neural networks (ijcnn)* (pp. 1-8). IEEE.
[21] Piloto, L., Liguori, S., Madjiheurem, S., Zgubic, M., Lovett, S., Tomlinson, H., ... & Witherspoon, S. (2024). Canos: A fast and scalable neural ac-opf solver robust to n-1 perturbations. *arXiv preprint arXiv:2403.17660*.
[22] B. Habib, E. Isufi, W. van Breda, A. Jongepier and J. L. Cremer, "Deep Statistical Solver for Distribution System State Estimation," in IEEE Transactions on Power Systems, vol. 39, no. 2, pp. 4039-4050, March 2024, doi: 10.1109/TPWRS.2023.3290358.

Outline

Reliability management and data in control rooms

1. Introduction to reliability management
2. Machine learning approaches
3. Security assessment with cost-sensitive supervised learning

Learning models for secure system operation

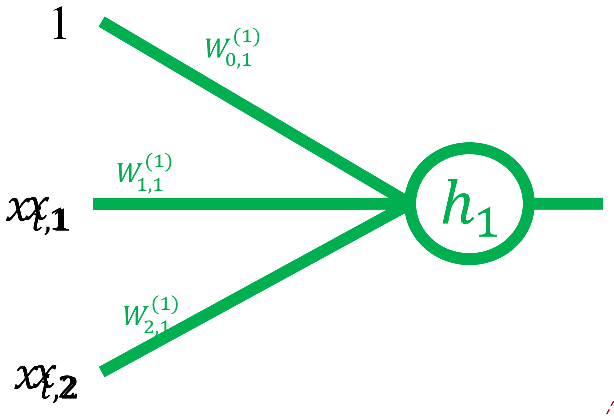
4. Learning with domain knowledge
5. State estimation with graph neural networks
6. Weakly-supervised learning for secure operation
7. Challenges applying ML to reliability

TU Delft

Neural networks

One neuron

Here simplified notation $x_{i,1}$ to x_1

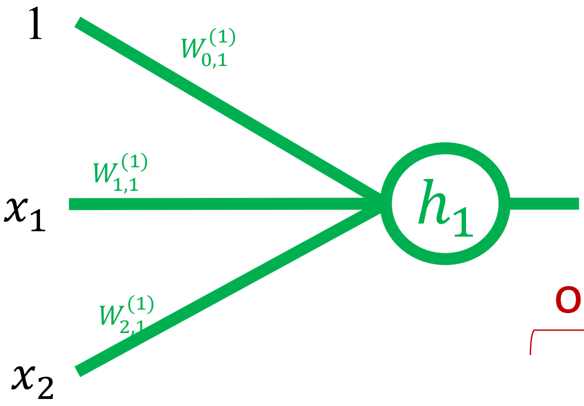


Activation function

$$h_1 = a(w_{0,1}^{(1)} + x_1 w_{1,1}^{(1)} + x_2 w_{2,1}^{(1)})$$

Neural networks

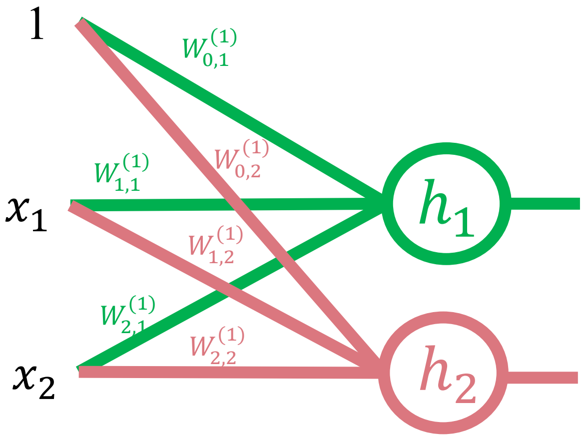
One neuron



$$h_1 = \overbrace{a(\overbrace{w_{0,1}^{(1)} + x_1 w_{1,1}^{(1)} + x_2 w_{2,1}^{(1)}}^{\text{in}_{h1}})}^{\text{out}_{h1}}$$

TU Delft

Compact model

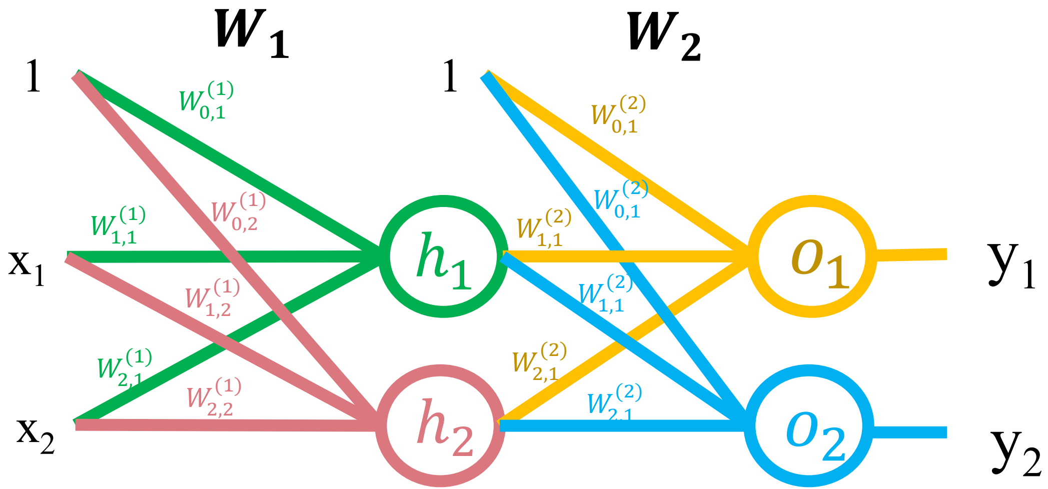


$$\mathring{\mathbf{a}}(\mathbf{W}_1^T \mathring{\mathbf{x}})$$

$$\boldsymbol{W}_1 = \begin{bmatrix} W_{0,1}^{(1)} & W_{0,2}^{(1)} \\ W_{1,1}^{(1)} & W_{1,2}^{(1)} \\ W_{2,1}^{(1)} & W_{2,2}^{(1)} \end{bmatrix}$$

$(N+1) \times u1$

Multiple layers



$\mathring{\mathbf{a}}(\mathbf{W}_1^T \mathring{\mathbf{x}})$ $\mathring{\mathbf{a}}(\mathbf{W}_2^T \mathring{\mathbf{a}}(\mathbf{W}_1^T \mathring{\mathbf{x}}))$

$$\boldsymbol{W_1} = \begin{bmatrix} W_{0,1}^{(1)} & W_{0,2}^{(1)} \\ W_{1,1}^{(1)} & W_{1,2}^{(1)} \\ W_{2,1}^{(1)} & W_{2,2}^{(1)} \end{bmatrix}$$

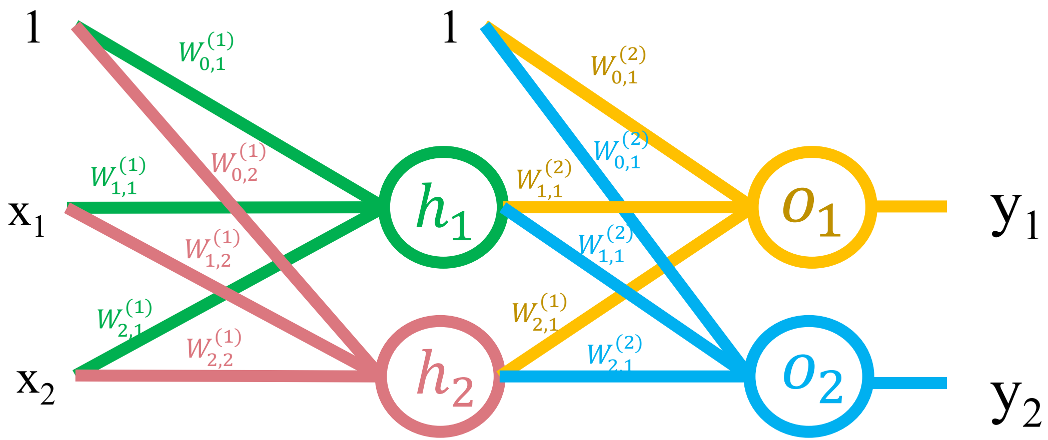
$(N+1) \times u1$

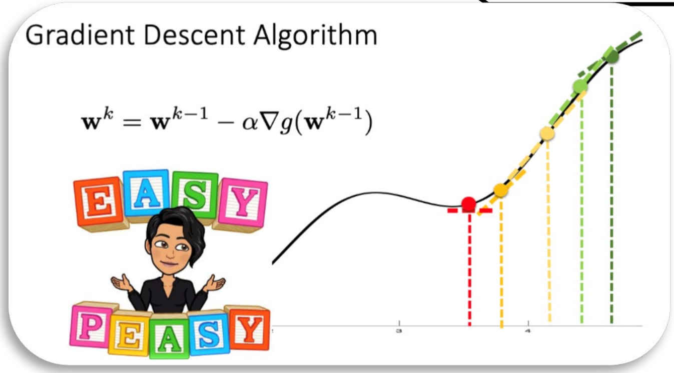
$$\boldsymbol{W_2} = \begin{bmatrix} W_{0,1}^{(2)} & W_{0,2}^{(2)} \\ W_{1,1}^{(2)} & W_{1,2}^{(2)} \\ W_{2,1}^{(2)} & W_{2,2}^{(2)} \end{bmatrix}$$

$(u1+1)$
$\times\ u2$

$$f(x) = \mathring{\mathbf{a}}(\mathbf{W}_2^T \mathring{\mathbf{a}}(\mathbf{W}_1^T \mathring{\mathbf{x}}))$$

Loss function





$$J = \frac{1}{|\Omega^{\mathrm{T}}|}\sum_{i\in\Omega^T}\left(\left(o_{i,1}-y_{i,1}\right)^2+\left(o_{i,2}-y_{i,2}\right)^2\right)$$

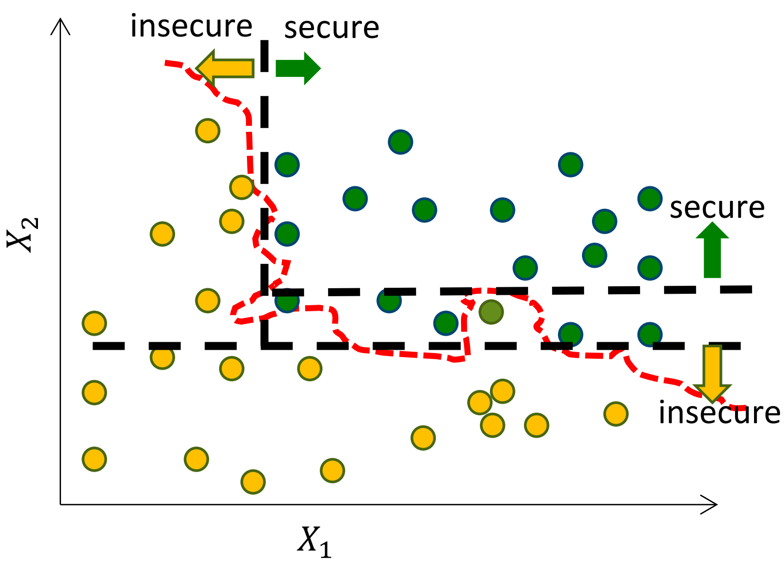
System operation

Experts are in charge to manually operate the power system based on **experience** and with **the support of tools**

Interpretable models

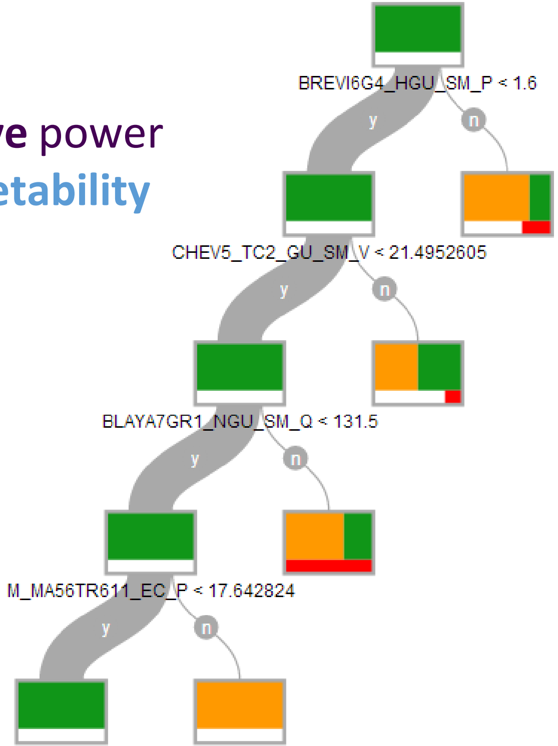
Decision Trees as a model?

Two-dimensional example

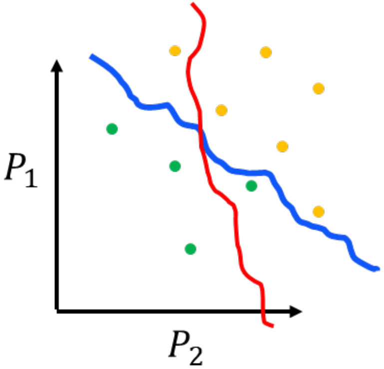


Decision trees:

- Limited **expressive** power
- Fantastic **interpretability**

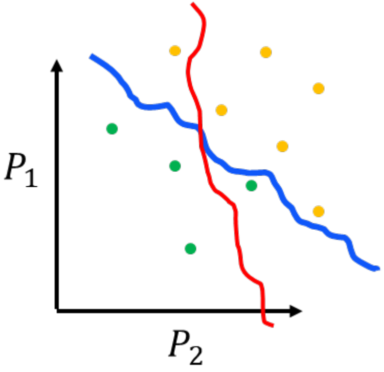


Metrics for classification



Predicted and actual security limits are different!

Metrics for classification

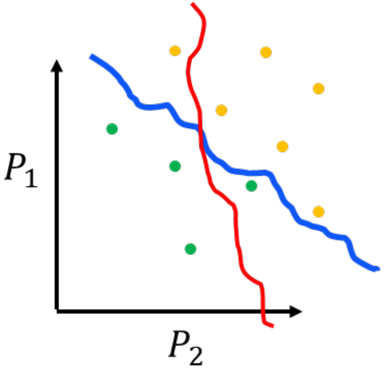


Predicted and actual security limits are different!

		Predicted Class: Positive	Predicted Class: Negative
True Class	Positive	TP	FN
	Negative	FP	TN

Confusion Matrix

Metrics for classification



Predicted and actual security limits are different!

Two types of accurate predictions:

TN: Is secure and we think it is secure (**GOOD**)

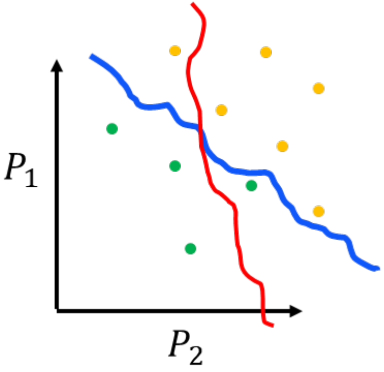
TP: Is insecure and we think it is insecure (**VERY GOOD!**)

True Class	Predicted Class: Positive	Predicted Class: Negative
Positive	TP	FN
Negative	FP	TN

Confusion Matrix

TU Delft

Metrics for classification



Predicted and actual security limits are different!

Predicted Class

		Positive	Negative
True Class	Positive	TP	FN
	Negative	FP	TN

Confusion Matrix

<u>**Two types of accurate predictions**</u>:

TN: Is secure and we think it is secure (**GOOD**)

TP: Is insecure and we think it is insecure (**VERY GOOD!**)

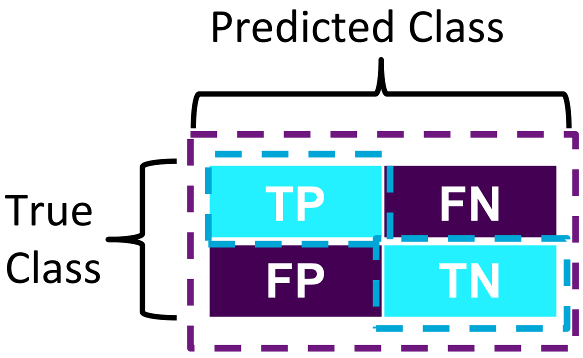
<u>**Two types of wrong predictions**</u>:

FP: Is secure but we think it is insecure (**BAD**)

FN: Is insecure but we think it is secure (**VERY BAD!**)

This can have a severe effect!

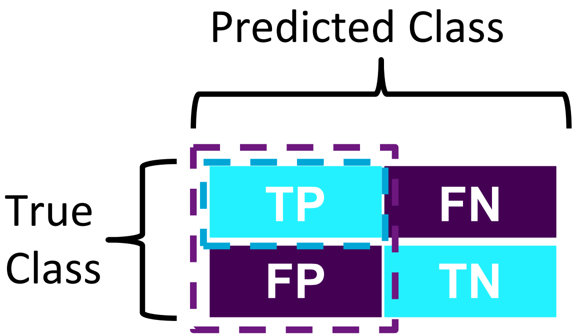
Metrics for classification



Ratio of correct predictions

$$\text{Accuracy} = \frac{N_{TP}+N_{TN}}{N_{FP}+N_{TP}+N_{FN}+N_{TN}}$$

Metrics for classification

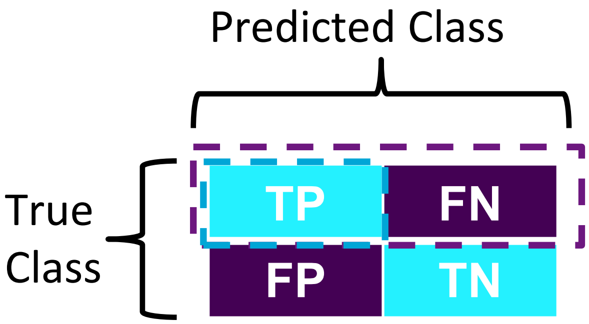


Ratio of correctly found insecure cases to predicted insecure predictions

$$\text{Accuracy} = \frac{N_{TP}+N_{TN}}{N_{FP}+N_{TP}+N_{FN}+N_{TN}}$$

$$\text{Precision} = \frac{N_{TP}}{N_{TP}+N_{FP}}$$

Metrics for classification



Ratio of correctly found insecure cases to all insecure cases

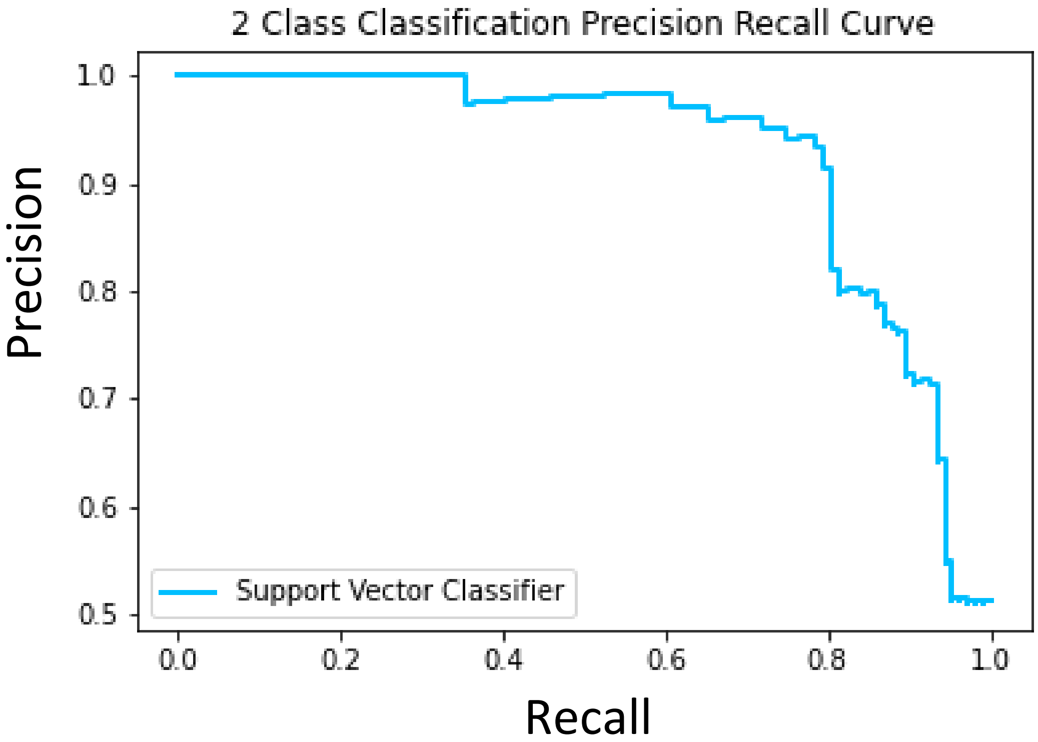
$$\text{Accuracy} = \frac{N_{TP}+N_{TN}}{N_{FP}+N_{TP}+N_{FN}+N_{TN}}$$

$$\text{Precision} = \frac{N_{TP}}{N_{TP}+N_{FP}}$$

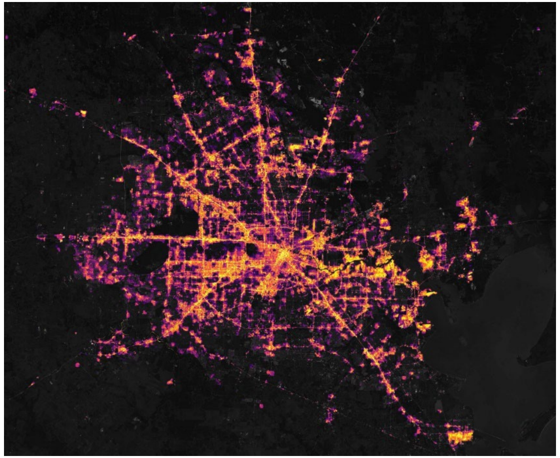
$$\text{Recall} = \frac{N_{TP}}{N_{TP}+N_{FN}}$$

Precision vs Recall

When do we observe the highest performance?



Blackout predictions: Precision or Recall?



Houston, Texas 07 Feb 2021

Houston, Texas 16 Feb 2021

$$\text{Precision} = \frac{N_{TP}}{N_{TP}+N_{FP}}$$

$$\text{Recall} = \frac{N_{TP}}{N_{TP}+N_{FN}}$$

TU Delft

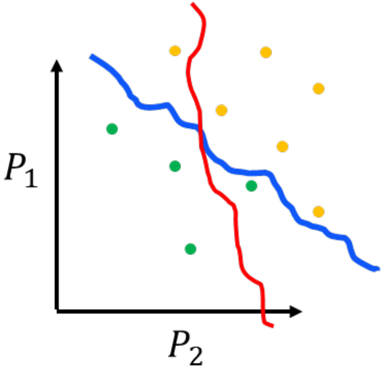
Cost skewness: $C_{FN} \gg C_{FP}$

Problem

- The two different false predictions have different costs.

- Damages from the blackouts were estimated at $195 billion
- Seconds away from a total power blackout in Texas

Metrics for classification



Predicted and actual security limits are different!

		Predicted Class: Positive	Predicted Class: Negative
True Class	Positive	0	C_{FN}
	Negative	C_{FP}	0

Two types of accurate predictions:

TN: Is secure and we think it is secure (**GOOD**)

TP: Is insecure and we think it is insecure (**VERY GOOD!**)

Two types of wrong predictions:

FP: Is secure but we think it is insecure (**BAD**)

FN: Is insecure but we think it is secure (**VERY BAD!**)

Two issues

- Cost-skewness: $C_{FN} \gg C_{FP}$
- Class imbalance: $\pi_{+} \ll \pi_{-}$

What a classifier can do

Classify points

- is x positive?

Rank points

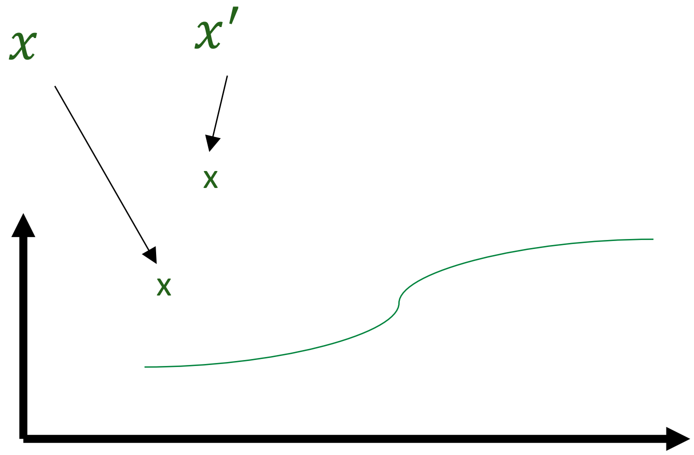
- Is x 'more positive' than x'?

Output a score $s(x)$

- 'How positive' is x?

Output a probability estimate $\hat{p}(x)$

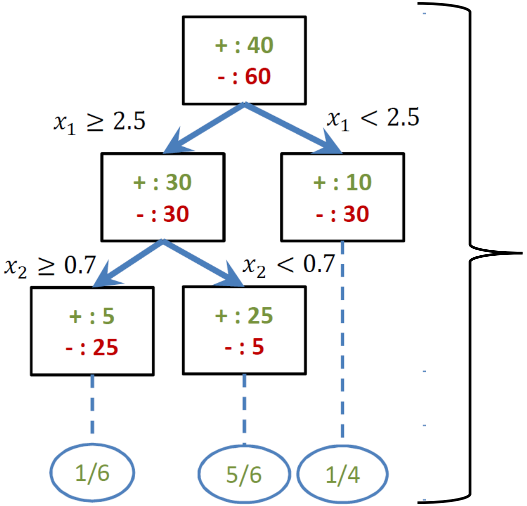
- What is the (estimated) probability that x is positive?



TU Delft

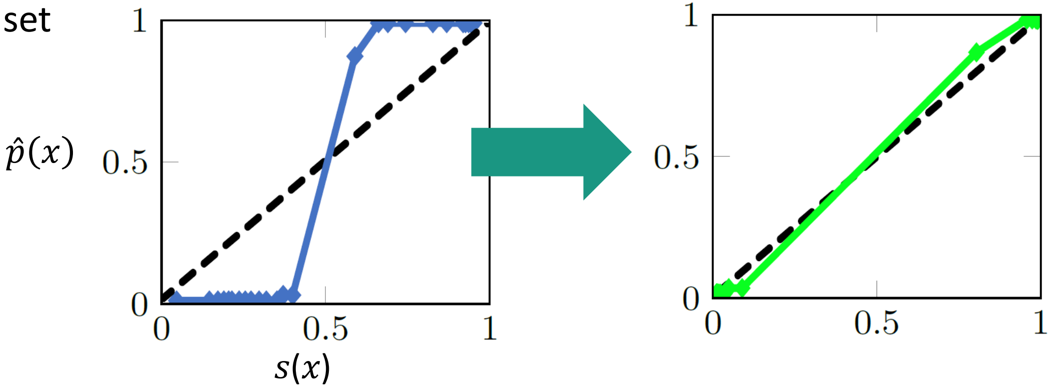
Probability estimation is not easy

- Scores $s(x) \in [0,1]$ as probability estimates $\hat{\boldsymbol{p}}(\boldsymbol{x})$? No!



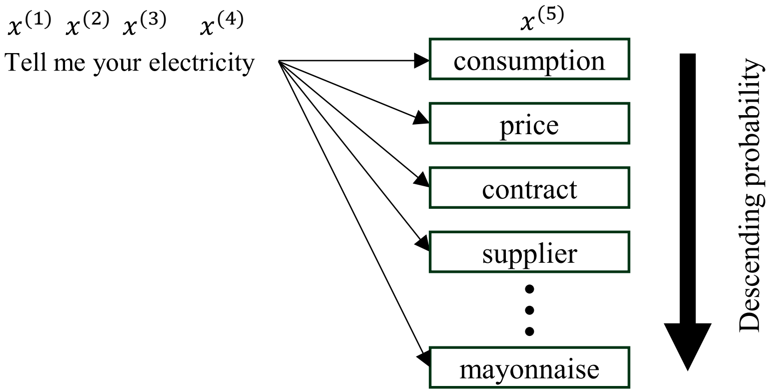
Based on training set

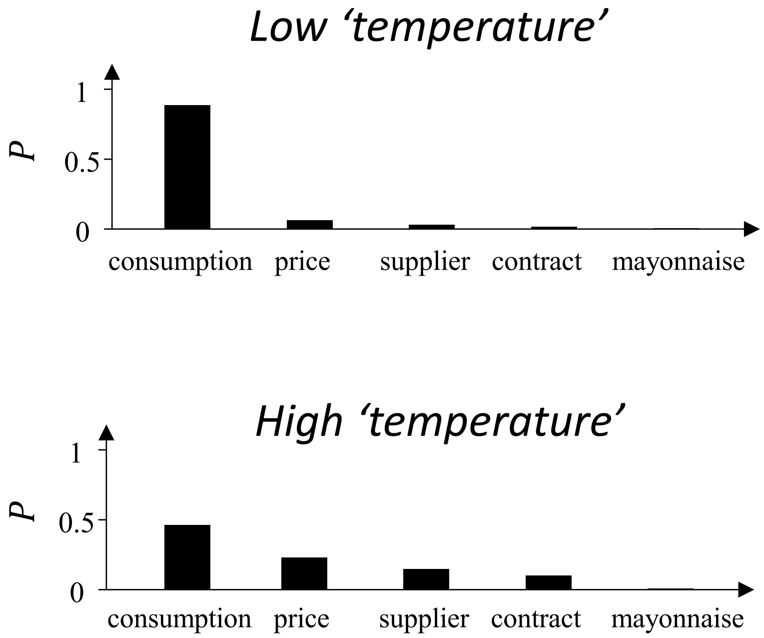
Platt scaling: Find $\hat{p}(x) = \frac{1}{1+e^{A\,s(x)+B}}$



[23] Nikolaou, N., Edakunni, N., Kull, M., Flach, P. and Brown, G., 2016. Cost-sensitive boosting algorithms: Do we really need them?. *Machine Learning*, *104*(2), pp.359-384.

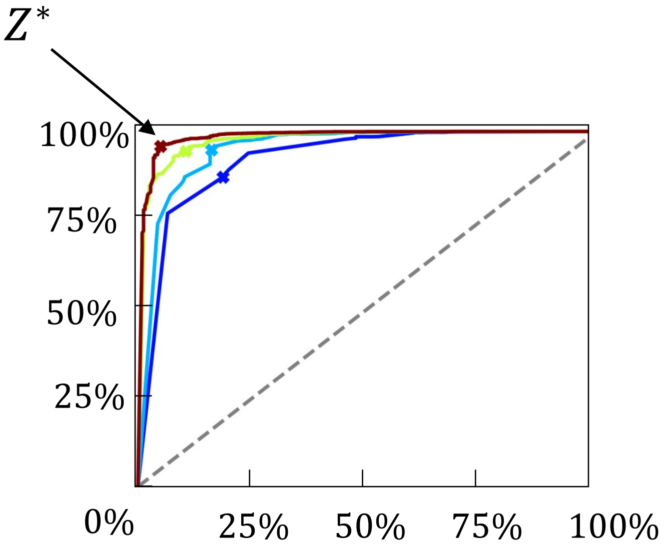
Calibration in Large Language Models





TU Delft

Cost-sensitive learning



$$TPR = \frac{TP}{\pi_+}$$

$$FPR = \frac{FP}{\pi_+}$$

$$Z^* = \frac{\Pi_- C_{FP}}{\Pi_- C_{FP} + \Pi_+ C_{FN}}$$

Probability of contingency

'Impact' of contingency

$\hat{\boldsymbol{p}}\,(\boldsymbol{x}) \geq \boldsymbol{Z}^*$ → predict secure

$\hat{\boldsymbol{p}}\,(\boldsymbol{x}) < \boldsymbol{Z}^*$ → predict insecure

[24] Jochen Cremer, Goran Strbac, "A Machine-learning based Probabilistic Perspective on Dynamic Security Assessment, *International Journal of Electrical Power and Energy Systems, 2021*

The risk of relying on machine learning

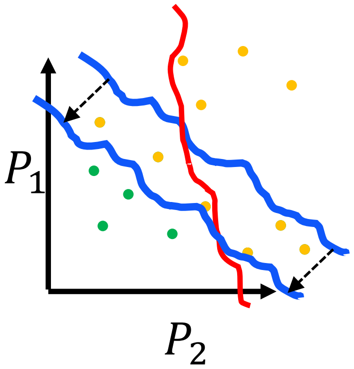
Step 1: Compute risks when predicting x_i as secure $\hat{y}_i = 1$ and insecure $\hat{y}_i$ =0

$$R_{\text{secure}} = p_i p_c \hat{\boldsymbol{p}}(\boldsymbol{x_i}) \boldsymbol{C_{FN}}$$

$$R_{\text{insecure}} = p_i(1 - p_c)(1 - \hat{\boldsymbol{p}}(\boldsymbol{x_i}))\boldsymbol{C_{FP}}$$

Step 2: Predict with lowest residual risk

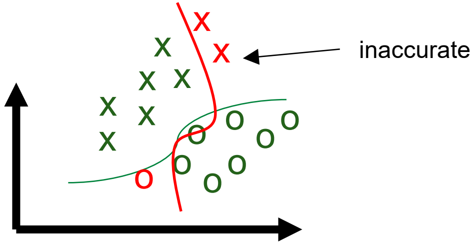
$$R_{\text{secure}} \bigvee R_{\text{insecure}}$$



[24] Jochen Cremer, Goran Strbac, “A Machine-learning based Probabilistic Perspective on Dynamic Security Assessment, *International Journal of Electrical Power and Energy Systems, 2021*

Minimize risks by hybrid approach

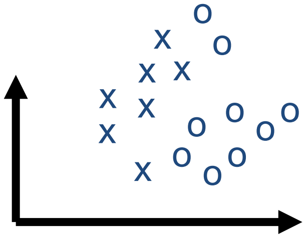
Machine Learning



- Fast
- Sometimes inaccurate

+

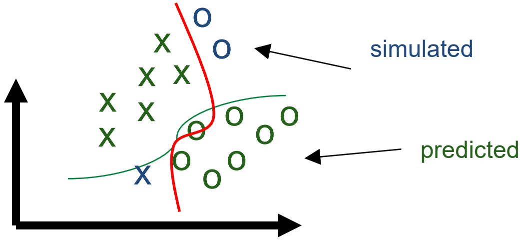
Simulator



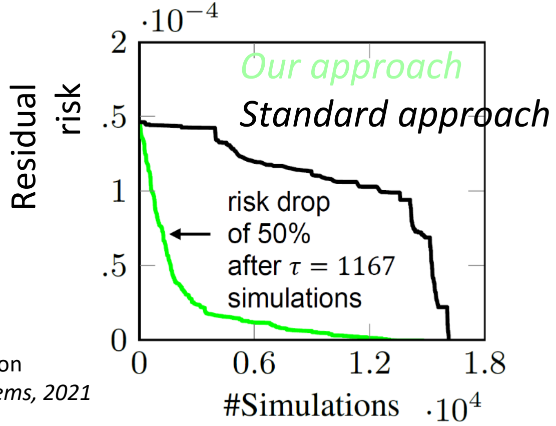
- Slow
- Always accurate

=

Probabilistic approach



Case study: French system



[24] Jochen Cremer, Goran Strbac, "A Machine-learning based Probabilistic Perspective on Dynamic Security Assessment, *International Journal of Electrical Power and Energy Systems, 2021*

Outline

Reliability management and data in control rooms

1. Introduction to reliability management
2. Machine learning approaches
3. Security assessment with cost-sensitive supervised learning

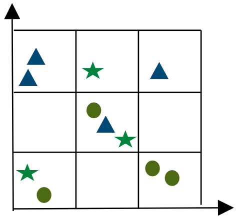
Learning models for secure system operation

4. Learning with domain knowledge
5. State estimation with graph neural networks
6. Weakly-supervised learning for secure operation
7. Challenges applying ML to reliability

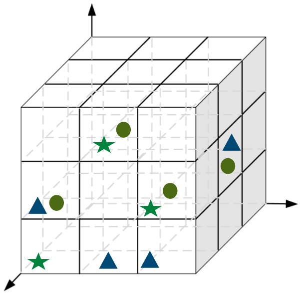
TU Delft

How to address curse of dimensionality?

1d: 3 regions



2d: 3^2 regions



3d: 3^3 regions

?

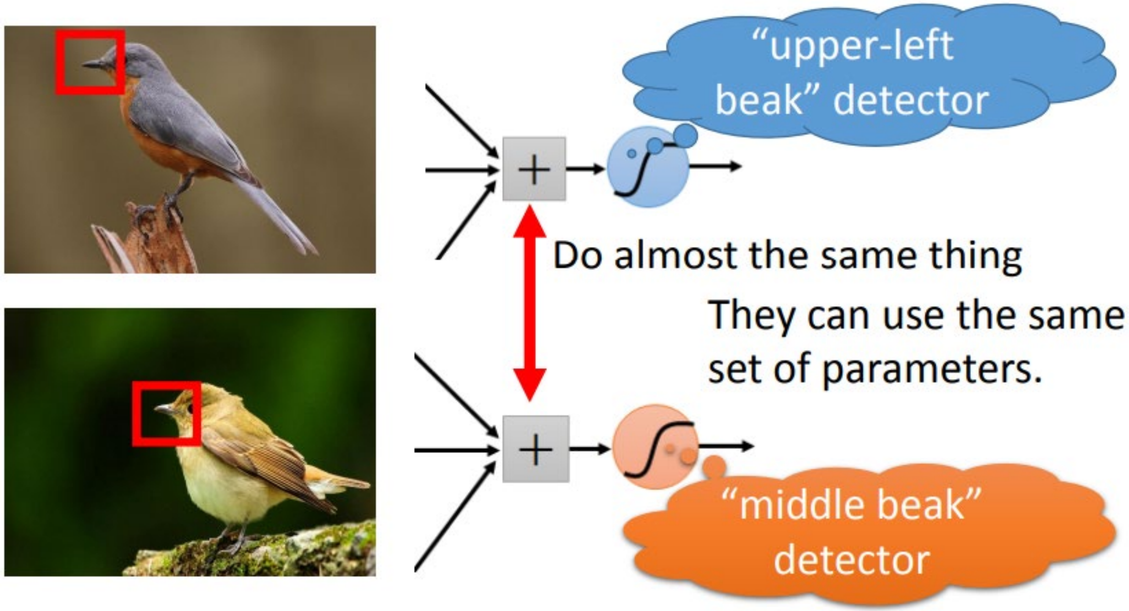
1000d: hopeless

As dimensionality grows: fewer samples per region.

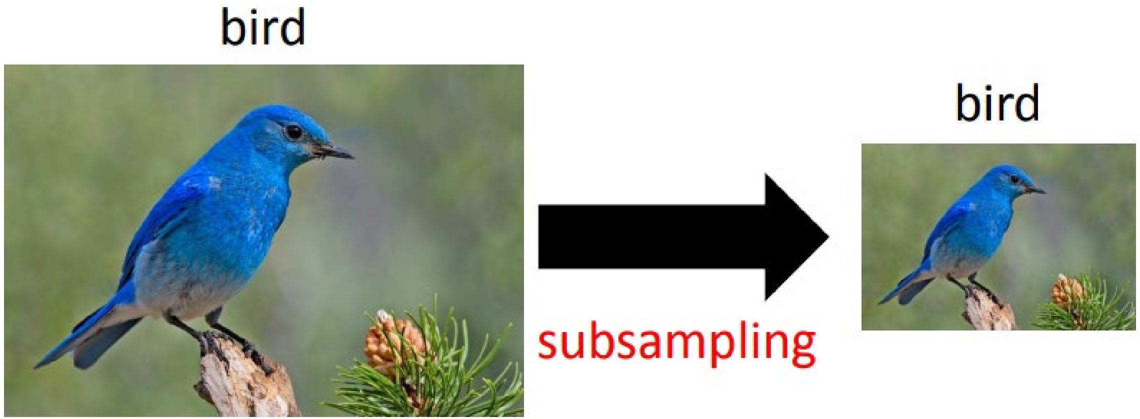
AI can well predict with images

Property 1: Some patterns are much smaller than the whole image. A neuron does not have to see the whole image to discover the pattern.

Property 2: The same patterns appear in different regions. (translated invariance)



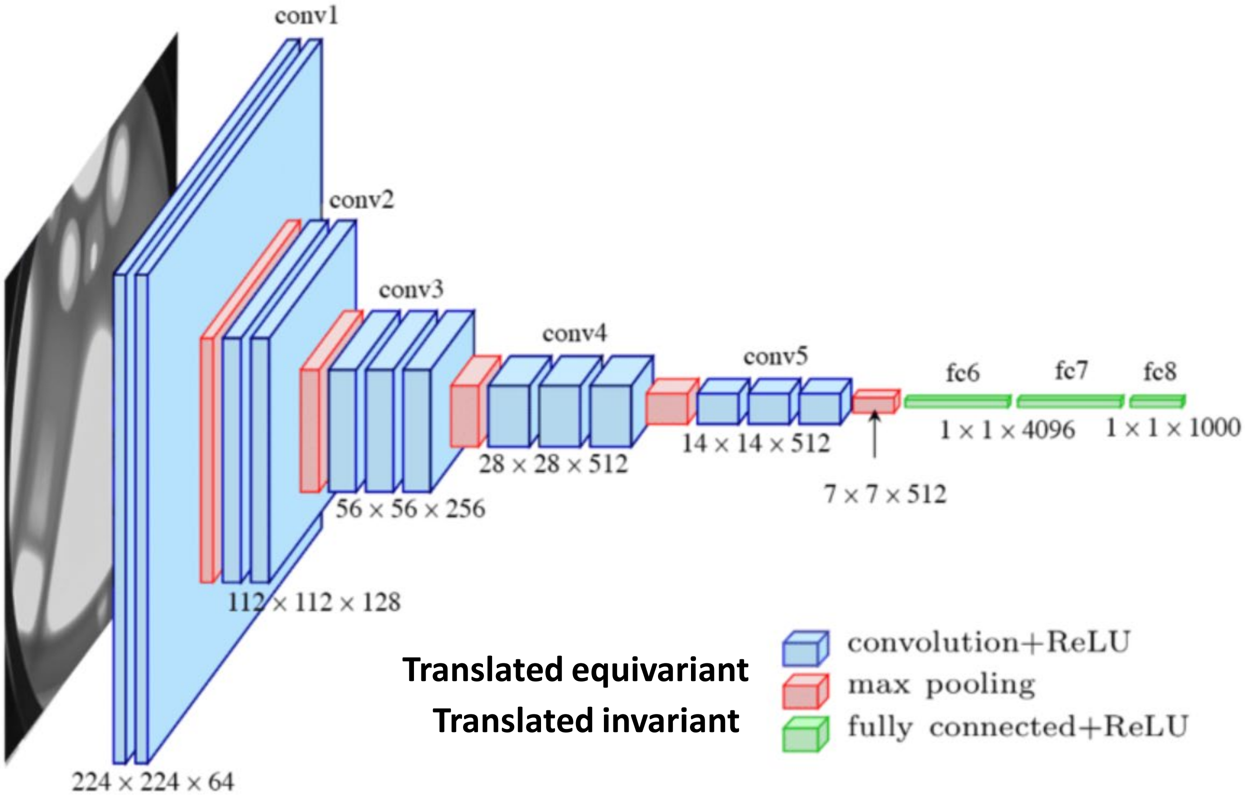
Property 3: Subsampling the pixels will not change the object. (Subsampling invariance)



We can subsample the pixels to make image smaller

→ Less parameters for the network to process the image

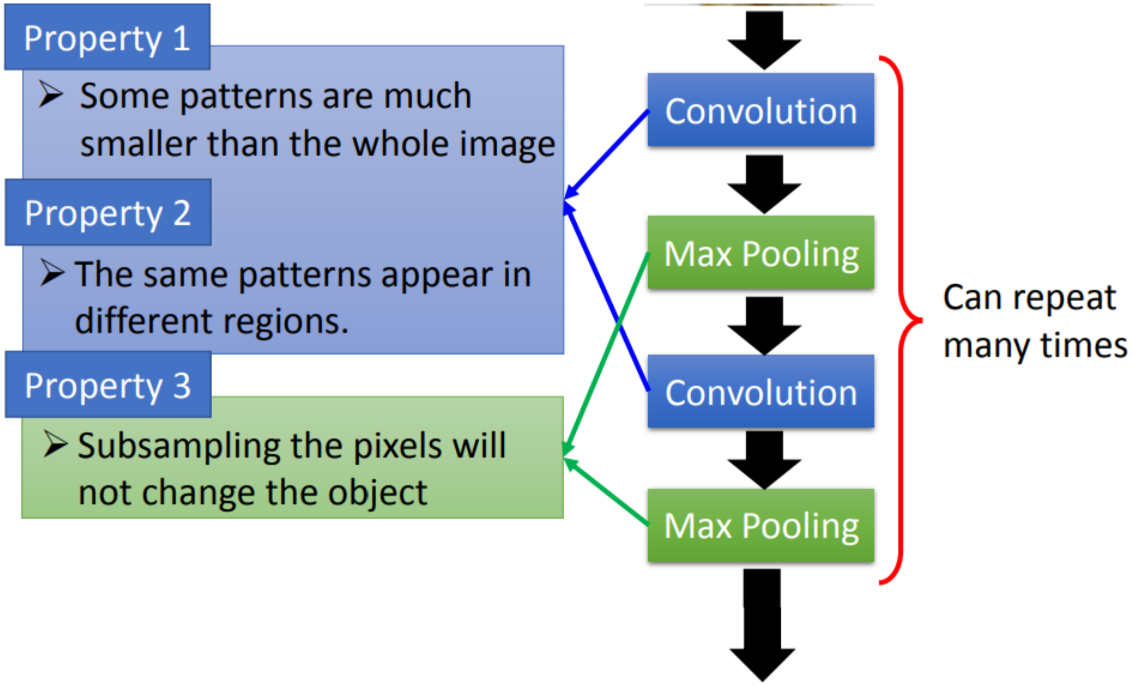
How can CNN make this happen?



Translated equivariant

Translated invariant

How can CNN make this happen?



CNN— Convolution layer

Stride=1

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
0	0	0	0	1	0
1	1	1	0	1	0
0	0	0	0	1	0

6×6 image

Those are the network parameters to be learned.

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

Matrix

3	-1	-3	-1
-2	2	-1	-3
-2	-4	0	1
-1	0	-2	-1

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

Matrix

Property 1 Each filter detects a small pattern (3 x 3).

CNN— Convolution layer

Stride=1

0	1	0	0	0	1
0	1	0	0	1	0
0	0	0	1	0	0
0	0	0	1	0	0
1	1	1	0	1	0
0	0	0	0	0	1

6×6 image

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

Matrix

0	1	-2	-1
-1	1	-1	-3
-1	-4	0	-1
-1	-1	-3	3

Property 2

The same patterns appear in different regions can be detected.

CNN— Max Pooling

Property 3

Subsampling the pixels will not change the object

6×6 image

convolution

3	-1	-3	-1
-2	2	-1	-3
-2	-4	0	1
-1	0	-2	-1

0	1	-2	-1
-1	1	-1	-3
-1	-4	0	-1
-1	-1	-3	3

Max pooling

3	-1
0	1

New images but smaller

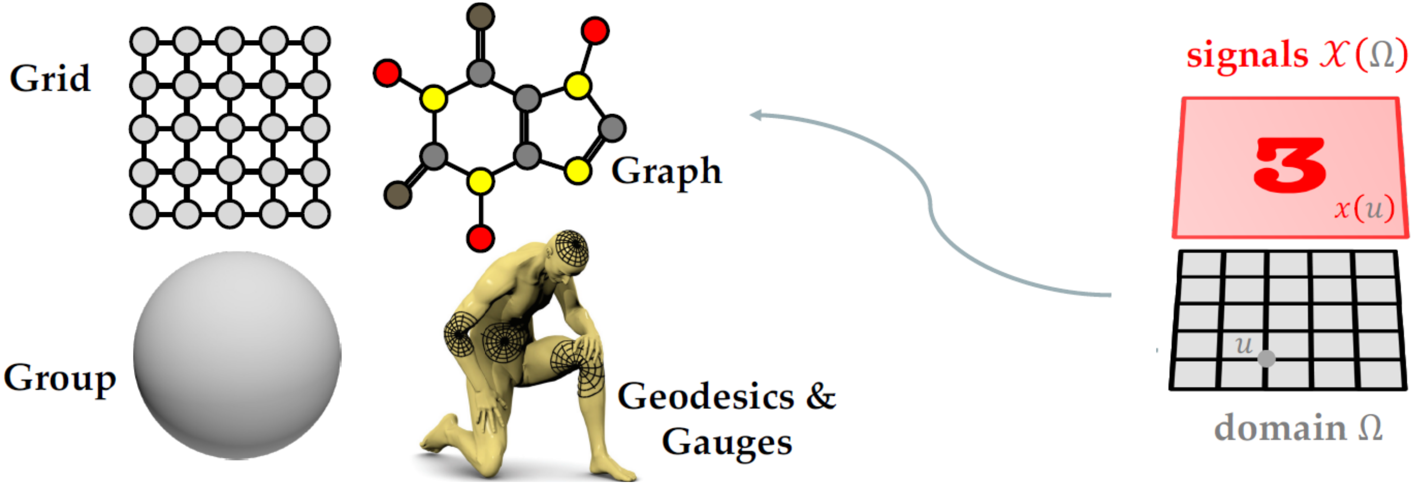
1	-1
-1	3

Not much image-like data in power system operation and planning…

Geometric deep learning (GDL)

In GDL, data are **signals** x on **geometric domains** Ω

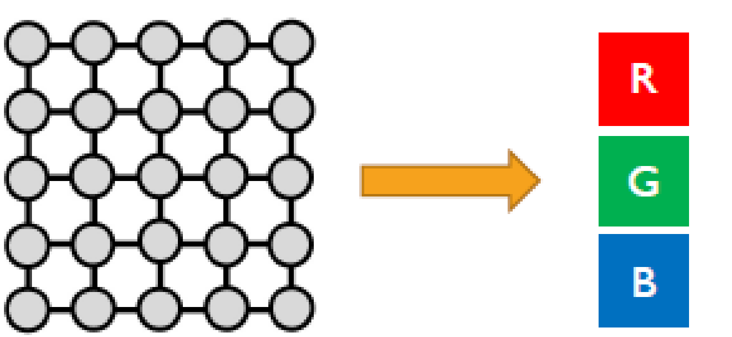
- The domain Ω is a set, possibly with additional structure
- A signal x on Ω is a function $\chi(\Omega, C) = \{x: \Omega \rightarrow C\}$
- C is a vector space whose dimensions are called channels



TU Delft

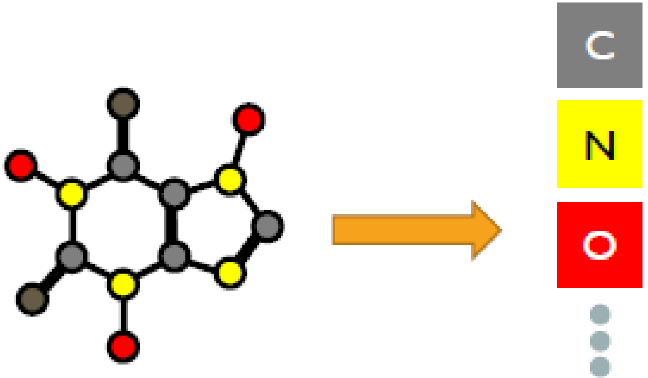
[25] Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*.

Graph Neural Networks (GNNs)



$\Omega = \mathbb{Z}_n \times \mathbb{Z}_n$ $\mathcal{C} = \mathbb{R}^3$

Example: $n \times n$ RGB image

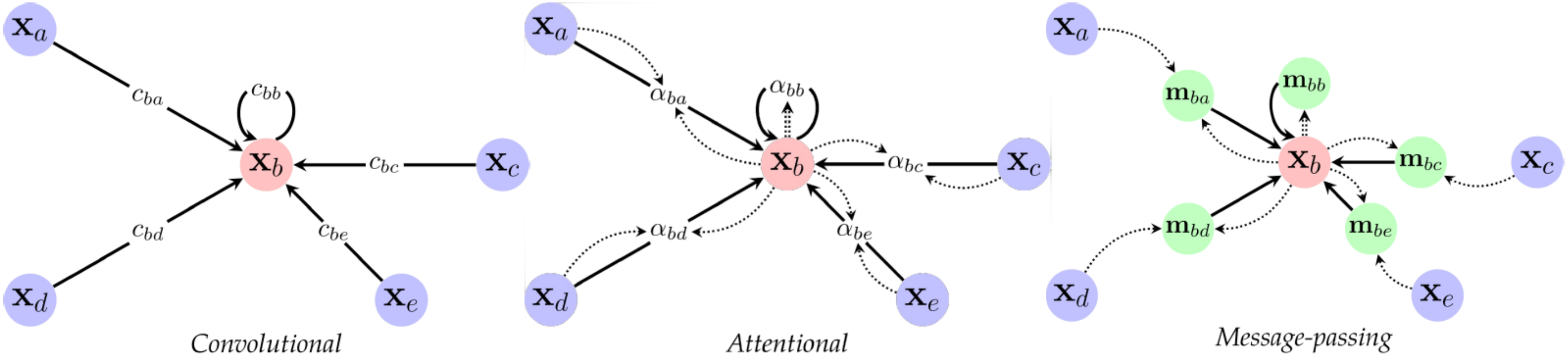


$\Omega = \{1, \ldots, n\}$ $\mathcal{C} = \mathbb{R}^m$

Example: molecular graph

[25] Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*.

The three ‘flavours’ of GNN layers



$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} c_{ij}\psi(\mathbf{x}_j)\right)$$

$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} a(\mathbf{x}_i, \mathbf{x}_j)\psi(\mathbf{x}_j)\right)$$

$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} \psi(\mathbf{x}_i, \mathbf{x}_j)\right)$$

Increasing order of generality: $convolutional \subseteq attentional \subseteq message - passing$

Outline

Reliability management and data in control rooms

TU Delft

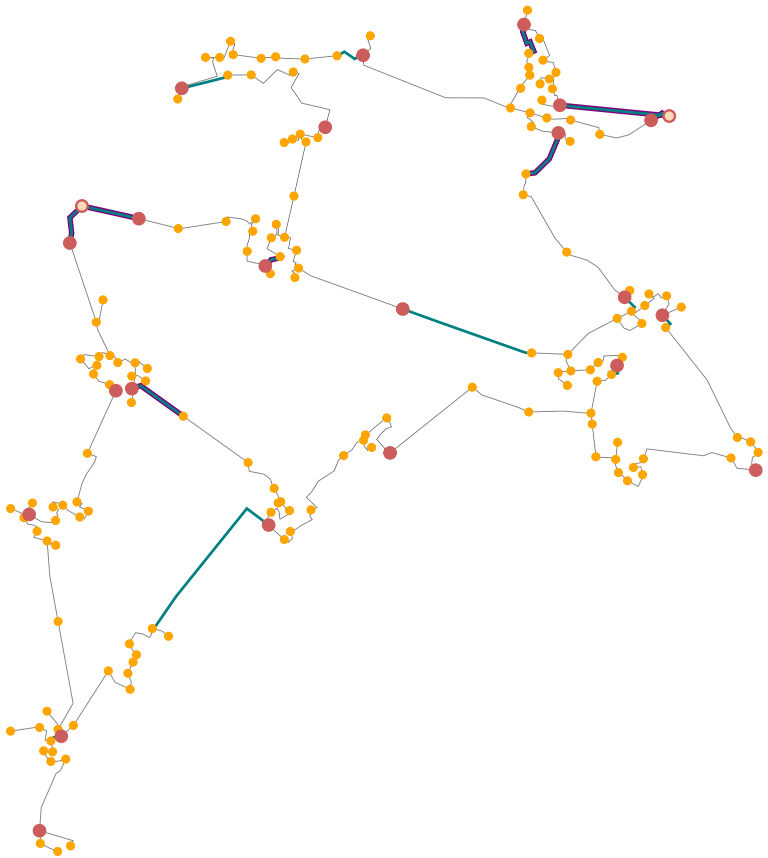
Distribution system state estimation

- Measurements $z \in \mathbb{R}^m$ with noise $\varepsilon \in \mathbb{R}^m$
- System state $x \in \mathbb{R}^{2n}$
- State estimation $f(z) \rightarrow x$
- Challenge: partial observable, scarce measurements $m \ll n$

TU Delft

[26] A. Primadianto and C. -N. Lu, "A Review on Distribution System State Estimation," in *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875-3883, Sept. 2017, doi: 10.1109/TPWRS.2016.2632156.

Distribution system (MV)



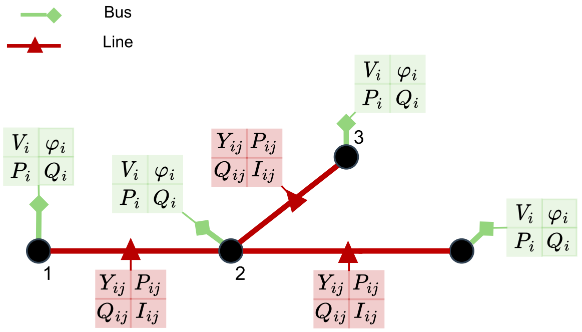
- Trafo
- Lines
- MV/LV buses
- HV buses
- Power flow measurement
- Voltage measurements

Model of the power flow

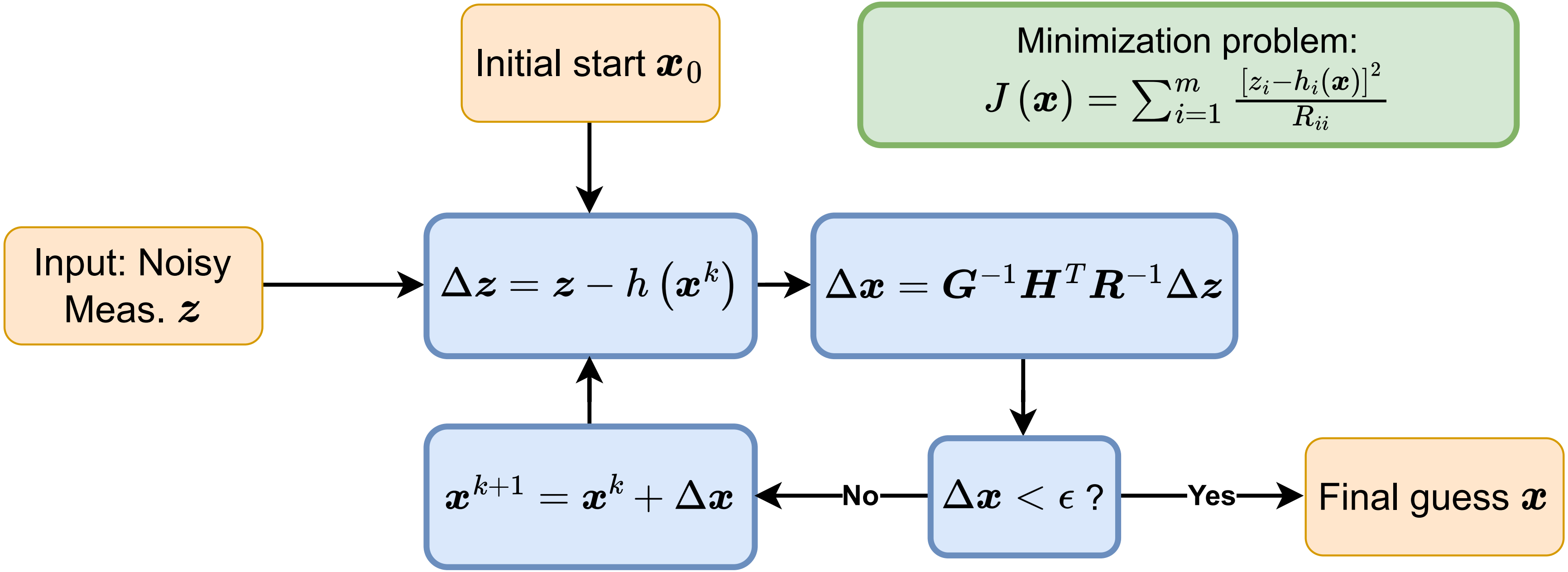
$$x = [V, \varphi]$$

$$z = h(x) + \varepsilon$$

$$h(x) = \begin{cases} V_i = V_i \\ \varphi_i = \varphi_i \\ P_{ij\rightarrow} = -V_i V_j [\mathbb{R}(Y_{ij}) \cos \Delta\varphi_{ij} + \mathbb{I}(Y_{ij}) \sin \Delta\varphi_{ij}] + V_i^2 \left[\mathbb{R}(Y_{ij}) + \frac{\mathbb{R}(Y_{sij})}{2}\right] \\ P_{ij\leftarrow} = V_i V_j [-\mathbb{R}(Y_{ij}) \cos \Delta\varphi_{ij} + \mathbb{I}(Y_{ij}) \sin \Delta\varphi_{ij}] + V_j^2 \left[\mathbb{R}(Y_{ij}) + \frac{\mathbb{R}(Y_{sij})}{2}\right] \\ Q_{ij\rightarrow} = V_i V_j [-\mathbb{R}(Y_{ij}) \sin \Delta\varphi_{ij} + \mathbb{I}(Y_{ij}) \cos \Delta\varphi_{ij}] - V_i^2 \left[\mathbb{I}(Y_{ij}) + \frac{\mathbb{I}(Y_{sij})}{2}\right] \\ Q_{ij\leftarrow} = V_i V_j [\mathbb{R}(Y_{ij}) \sin \Delta\varphi_{ij} + \mathbb{I}(Y_{ij}) \cos \Delta\varphi_{ij}] - V_j^2 \left[\mathbb{I}(Y_{ij}) + \frac{\mathbb{I}(Y_{sij})}{2}\right] \\ I_{ij\rightarrow} = -\left|\frac{P_{ij\rightarrow} - jQ_{ij\rightarrow}}{\sqrt{3} V_i e^{-j\varphi_i}}\right| \\ I_{ij\leftarrow} = -\left|\frac{P_{ij\leftarrow} - jQ_{ij\leftarrow}}{\sqrt{3} V_j e^{-j\varphi_j}}\right| \\ P_i = -\sum_{j \in N_x(i)} P_{ij\leftarrow} + P_{ij\rightarrow} \\ Q_i = -\sum_{j \in N_x(i)} Q_{ij\leftarrow} + Q_{ij\rightarrow} \end{cases}$$



Weighted least squares method



How accurate is x?
Convergence?

TU Delft

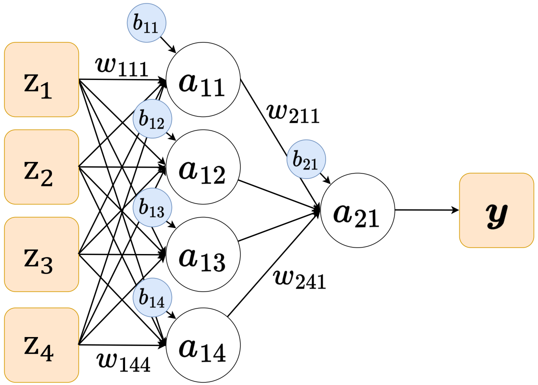
Statistical learning

- Training set $\Omega^T = \{(z_1, y_1), (z_2, y_2) \dots (z_t, y_t)\}$ with t samples

- Inference problem is to find a function $f: Z \to Y$ such that $f(z) \sim y$

- Common loss function $\mathrm{L}(f(z), y)$ for regression is the square loss

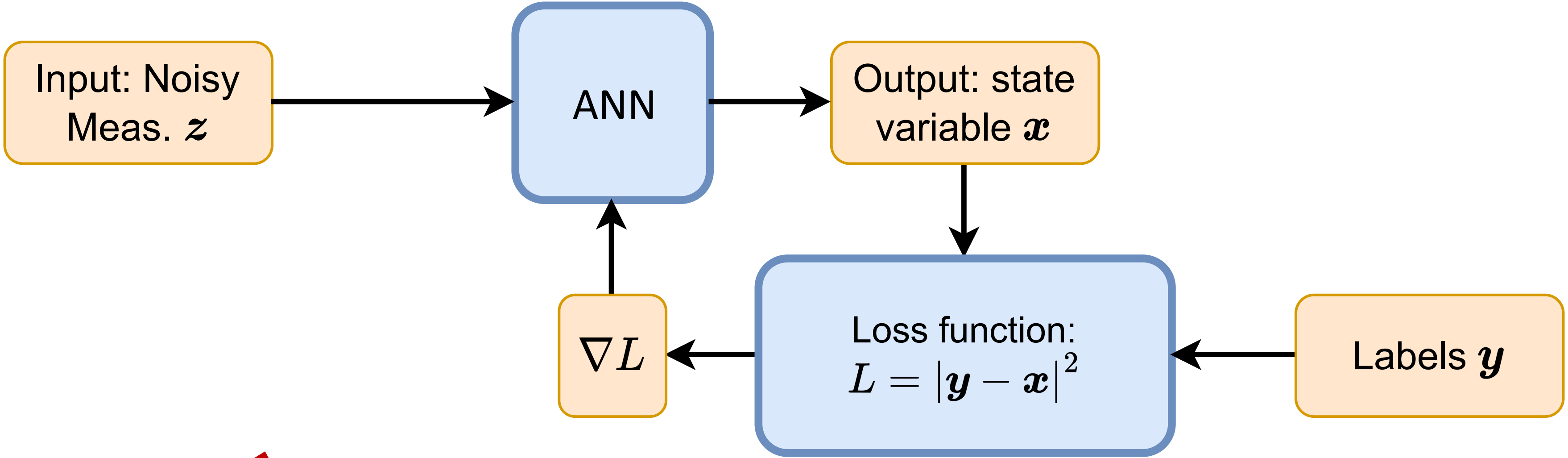
Artificial Neural Network (ANN)



$f_\theta: Z \to Y$

Note: this notation differs to conventional statistical learning where x is feature

Supervised learning for state estimation



- Labels are not known
- Newton's method generates label with "errors" $\hat{y} = y + \gamma^N$

Weakly-supervised learning

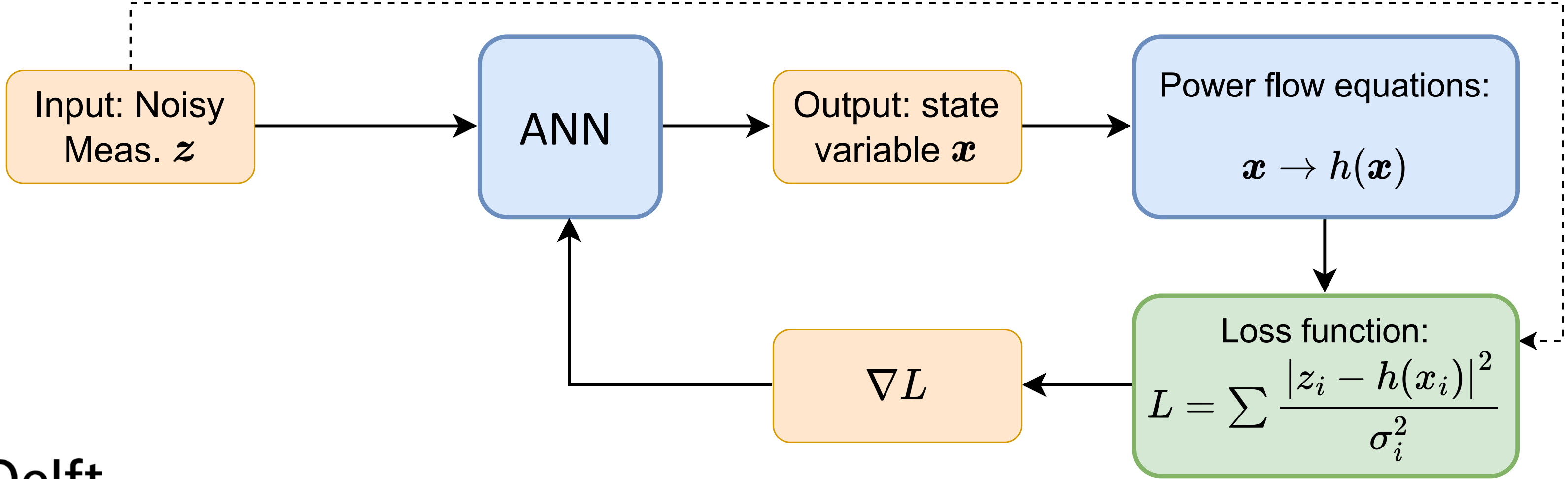
- Inaccurate input and output
- Learn with inaccurate labels $\Omega^T = \{(z_1, \hat{y}_1), (z_2, \hat{y}_2) \dots (z_t, \hat{y}_t)\}$
- Design a loss function $\mathrm{L}(f(z), \hat{y})$
- Objective: learning $f: Z \rightarrow Y$ such that $f(z) \sim y$

[27] Y. -F. Li, L. -Z. Guo and Z. -H. Zhou, "Towards Safe Weakly Supervised Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 334-346, 1 Jan. 2021, doi: 10.1109/TPAMI.2019.2922396.

Weakly-supervised learning for state estimation

Benjamin Habib

- ANN $f(z) \rightarrow x$
- Measurement function using power flow equations $\mathrm{h}(\mathrm{x}) \rightarrow \hat{z}$

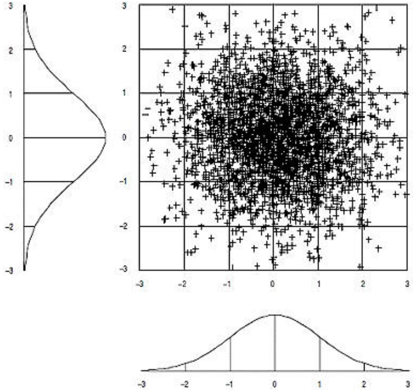


TU Delft

[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, ”Deep Statistical Solver for Distribution System State Estimation, ” *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358*.

Respect the structure of the domain

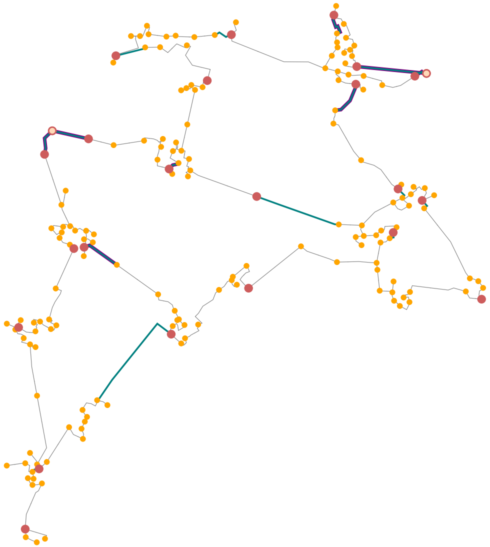
Noisy measurements



Power flow equations

$$h(x) = \begin{cases} V_i = V_i \\ \varphi_i = \varphi_i \\ P_{ij\rightarrow} = -V_i V_j \left[\mathbb{R}(Y_{ij}) \cos \Delta\varphi_{ij} + \mathbb{I}(Y_{ij}) \sin \Delta\varphi_{ij}\right]\big] + V_i^2 \left[\mathbb{R}(Y_{ij}) + \frac{\mathbb{R}(Y_{sij})}{2}\right] \\ P_{ij\leftarrow} = V_i V_j \left[-\mathbb{R}(Y_{ij}) \cos \Delta\varphi_{ij} + \mathbb{I}(Y_{ij}) \sin \Delta\varphi_{ij}\right]\big] + V_j^2 \left[\mathbb{R}(Y_{ij}) + \frac{\mathbb{R}(Y_{sij})}{2}\right] \\ Q_{ij\rightarrow} = V_i V_j \left[-\mathbb{R}(Y_{ij}) \sin \Delta\varphi_{ij} + \mathbb{I}(Y_{ij}) \cos \Delta\varphi_{ij}\right]\big] - V_i^2 \left[\mathbb{I}(Y_{ij}) + \frac{\mathbb{I}(Y_{sij})}{2}\right] \\ Q_{ij\leftarrow} = V_i V_j \left[\mathbb{R}(Y_{ij}) \sin \Delta\varphi_{ij} + \mathbb{I}(Y_{ij}) \cos \Delta\varphi_{ij}\right]\big] - V_j^2 \left[\mathbb{I}(Y_{ij}) + \frac{\mathbb{I}(Y_{sij})}{2}\right] \\ I_{ij\rightarrow} = -\left|\frac{P_{ij\rightarrow} - jQ_{ij\rightarrow}}{\sqrt{3} V_i e^{-j\varphi_i}}\right| \\ I_{ij\leftarrow} = -\left|\frac{P_{ij\leftarrow} - jQ_{ij\leftarrow}}{\sqrt{3} V_j e^{-j\varphi_j}}\right| \\ P_i = -\sum_{j \in N_x(i)} P_{ij\leftarrow} + P_{ij\rightarrow} \\ Q_i = -\sum_{j \in N_x(i)} Q_{ij\leftarrow} + Q_{ij\rightarrow} \end{cases}$$

Topology



[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, "Deep Statistical Solver for Distribution System State Estimation," *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358*.

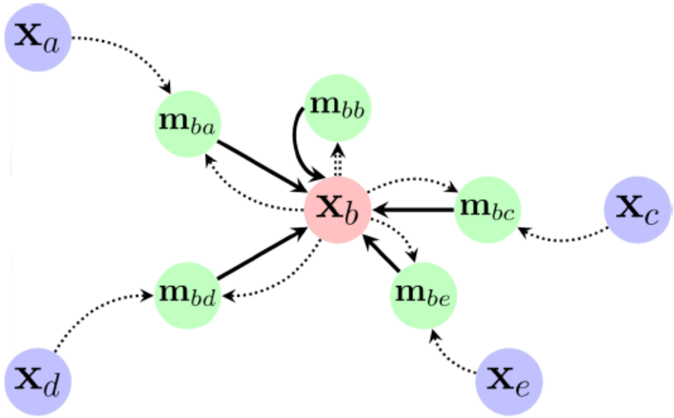
Locality on graphs: Neighbourhoods

- Consider graph $G = (V, E)$ where $E \subseteq V \times V$
- Adjacency matrix A with

$$a_{ij} = \begin{cases} 1, & (i,j) \in E \\ 0, & (i,j) \notin E \end{cases}$$

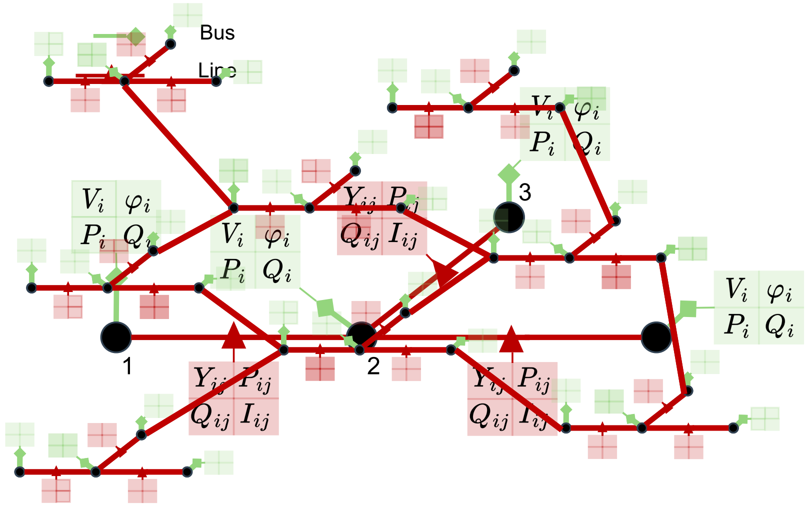
- (1-hop) neighbourhood $N_i = \{j: (i,j) \in E \cup (j,i) \in E\}$ for a node i
- Neighbourhood features $\boldsymbol{X}_{N_i} = \{\{x_j : j \in N_i\}\}$
- Local function, $\phi(x_i, \boldsymbol{X}_{N_i})$, operating over them.

Convolutional layers & message passing



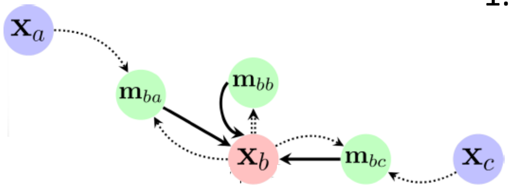
$$\eta_i = \phi\left(x_i, \bigoplus_{j \in N_i} \psi(x_i, x_j)\right)$$

State estimation



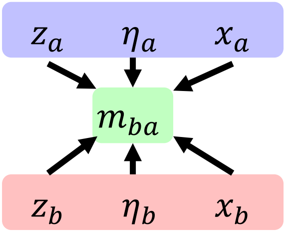
[28] Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., ... & Pascanu, R. (2018). Relational inductive biases, deep learning, and graph networks. *arXiv e-prints*, arXiv-1806.

Deep statistical solver



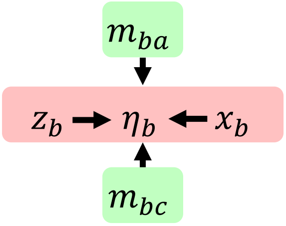
0. Initialize $x = x^0, \eta = \eta^0$

1. update edges



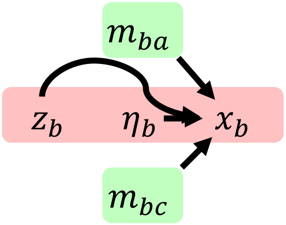
$$m_{ba} \leftarrow m_{ba} + \Delta t \times [\phi_\theta^{ba}(t, z_a, \eta_a, x_a) + \phi_\theta^{ba}(t, z_b, \eta_b, x_b)]$$

2. update vertices



$$\eta_b \leftarrow \eta_b + \Delta t \times \phi_\theta^b(t, z_b, \eta_b, x_b, m_{ba}, m_{bc}, m_{bb})$$

3. update label



$$x_b \leftarrow x_b + \Delta t \times \phi_\theta^{bx}(t, z_b, \eta_b, x_b, m_{ba}, m_{bc}, m_{bb})$$

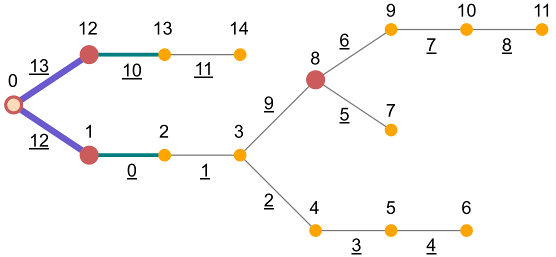
Perform $t = 1, \ldots T$

[14] Balthazar Donon, Liu, Z., Liu, W., Guyon, I., Marot, A., & Schoenauer, M. (2020). Deep statistical solvers. *Advances in Neural Information Processing Systems*, *33*, 7910-7921.

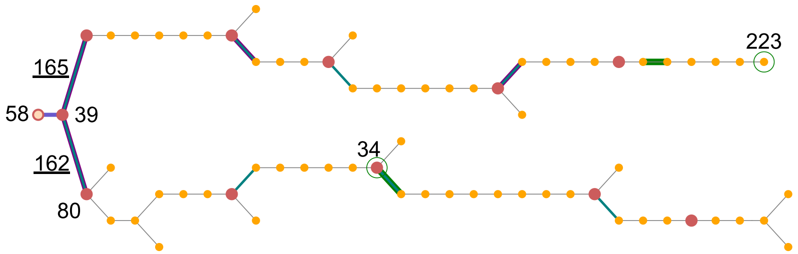
Benjamin Habib

Case study: power systems

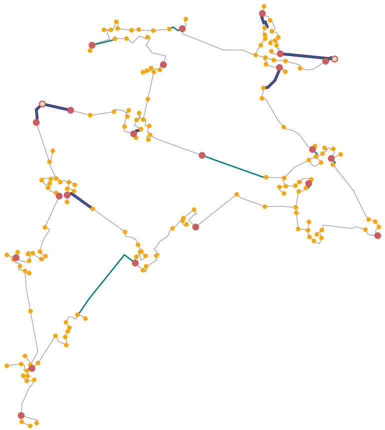
14-bus CIGRE MV grid from



70-bus Oberrhein MV sub-grid from



179-bus Oberrhein MV grid from



- Trafo
- Lines
- MV/LV buses
- HV buses
- Power flow measurement
- Voltage measurements
- Focus bus

[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and J. L. Cremer, "Deep Statistical Solver for Distribution System State Estimation," in IEEE Transactions on Power Systems, doi: 10.1109/TPWRS.2023.3290358.

Case study settings

Data generation

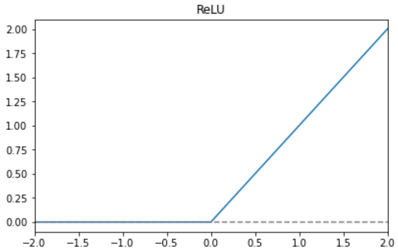
- 8640 days, with each 24 hours, +/- 15% around Gaussian on loads
- Balanced system, pandapower, AC power flow
- Measurement noise
  - $0.5\% - 2\%$ for the voltage and current measurement
  - $1\% - 5\%$ for the active and reactive power measurement
  - Pseudomeasurement were generic load profiles
- Baselines
  - Weighted least square (WLS)
  - Feedforward Neural Network (FFNN)
  - supervised DSS^2

Model & hyperparameters

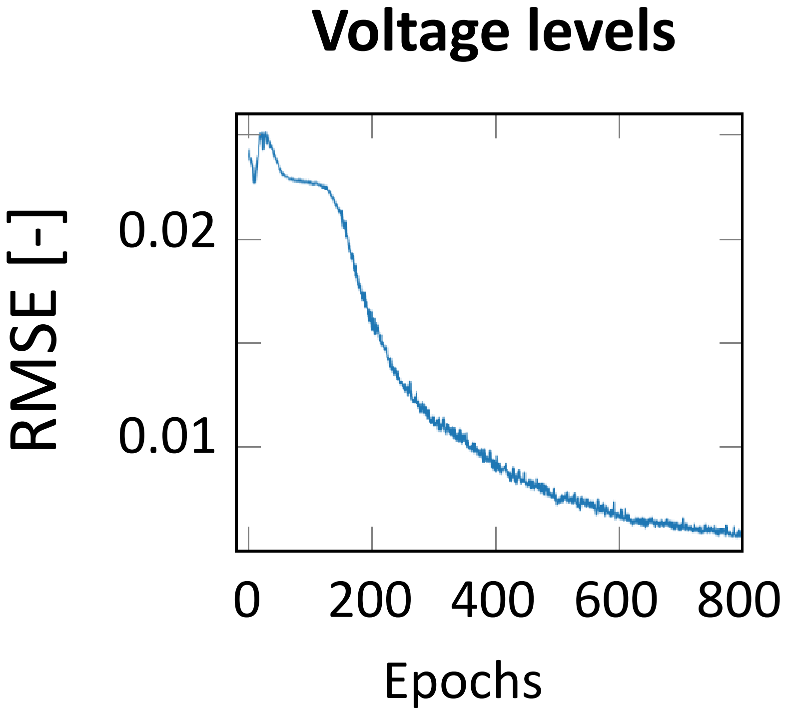
- Hyper-Heterogeneous Multi GNN
- Training 80%, validation 10%, testing 10%
- Grid search on learning rate λ, layer dimensions d, and layer numbers

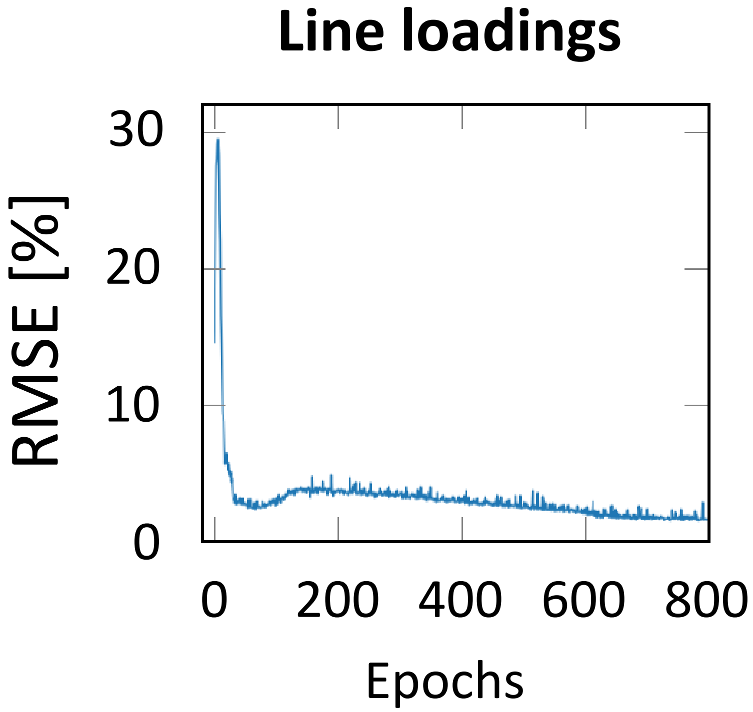
Assume stable system

$$L(\boldsymbol{z}, \boldsymbol{x}) = \sum_{i \in m} \frac{|z_i - h_i(\boldsymbol{x})|^2}{R_{ii}} + \lambda[\text{ReLU}(V - 1.05) + \text{ReLU}(0.95 - V)$$
$$+ \text{ReLU}(\text{ loading } - 100) + \text{ReLU}(\varphi - 0.25) + \text{ReLU}(-0.25 - \varphi)]$$



Training performance 14-bus system

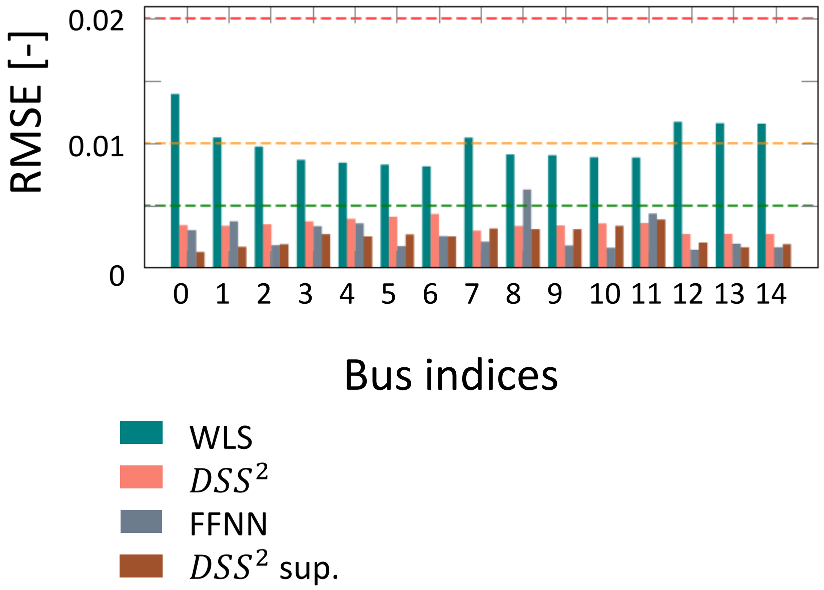




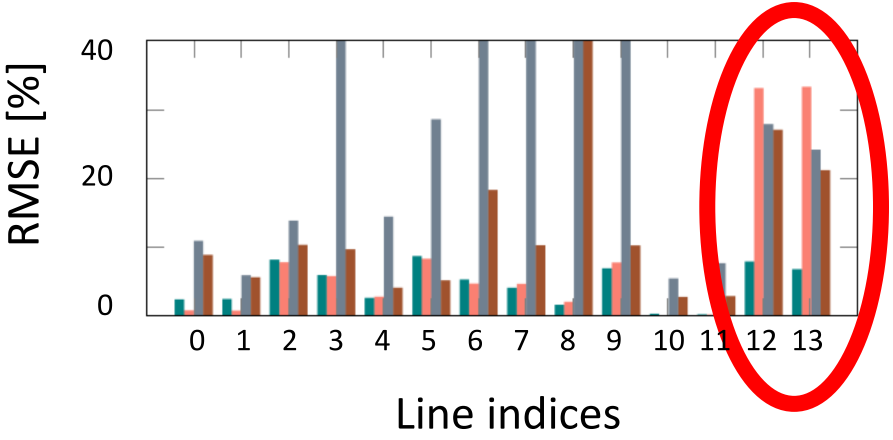
[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, ”Deep Statistical Solver for Distribution System State Estimation, ” *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358*.

State estimation 14-bus system

Voltage levels



Line loadings



Model inaccuracies: assumed transformer = lines!

TU Delft

[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, "Deep Statistical Solver for Distribution System State Estimation, " *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358.*

Accuracy

	14-bus system			
Metric \ approach	WLS	ANN	sup DSS^2	DSS^2
Voltage RMSE $[10^{-3}]$	10	3	**3**	3
Line loading RMSE [%]	**3**	42	13	4
Trafos loading RMSE [%]	**5**	39	14	8

* with increased convergence rate
** with higher tolerance and more iterations

[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, "Deep Statistical Solver for Distribution System State Estimation," *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358.*

Convergence

	14-bus system				70-bus Oberrhein			179-bus Oberrhein	
Metric \ approach	WLS	ANN	sup DSS^2	DSS^2	WLS	WLS*	DSS^2	WLS**	DSS^2
Voltage RMSE $[10^{-3}]$	10	3	3	3	31	6	2	10	2
Line loading RMSE [%]	3	42	13	4	17	15	2	6	3
Trafos loading RMSE [%]	5	39	14	8	39	24	3	4	4
Convergence [%]	100	100	100	100	25	100	100	53	100

- WLS did not converge in some instances (25%-50%)
- DSS^2 always 'converges' (produces a label)

* with increased convergence rate
** with higher tolerance and more iterations

[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, "Deep Statistical Solver for Distribution System State Estimation, " *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358.*

Computational ‘prediction’ time [ms]

	14-bus system				70-bus Oberrhein			179-bus Oberrhein	
Metric \ approach	WLS	ANN	sup DSS^2	DSS^2	WLS	WLS*	DSS^2	WLS**	DSS^2
Voltage RMSE $[10^{-3}]$	10	3	3	3	31	6	2	10	2
Line loading RMSE [%]	3	42	13	4	17	15	2	6	3
Trafos loading RMSE [%]	5	39	14	8	39	24	3	4	4
Convergence [%]	100	100	100	100	25	100	100	53	100
Computational time $[ms]$	86	**4**	5	6	123	161	**26**	1212	**58**

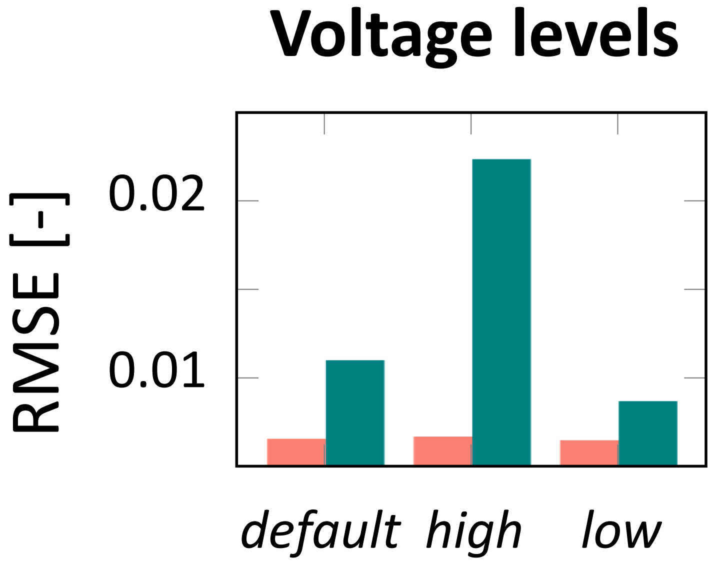
~10 ~2

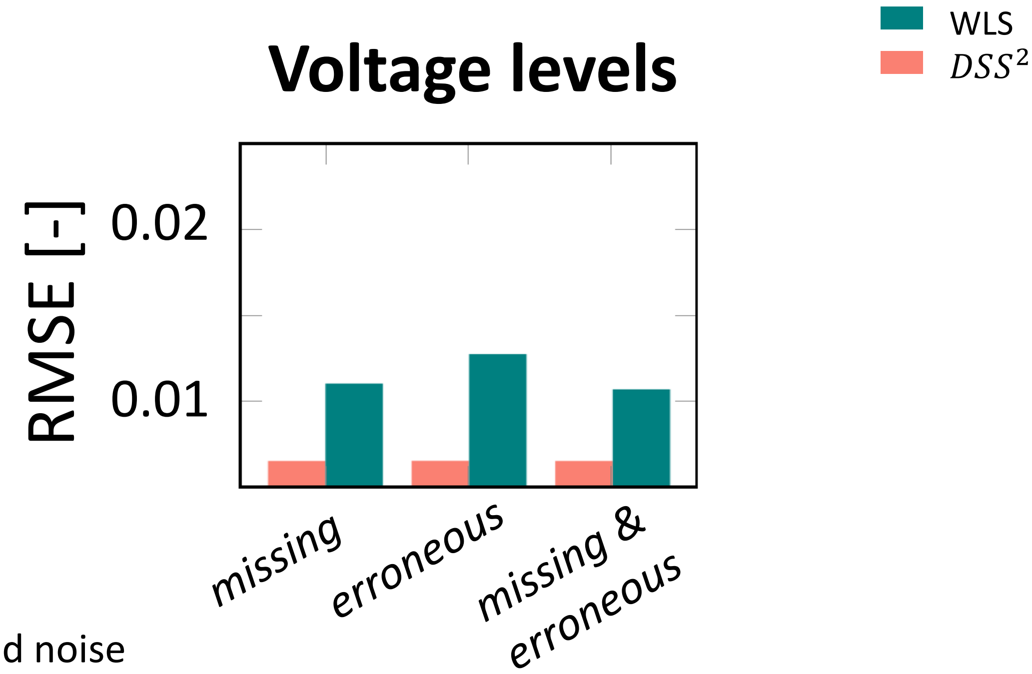
- WLS increases significantly with system size
- DSS^2 increases moderately with system size

* with increased convergence rate
** with higher tolerance and more iterations

[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, ”Deep Statistical Solver for Distribution System State Estimation, ” *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358.*

Noise and missing, erroneous data

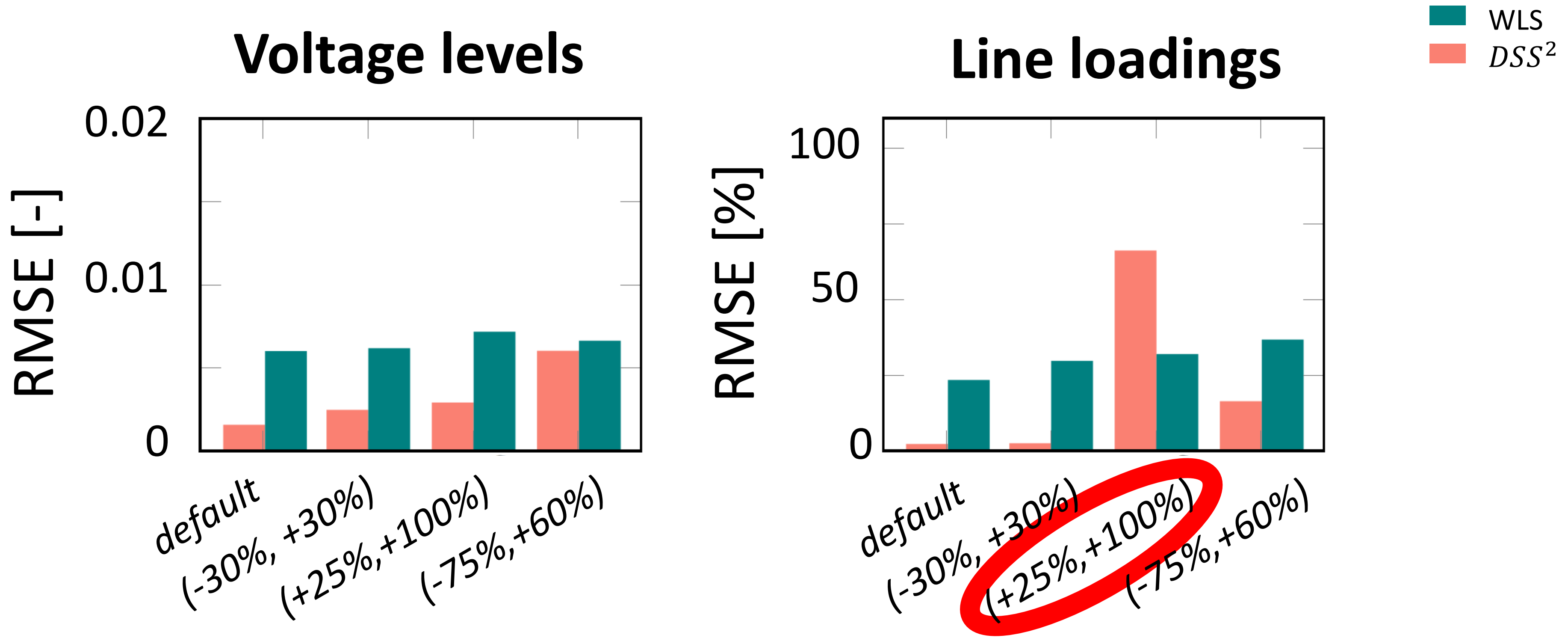




- DSS^2 successfully cancelled noise
- DSS^2 was not trained to handle such events
- GNN architecture increased the interpolation capabilities by incorporating the data symmetries w.r.t. the underlying graph

[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, "Deep Statistical Solver for Distribution System State Estimation, " *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358.*

Increase in (generation, load)



Limitation: high load levels

[22] B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, "Deep Statistical Solver for Distribution System State Estimation," *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358.*

Outline

Reliability management and data in control rooms

1. Introduction to reliability management
2. Machine learning approaches
3. Security assessment with cost-sensitive supervised learning

Learning models for secure system operation

4. Learning with domain knowledge
5. State estimation with graph neural networks
6. Weakly-supervised learning for secure operation
7. Challenges applying ML to reliability

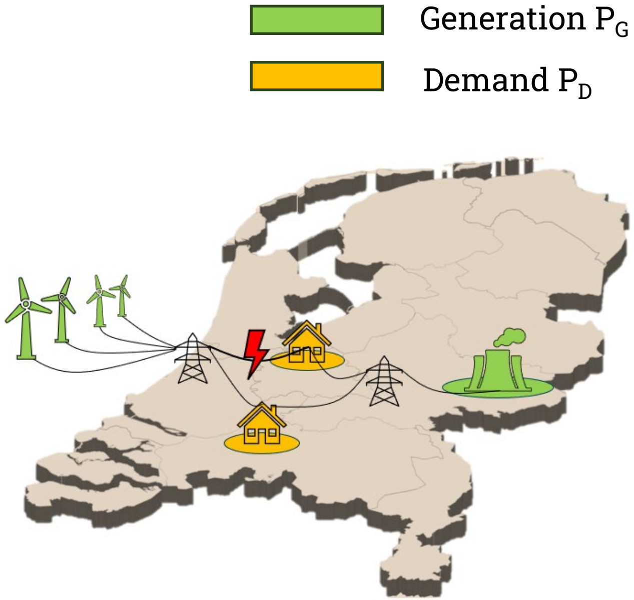
TU Delft

Bastien Giraud

Problem overview

- Growing grid complexity
  - ➢ Challenging to maintain N-1 security
- Increasing number of unforeseen weather events
- Need for N-k considerations to increase reliability

- **Problem:** Conventional approaches don’t scale well with the number of simultaneous outages k



TU Delft

Security constrained optimal power flow (SCOPF)

Objective:	minimize cost	$\min_{n \in \Omega^G} \Sigma\, c_n P_{G_n}$
Constraints:	In = out	$\boldsymbol{B} \cdot \boldsymbol{\delta} = \boldsymbol{P_G} - \boldsymbol{P_D}$
	Generator limits	$P_{G_n}^{min} < P_{G_n} < P_{G_n}^{max}\ \forall n \in \Omega^G$
	Line flow limits	$F_l^{min} < F_l < F_l^{max}\ \forall l \in \Omega^L$
Contingency Constraints:	Line flow limits	$F_l^{min} < F_l^c < F_l^{max}\ \forall l \in \Omega^L, \forall c \in \Omega^C$

Combinatorial complexity

TU Delft

Conventional approaches

Outaged line

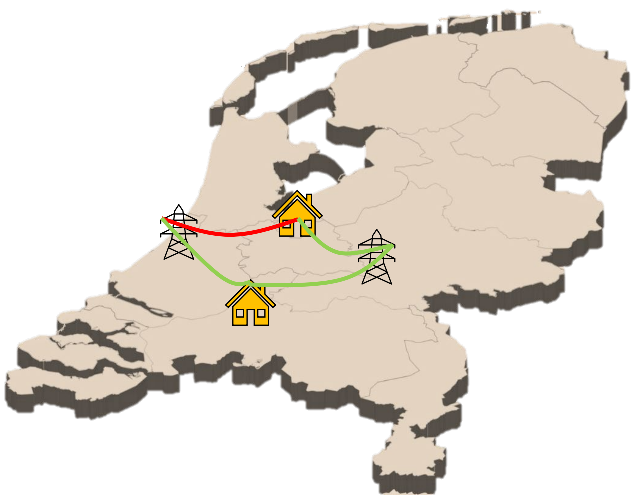
Line flow changes

Solving a **large optimization** problem can be slow

- Benders decomposition
- Column and constraint generation algorithm with robust optimization
- Line outage distribution factors (**LODF**)

Machine learning approaches often rely on **labeled** training data

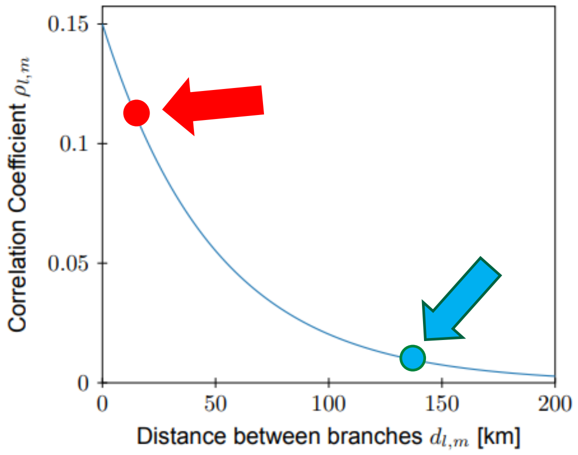
- Intractable for increasing k

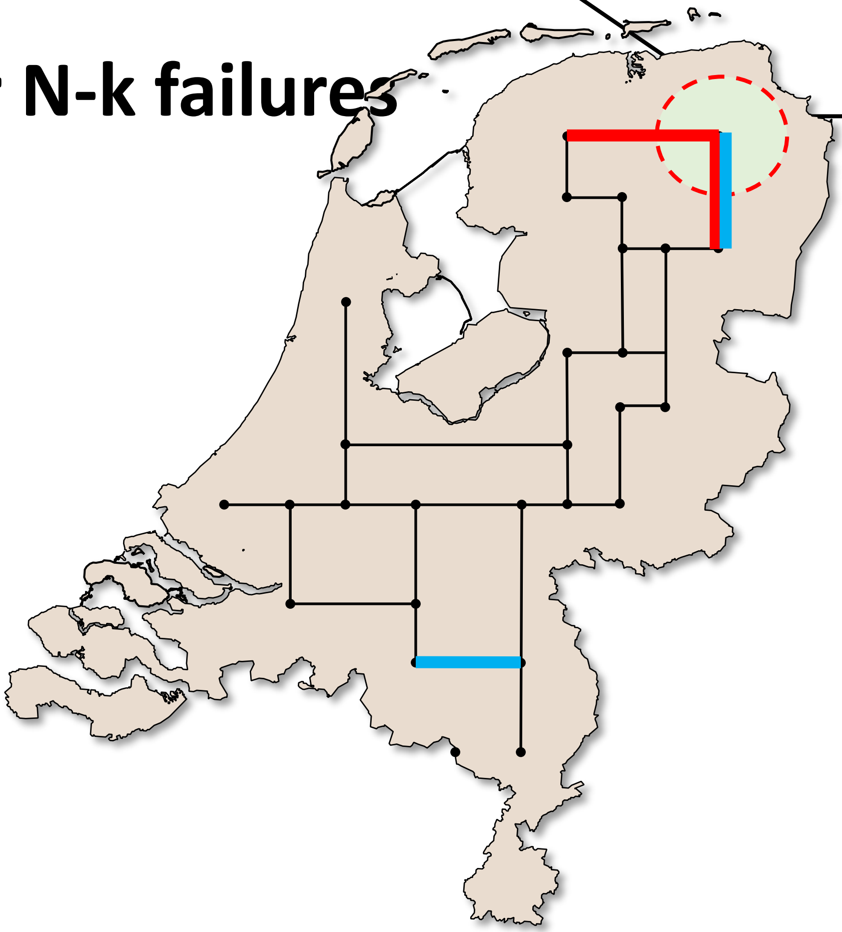


$$F^c = F^0 + LODF_{N-k} \times F^0$$

Probabilistic security for N-k failures

- Compute probabilities of all contingencies
- Spatial correlation between line outages
- Compute joint probabilities using a copula analysis





TU Delft

Proposed constraint-driven approach

Main advantages

- Weakly-supervised -> so **no labeled data** needed
- Never actually solve an SCOPF

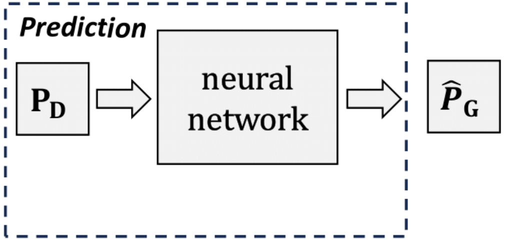
Contributions

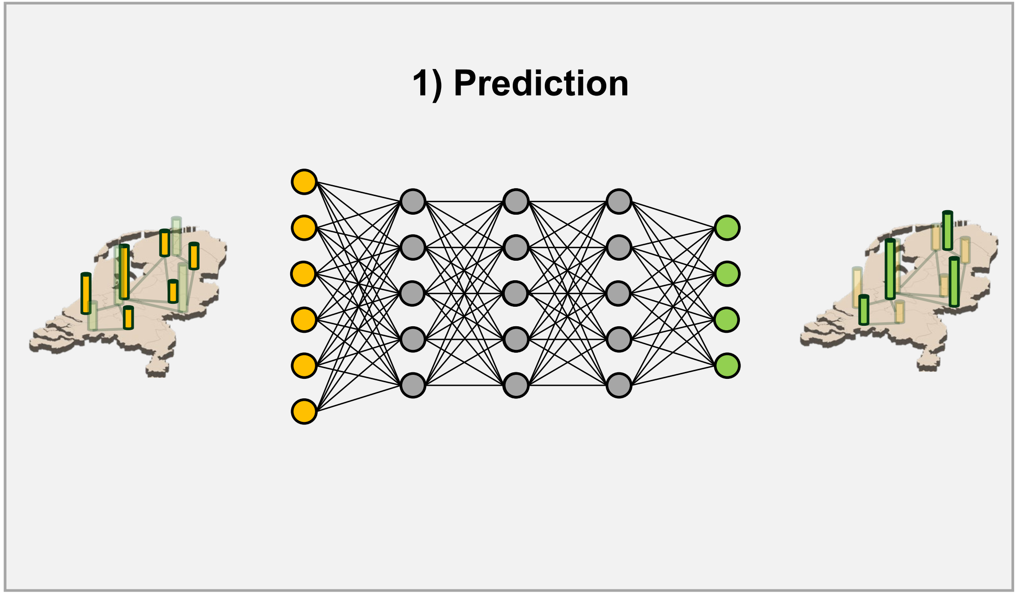
- The deterministic constraint-driven approach to approximate N-k SCOPFs, considering all line contingencies using LODFs.
- The computational graph memory reduction for fast and efficient implementation.
- The probabilistic security assessment to formulate a N-k risk-based security criterion, providing an alternative to the current deterministic N-1 security criterion.

LODF = line outage distribution factor
SCOPF = security constrainted optimal power flow

[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

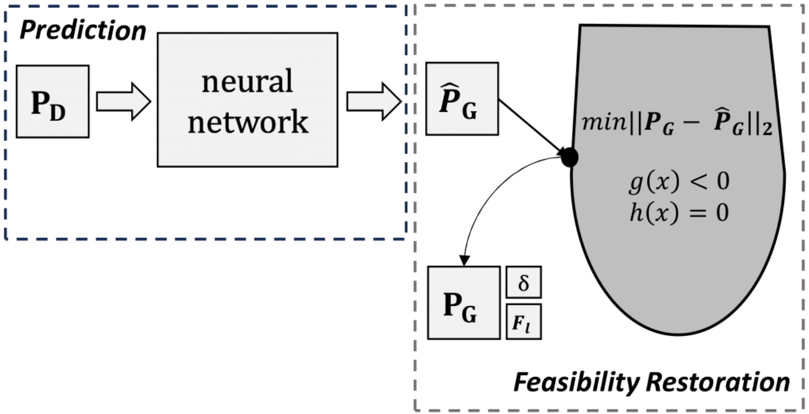
Proposed constraint-driven approach





[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Proposed constraint-driven approach



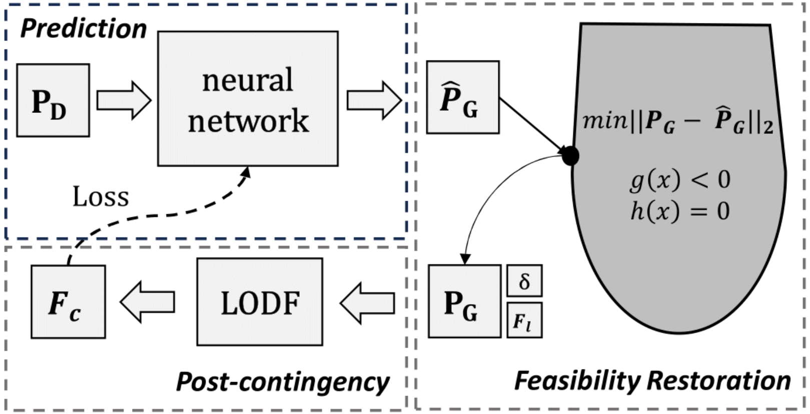
2) Feasibility Restoration

- With $\hat{P}_{G_n}$ compute predicted line flow $\hat{F}_l^0$
- Prediction might **violate** DC PF equations
- Map prediction to feasible region constrained by DC PF equations

DC PF = DC power flow

[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Proposed constraint-driven approach



3) Post-contingency

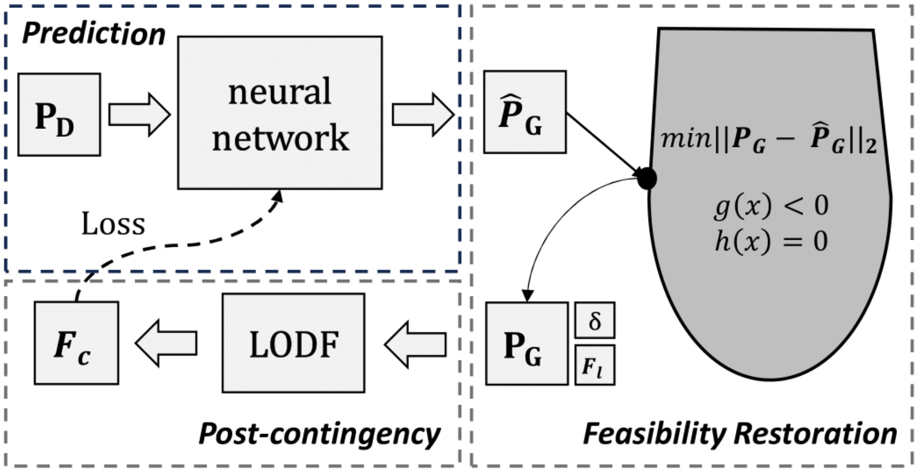
$$\boldsymbol{F^c = F^0 + LODF_{N-k} \times F^0}$$

$$F_l^{min} < F_l^c < F_l^{max}\ \forall l \in \Omega^L, \forall c \in \Omega^C$$

LODF = line outage distribution factor

[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Proposed constraint-driven approach



1) Dispatch cost — $\lambda_c \sum \boldsymbol{P_G c_G}$

2) Line flow violation pre-contingency — $\lambda_0 \left\| ReLU(|\boldsymbol{\hat{F}^0}| - \boldsymbol{F^{max}}) \right\|_1$

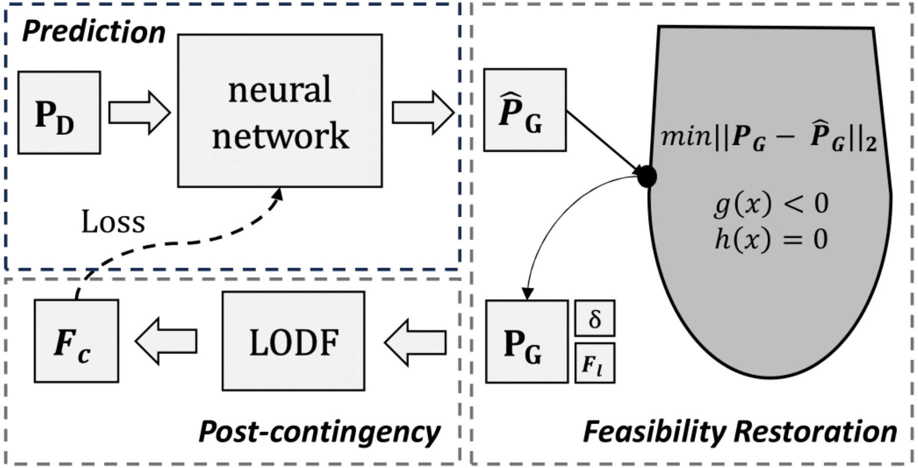
3) Line flow violation post-contingency — $\lambda_1 \| ReLU(|\boldsymbol{F^c}| - \boldsymbol{F^{max}}) \|_1$

4) Power imbalance — $\lambda_2 \left\| \sum \boldsymbol{\hat{P}_G} - \sum \boldsymbol{P_D} \right\|_1$

$$Loss = \lambda_c \sum \boldsymbol{P_G c_G} + \lambda_0 \left\| ReLU(|\boldsymbol{\hat{F}^0}| - \boldsymbol{F^{max}}) \right\|_1 + \lambda_1 \| ReLU(|\boldsymbol{F^c}| - \boldsymbol{F^{max}}) \|_1 + \lambda_2 \left\| \sum \boldsymbol{\hat{P}_G} - \sum \boldsymbol{P_D} \right\|_1$$

[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Proposed constraint-driven approach



1) Dispatch cost $\lambda_c \sum P_G c_G$

2) Line flow violation pre-contingency $\lambda_0 \left\| ReLU(|\hat{F}^0| - F^{max}) \right\|_1$

3) Line flow violation post-contingency $\lambda_1 \| \pi_{N-k} \cdot ReLU(|F^c| - F^{max}) \|_1$

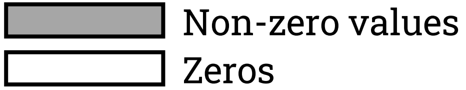
4) Power imbalance $\lambda_2 \left\| \sum \hat{P}_G - \sum P_D \right\|_1$

$$Loss = \lambda_c \sum P_G c_G + \lambda_0 \left\| ReLU(|\hat{F}^0| - F^{max}) \right\|_1 + \lambda_1 \| \pi_{N-k} \cdot ReLU(|F^c| - F^{max}) \|_1 + \lambda_2 \left\| \sum \hat{P}_G - \sum P_D \right\|_1$$

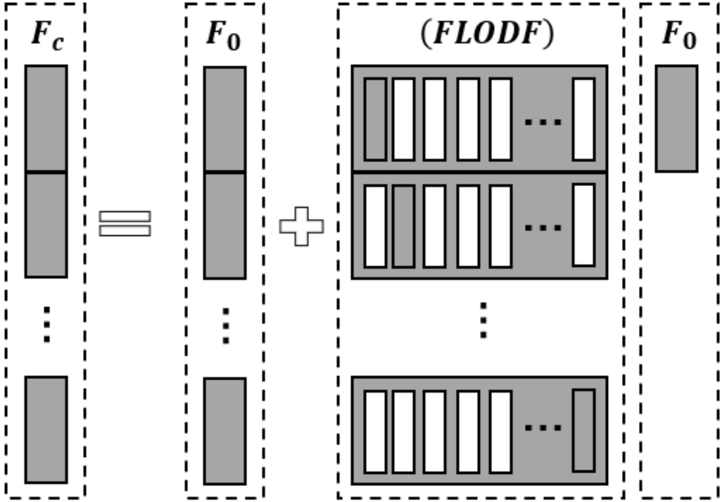
π_{N-k} = matrix of contingency probabilities

[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Sparsity LODF matrix



$$F^c = F^0 + LODF_{N-k} \times F^0$$



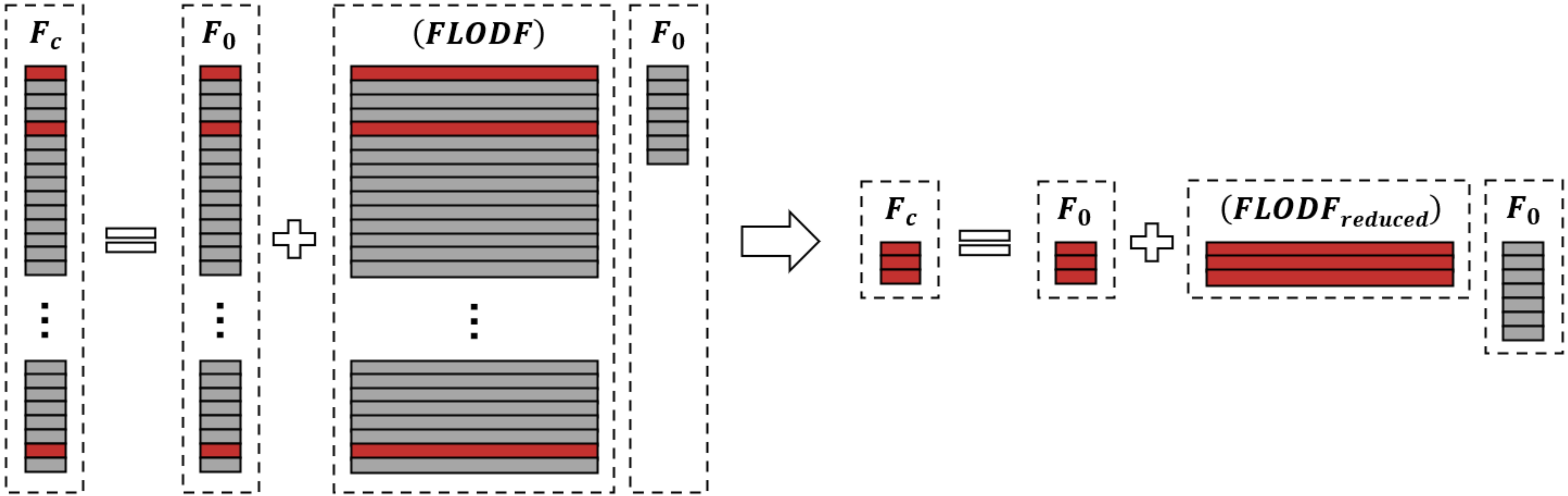
FLODF = 'Full LODF'

k	1	2	3
39-bus sparsity [%]	98.5	97.6	97.5
118-bus sparsity [%]	99.6	99.3	99.0

LODF = line outage distribution factor

[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

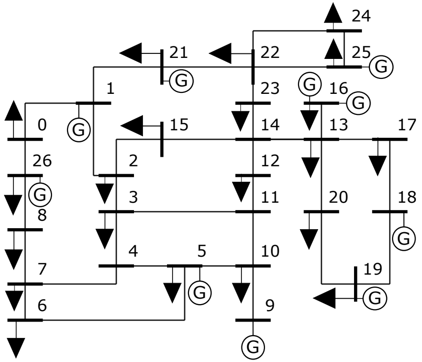
Reducing the graph

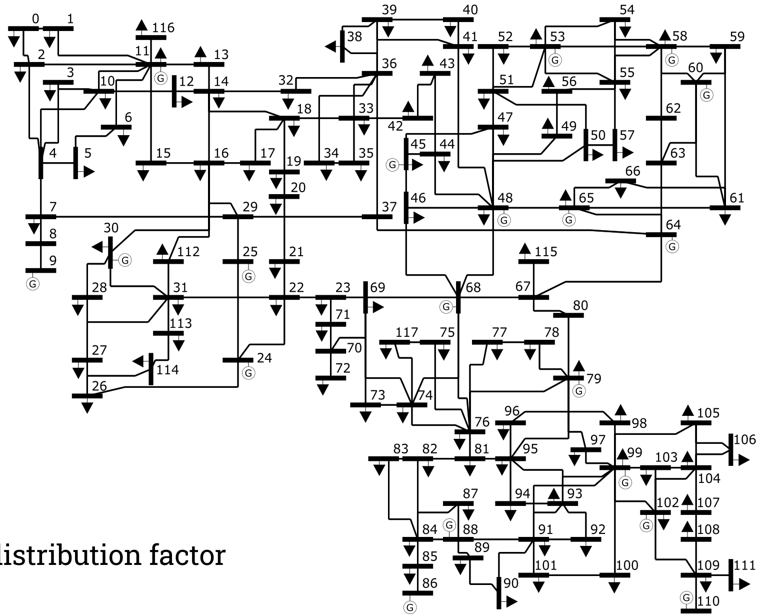


[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Case studies

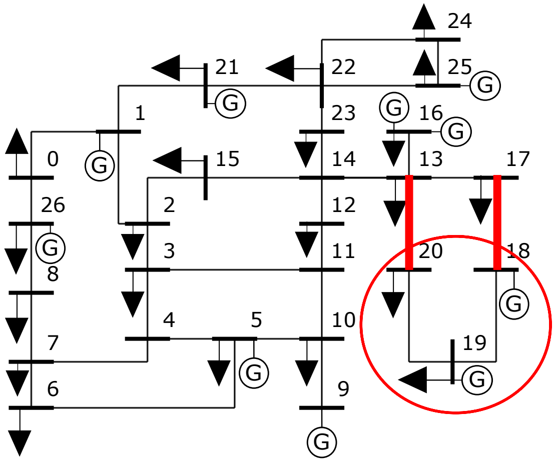
- IEEE 39-bus and 118-bus test systems
- k = {1,2,3}
- Baseline: iterative contingency screening with LODFs
- Code: https://github.com/TU-Delft-AI-Energy-Lab/Constraint-Driven-SCOPF

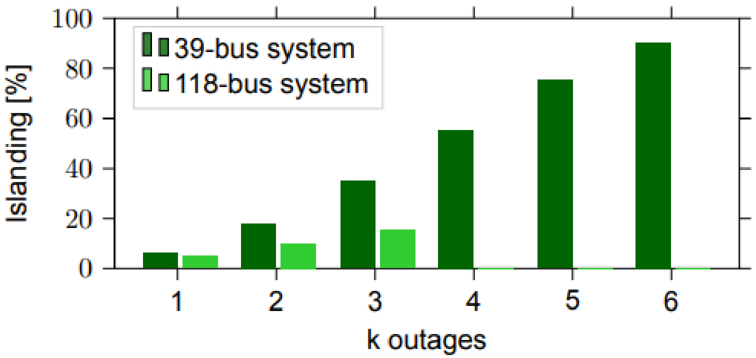




LODF = line outage distribution factor

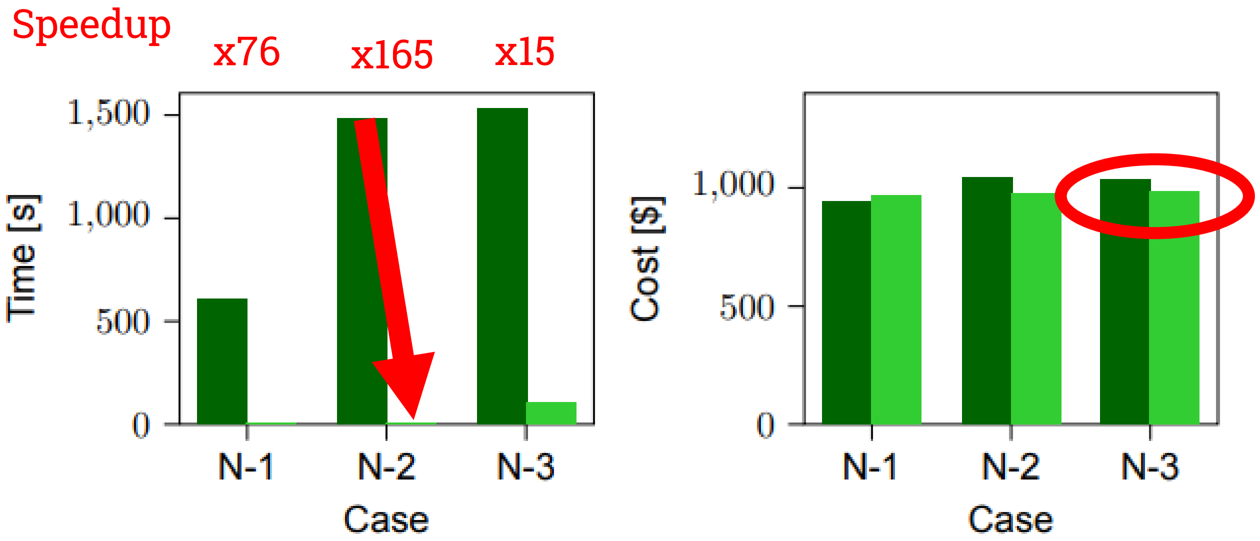
Islanding





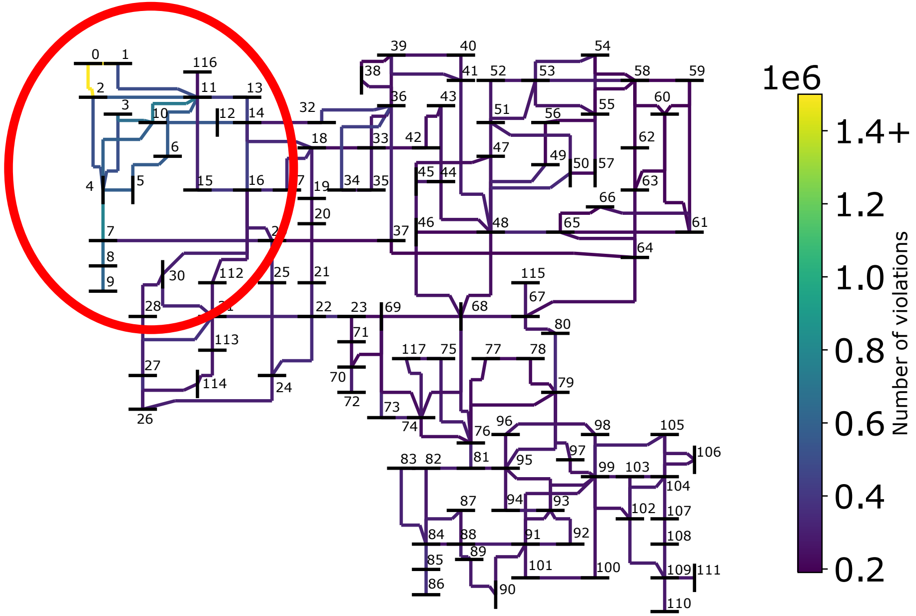
Removing islanding cases

Performance 118-bus system

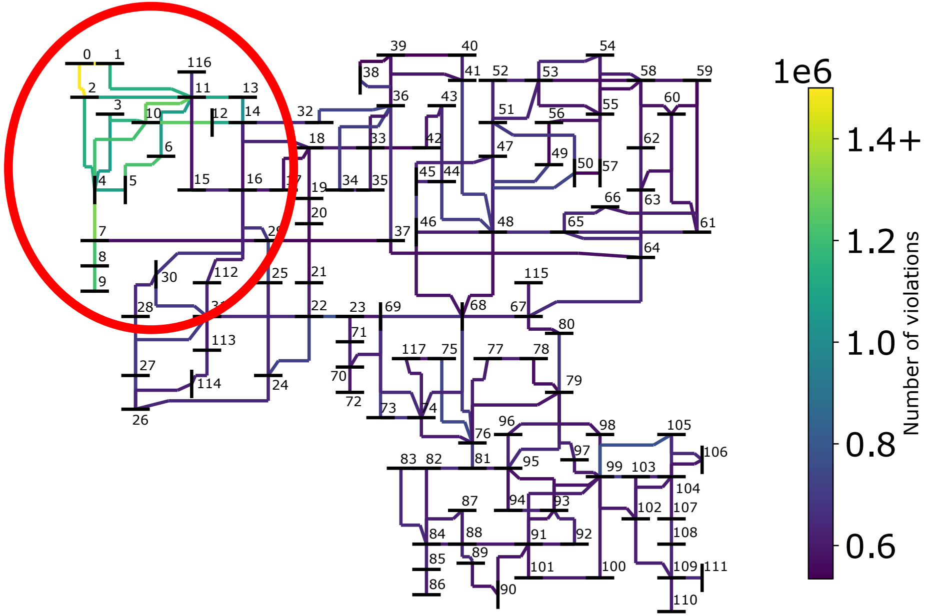


[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

N-3 proposed approach



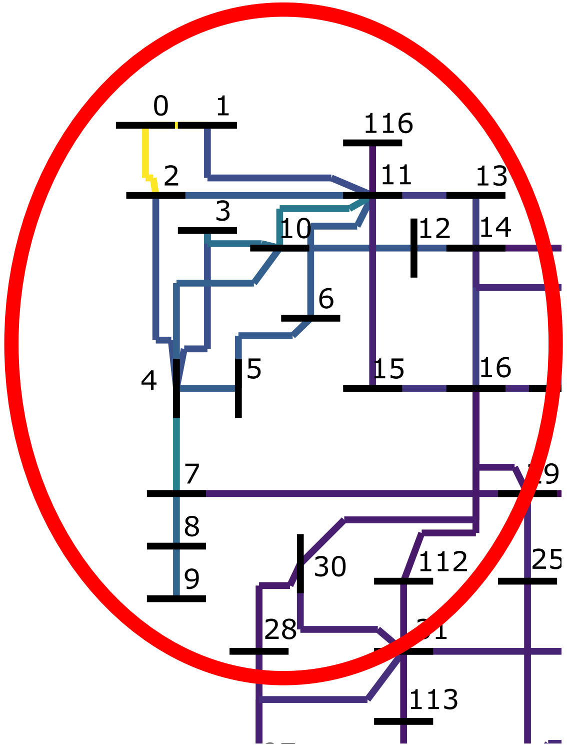
N-3 baseline

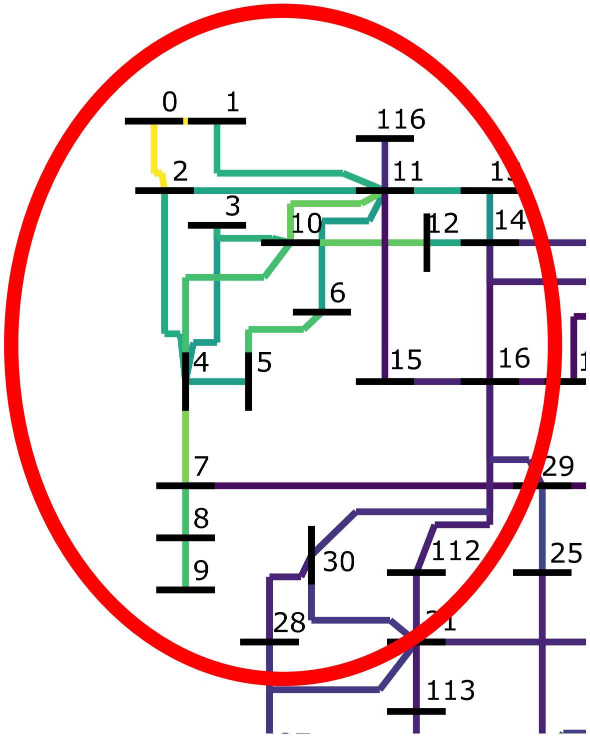


TU Delft

[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

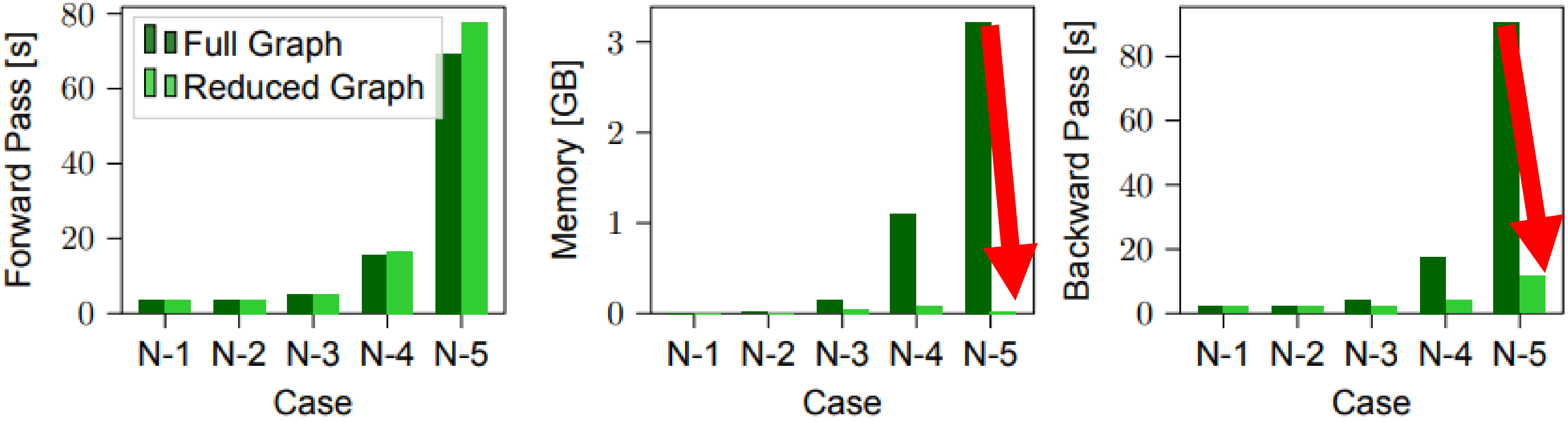
N-3 proposed approach





[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Reducing computational graph



Reduction in memory

Reduction in computation time

[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Outline

Reliability management and data in control rooms

1. Introduction to reliability management
2. Machine learning approaches
3. Security assessment with cost-sensitive supervised learning

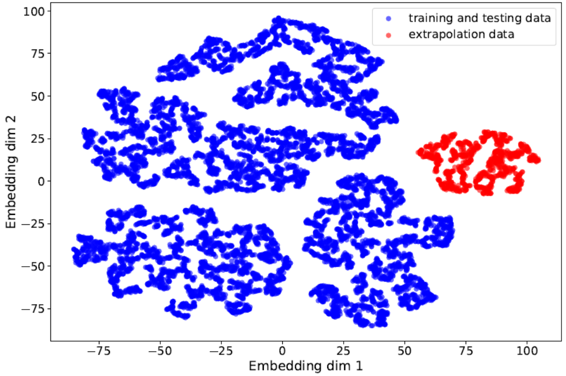
Learning models for secure system operation

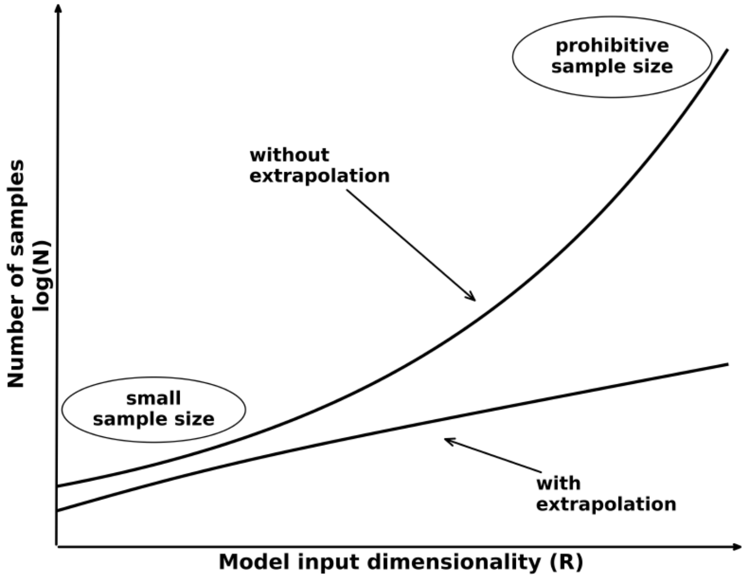
4. Learning with domain knowledge
5. State estimation with graph neural networks
6. Weakly-supervised learning for secure operation
7. Challenges applying ML to reliability

TU Delft

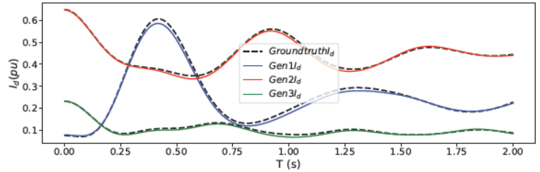
Generalisation to changes in s or m

The model performs well not just on training data, but on ***unseen scenarios*** *— new grid states, topologies, contingencies, or time horizons.*



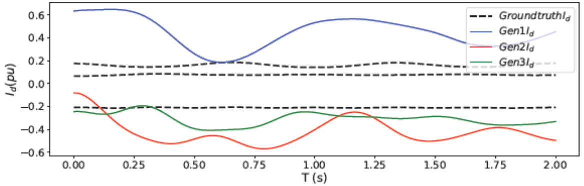


Extrapolation in continuous domain



(a) I_d current trajectory

Extrapolation in nonlinear domain (discrete)

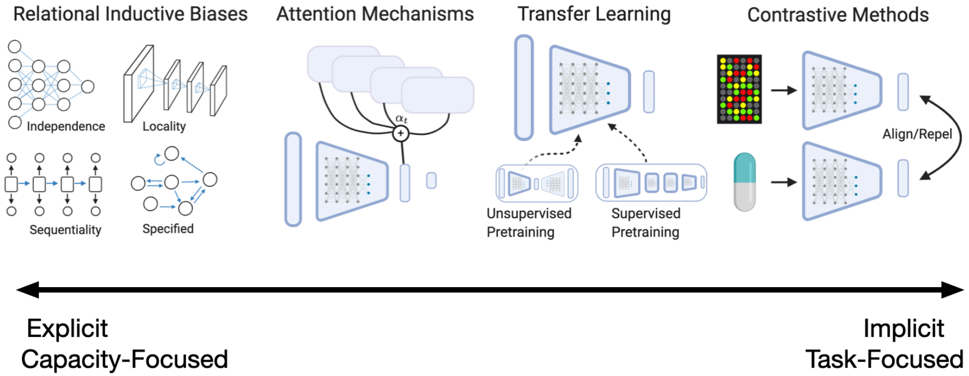


(a) I_d current trajectory

[10] Olayiwola Arowolo, Jochen Stiasny, Jochen Cremer, "Exploring the Extrapolation Performance of Machine Learning Models for Power System Time Domain Simulations", *12th Bulk Power System Dynamics and Control Symposium and Sustainable Energy, Grids and Networks Journal, 2025*

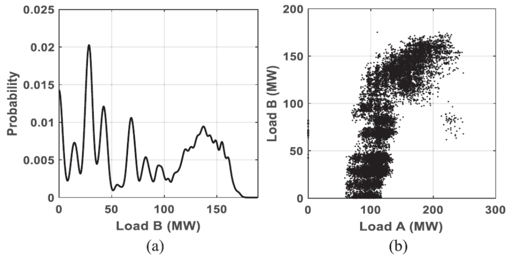
Challenge: Data-efficiency

- **Data efficiency is critical**
- Embedding **inductive bias** and learning **task-aware representations** helps supervised models generalise better — even with limited labels.

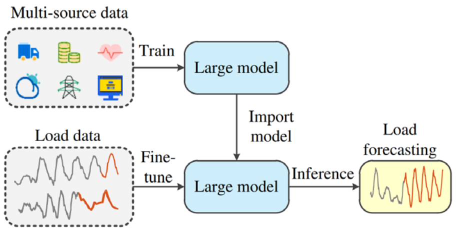


Sampling synthetic data & use real-data

Snapshot sampling



Time-series foundational models

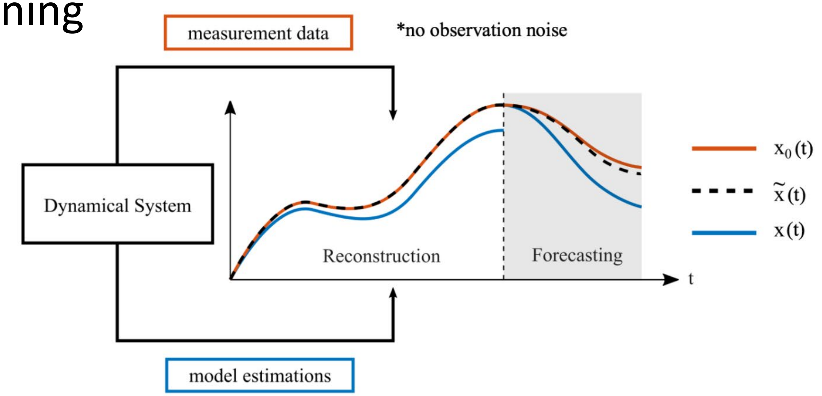


[29] https://sgfin.github.io/2020/06/22/Induction-Intro/
[30] Konstantelos, I., Sun, M., Tindemans, S. H., Issad, S., Panciatici, P., & Strbac, G. (2018). Using vine copulas to generate representative system states for machine learning. IEEE Transactions on Power Systems, 34(1), 225-235.
[31] Al-Amin Bugaje, Jochen L. Cremer, Goran Strbac, “Split-based Sequential Sampling for Realtime Security Assessment”, *International Journal of Electrical Power & Energy Systems, 2022*
[32] A. Venzke, D.K. Molzahn, S. Chatzivasileiadis,(2019). Efficient Creation of Datasets for Data-Driven Power System Applications, arXiv:1910.01794
[33] Liao, W., Wang, S., Yang, D., Yang, Z., Fang, J., Rehtanz, C., & Porté-Agel, F. (2025). TimeGPT in load forecasting: A large time series model perspective. *Applied Energy*, *379*, 124973.

Model inaccuracy $s \neq m$ (data quality issues)

"All models are wrong, but some are useful", George E. P. Box

- Example challenges
  - Distribution: Inaccurate transformer-tap positions
  - Transmission: Converter-based control models are unkown
- Possible techniques: Parameter estimation to develop probabilistic and deterministic models, discrepancy learning



[34] Ebers, M. R., Steele, K. M., & Kutz, J. N. (2024). Discrepancy modeling framework: Learning missing physics, modeling systematic residuals, and disambiguating between deterministic and random effects. *SIAM Journal on Applied Dynamical Systems*, *23*(1), 440-469.

Conclusions

- For many decades, AI has been investigated for power system reliability -> demonstrating promising ideas

- Promising: New techniques, availability of data, models, industry R&D commitments

Open research challenges

- Handling changes in data, and model inaccuracy -> Adaptive GNNs
- Curse of dimensionality -> Self-supervised learning
- Addressing risks, confidence, and trust in ML models
- A large amount of data is needed
- Integrating various concepts

Thank you

Speaker

Jochen Cremer

Associate Professor IEPG, TU Delft

Web: https://www.tudelft.nl/ai/delft-ai-energy-lab

Personal www.jochen-cremer.com

Email: j.l.cremer@tudelft.nl

Code: https://github.com/TU-Delft-AI-Energy-Lab

TU Delft

References & code

Weakly supervised learning for power systems (**Example code:** https://github.com/TU-Delft-AI-Energy-Lab/Deep-Statistical-Solver-for-Distribution-System-State-Estimation)

- Bastien Giraud, Ali Rajaei, Jochen L. Cremer "Constraint-Driven Deep Learning for N-k Security Constrained Optimal Power Flow", *Electric Power System Research and 2024 IEEE Power System Computation Conference [**Code**:* https://github.com/TU-Delft-AI-Energy-Lab/Constraint-Driven-SCOPF]

- B. Habib, E. Isufi, W. v. Breda, A. Jongepier and Jochen L. Cremer, "Deep Statistical Solver for Distribution System State Estimation, " *IEEE Transactions on Power Systems, 2023, doi: 10.1109/TPWRS.2023.3290358*.

Cost-sensitive learning

- Dariush Wahdany, Carlo Schmitt, Jochen L. Cremer, "More than Accuracy: End-To-End Wind Power Forecasting that Optimises the Energy System", *Electric Power System Research, 2023*
- A. Bugaje, J. L. Cremer, M. Sun, G. Strbac, "Selecting DT Models for Security Assessment using ROC- and Cost-Curves", *Energy and AI*, 2021: 100110.
- J. L. Cremer, G. Strbac, "A Machine-learning based Probabilistic Perspective on Dynamic Security Assessment" *International Journal of Electrical Power & Energy System, 2020*

Interpretable models (**Example code:** https://github.com/JochenC/From-optimization-based-machine-learning-to-interpretable-security-rules-for-operation)

- J. L. Cremer, I. Konstantelos, G. Strbac, "From Optimization-based Machine Learning to Interpretable Security Rules for Operation", *IEEE Transactions on Power Systems, 2019*
- J. L. Cremer, I. Konstantelos, S. H. Tindemans, G. Strbac, "Data-driven Power System Operation: Exploring the Balance between Cost and Risk", *IEEE Transactions on Power Systems, 2018*

Fast training of models

- Mert Karaçelebi, Jochen L. Cremer "Online Neural Dynamics Forecasting for Power System Security", *International Journal of Electrical Power & Energy Systems 2025*
- Mert Karaçelebi, Jochen L. Cremer, "Predicting Power System Frequency with Neural Ordinary Differential Equations", *12th Bulk Power System Dynamics and Control Symposium and Sustainable Energy, Grids and Networks Journal, 2025*

Generalisation challenge:

- Olayiwola Arowolo, Jochen Stiasny, Jochen Cremer, "Exploring the Extrapolation Performance of Machine Learning Models for Power System Time Domain Simulations", *12th Bulk Power System Dynamics and Control Symposium and Sustainable Energy, Grids and Networks Journal, 2025*

Thank you for your attention

Jochen Cremer
j.l.cremer@tudelft.nl

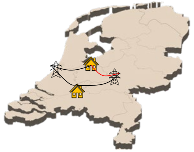
Reliability Indicators

Loss of load expectation (LOLE)

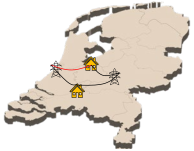
- Expected amount of time demand can not be supplied [h/y]
- ∑ (Outage duration x probability)

Expected energy not supplied (EENS)

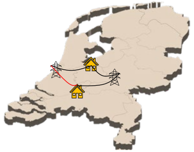
- Amount of energy expected not to be supplied during that period [MWh/y]
- ∑ (Energy not supplied x probability)



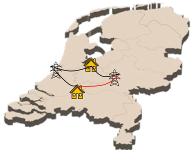
$t_1 * \pi_1 = 2$ h/y
$E_1 * \pi_1 = 3$ MWh/y



$t_1 * \pi_1 = 0$ h/y
$E_1 * \pi_1 = 0$ MWh/y



$t_1 * \pi_1 = 0$ h/y
$E_1 * \pi_1 = 0$ MWh/y



$t_1 * \pi_1 = 3$ h/y
$E_1 * \pi_1 = 4$ MWh/y

+

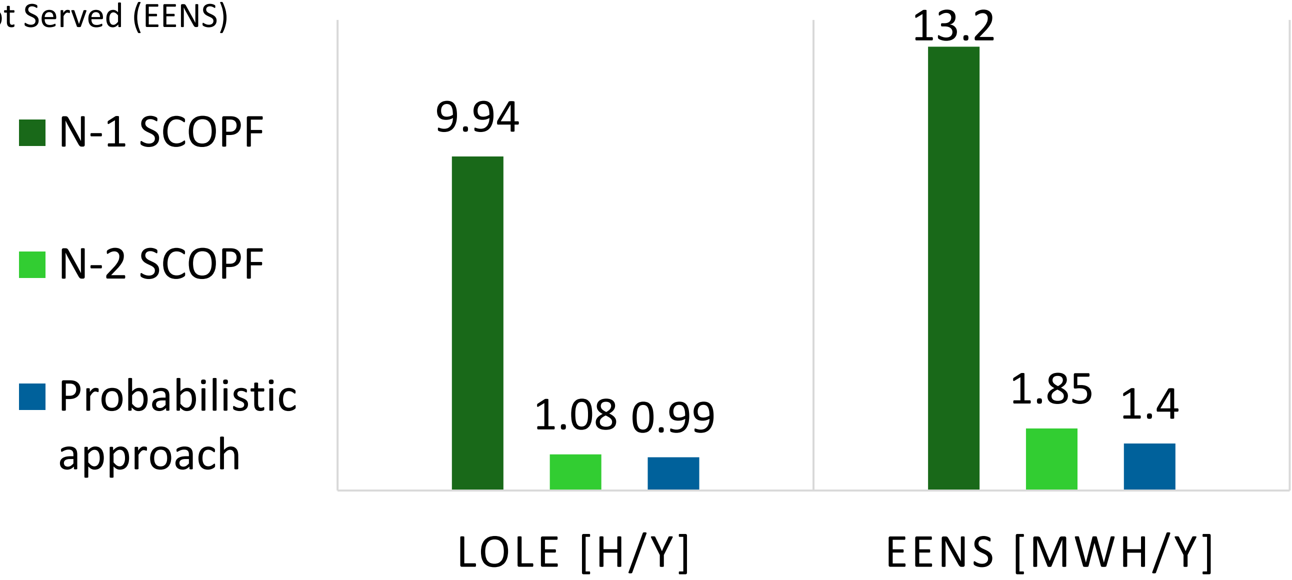
LOLE = 5 h/y
EENS = 7 MWh/y

Probabilistic security assessment

Proposed probabilistic security enhances reliability

Compare reliability indices

- Loss of Load Expected (LOLE)
- Expected Energy not Served (EENS)

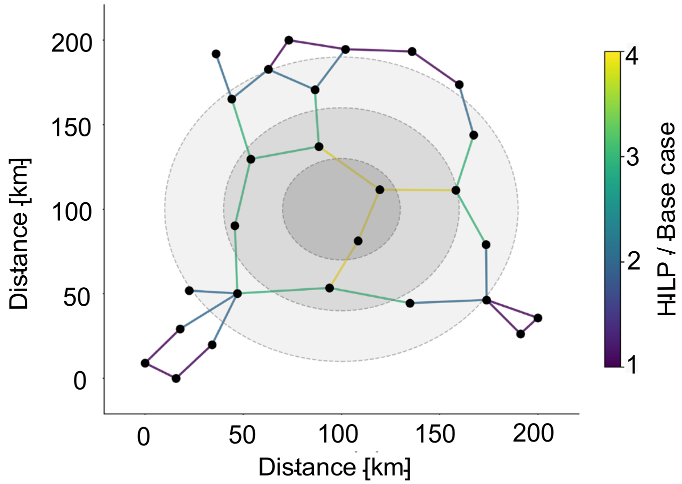


TU Delft

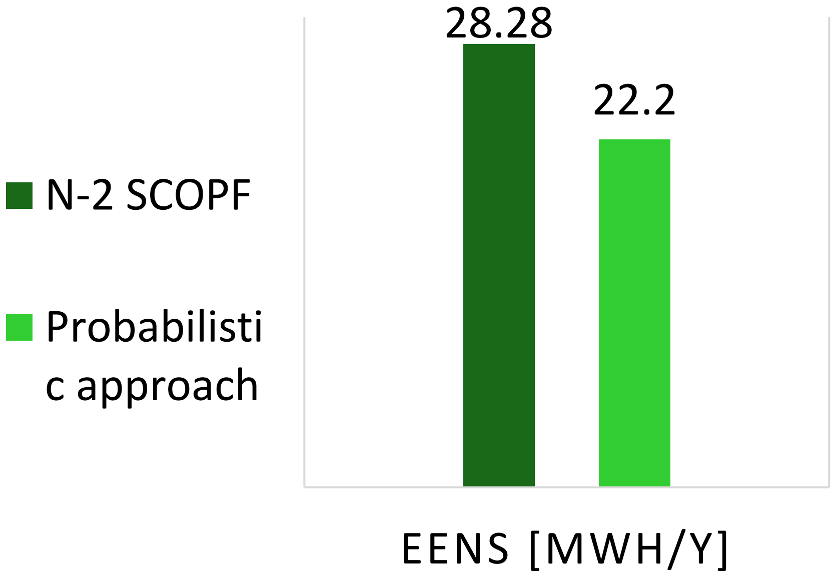
[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Extreme event

- Individual probabilities change due to an earthquake
- Recompute joint probabilities
- Recompute reliability indices



Potential for increased resiliency

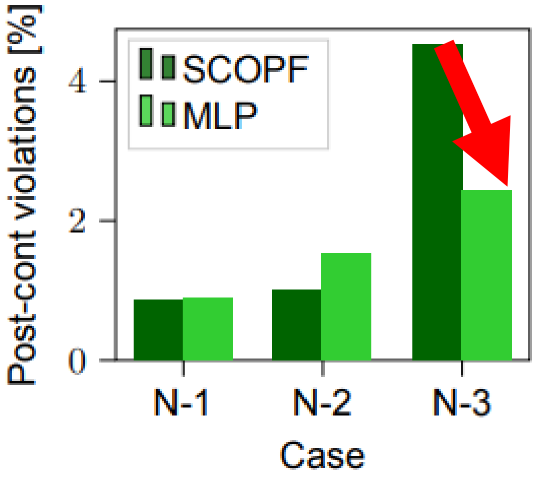


[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Performance 118-bus system

Proposed approach
Baseline

- Evaluate ability to identify line violations
- **Only** consider single, double or triple line outages
- Post-cont violations [%] indicates the percentage of samples where line violations occur



[16] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for nk security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.