All AI models are wrong, but some are useful … for power systems

Dr. Jochen L. Cremer
Associate Professor
www.jochen-cremer.com

TU Delft

Credits & team

Olayiwola Arowolo

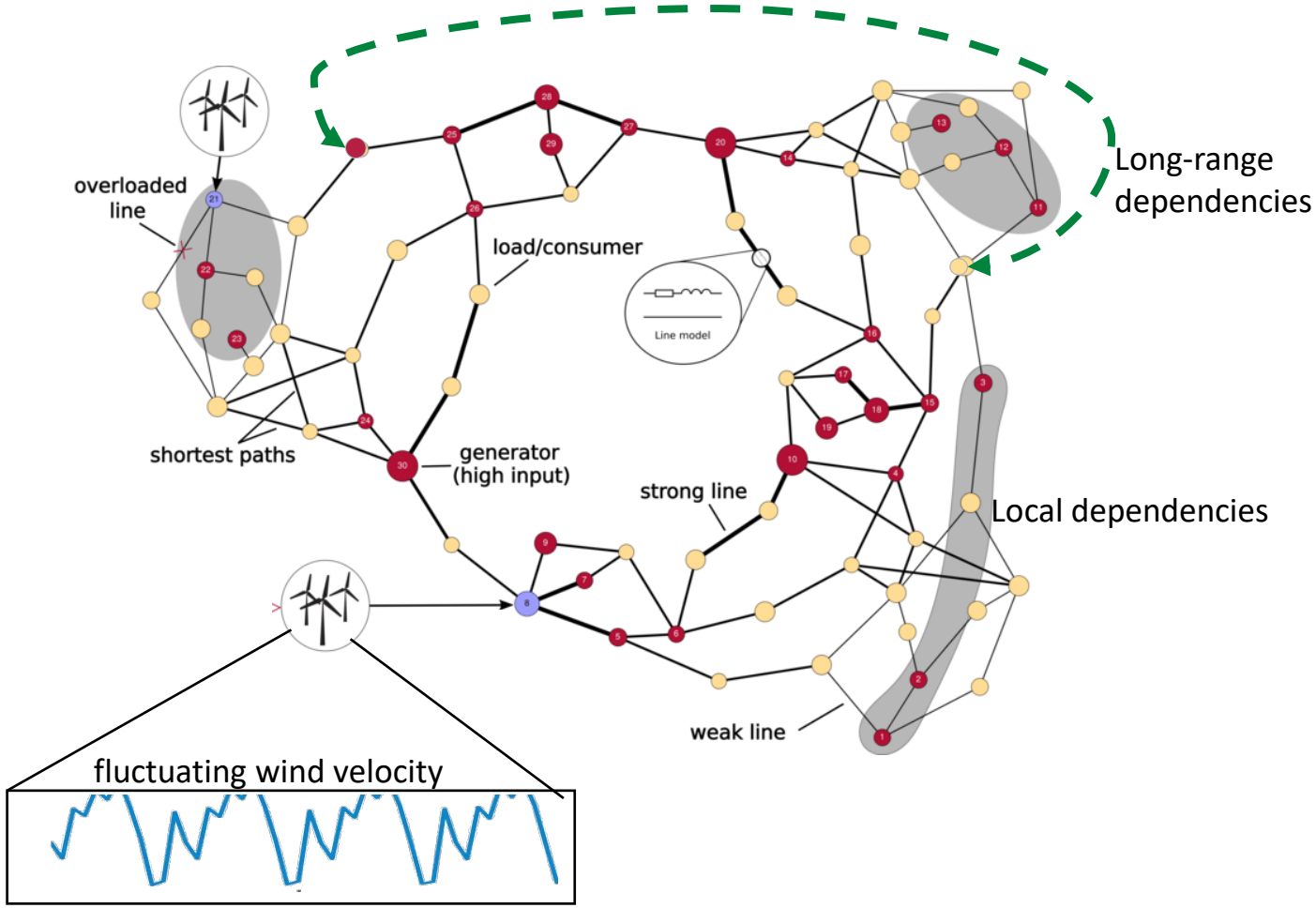
Maosheng Yang

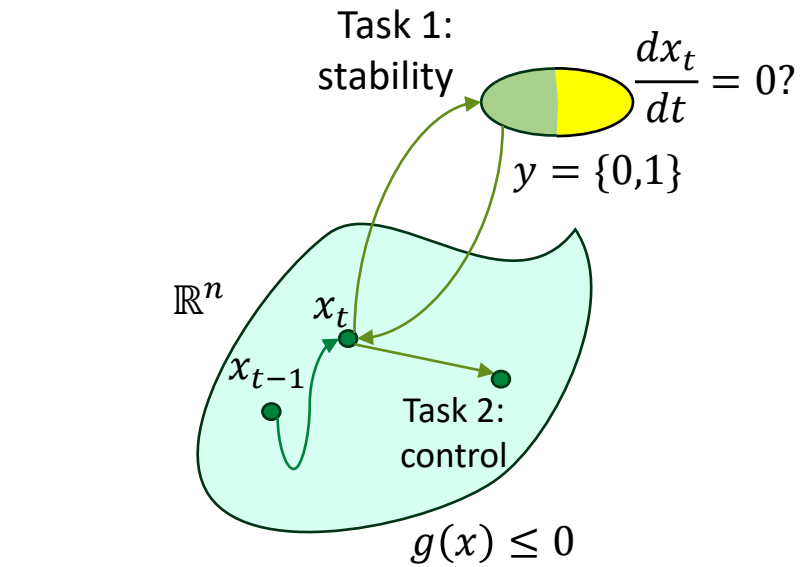
Jochen Stiasny

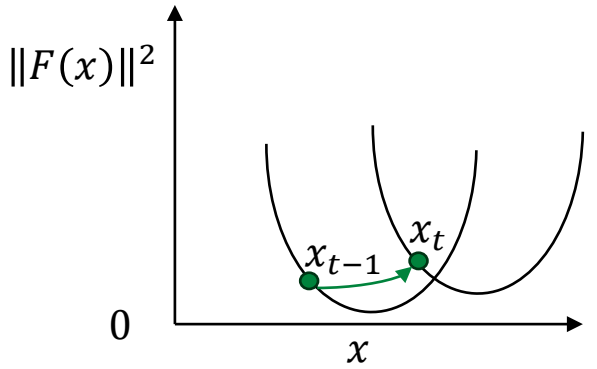
Haiwei Xie

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

Simplifying complex systems like the grid

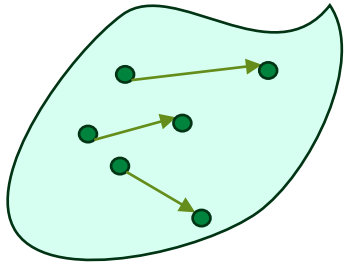






Supervised Learning for Surrogate Models

Notation: Power system s, model m, parameter x



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

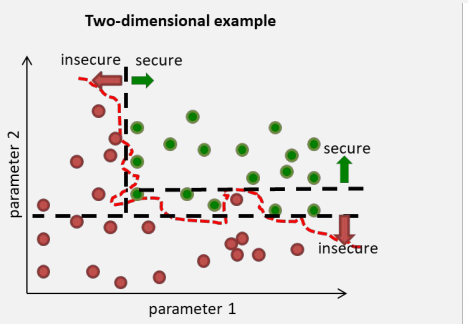
Surrogate approach

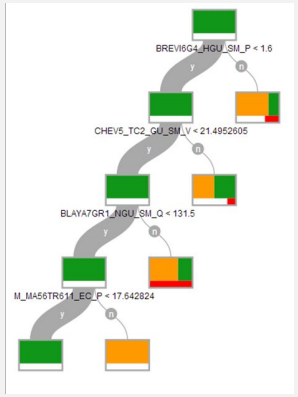
1. Generate a training dataset $\Omega^T = \{(x_i, y_i)\}_{i=1}^N$ where $y_i = m(x_i)$ from the full simulator
2. Train surrogate $f(x) \rightarrow \hat{y}$ with supervised loss $\sum_{i \in \Omega^T} \|y_i - \hat{y}_i\|$
3. Use $f(x_j)$ for new j $\notin \Omega^T$

Benefit: speed at inference

Applications

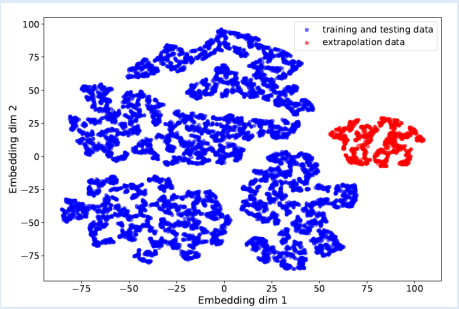
- Real-time dynamic security assessment ([1,2] and many others)





Challenges

- Out of distribution risks: What if s and m changes? e.g., topology changes
- What if the model is uncertain $s \neq m$? e.g., inverter-based controls
- Need large, representative training data



[1] Wehenkel, Louis, and Mania Pavella. "Decision trees and transient stability of electric power systems." *Automatica* 27.1 (1991): 115-134.
[2] 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.
[3] 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 generalisation

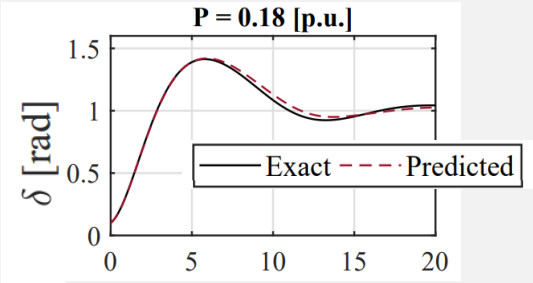
Physics-informed approach

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**

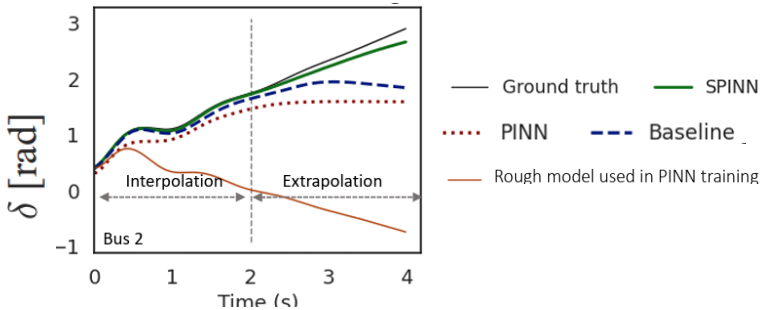
Applications

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



Challenges

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



[4] 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

[5] Stiasny, Jochen, Baosen Zhang, and Spyros Chatzivasileiadis. "PINNSim: A simulator for power system dynamics based on physics-informed neural networks." Electric Power Systems Research 235 (2024): 110796.

[6] 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.

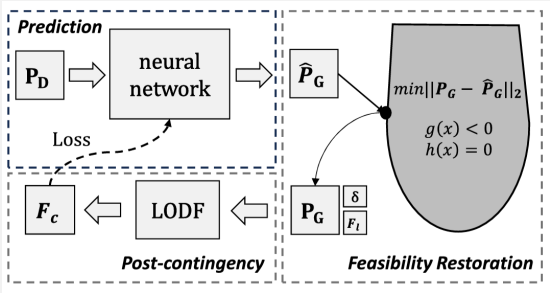
Approach

1. Generate many inputs $\Omega^T = \{(x_i)\}_{i=1}^N$
2. Model task loss $\sum_{i \in \Omega^T} \mathcal{L}(\tilde{m}(f(x_i))$
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[7]
- Replace conventional solvers with NN [8]
- Distribution system state estimation [9]
- N-k security constrained OPF [10]



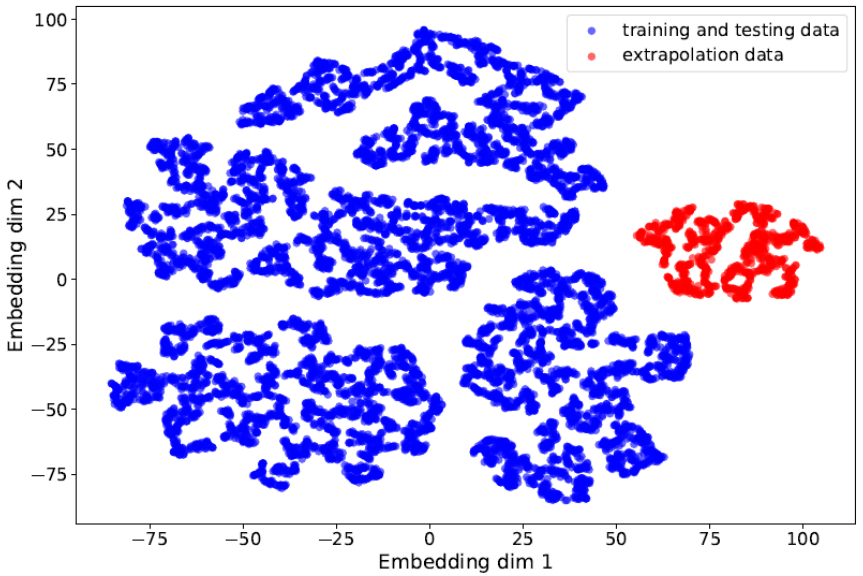
Challenges

- 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 generalisation

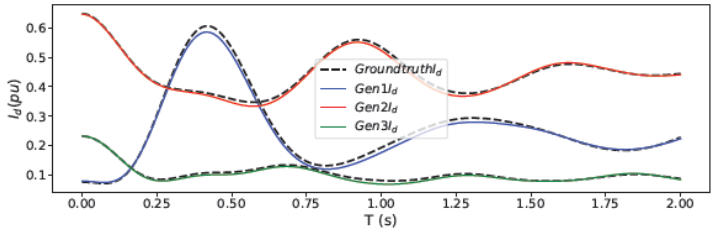
[7] Donti, P., Amos, B., & Kolter, J. Z. (2017). Task-based end-to-end model learning in stochastic optimization. *Advances in neural information processing systems*, *30*.
[8] Donon, B., Liu, Z., Liu, W., Guyon, I., Marot, A., & Schoenauer, M. (2020). Deep statistical solvers. *Advances in Neural Information Processing Systems*, *33*, 7910-7921.
[9] 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.
[10] 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.

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.*

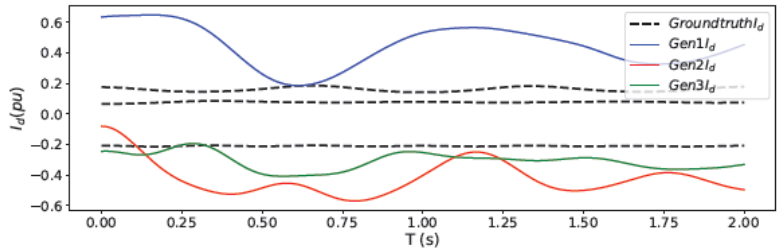


Extrapolation in continuous domain



(a) I_d current trajectory

Extrapolation in nonlinear domain (discrete)

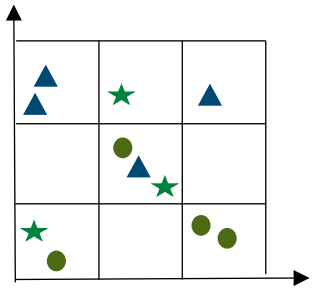


(a) I_d current trajectory

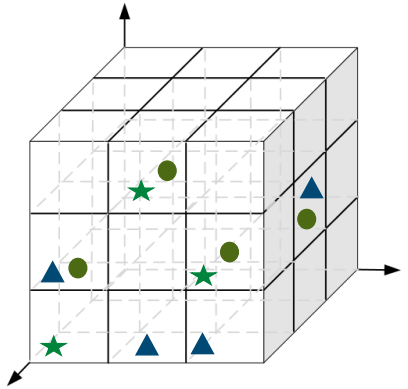
[11] Olayiwola Arowolo, Jochen Stiasny, Jochen Cremer, "Exploring the Extrapolation Performance of Machine Learning Models for Power System Time Domain Simulations", *12th Bulk Power System Dynamics and Control Symposium and Sustainable Energy, Grids and Networks Journal, 2025*

Curse of Dimensionality

1d: 3 regions



2d: 3^2 regions



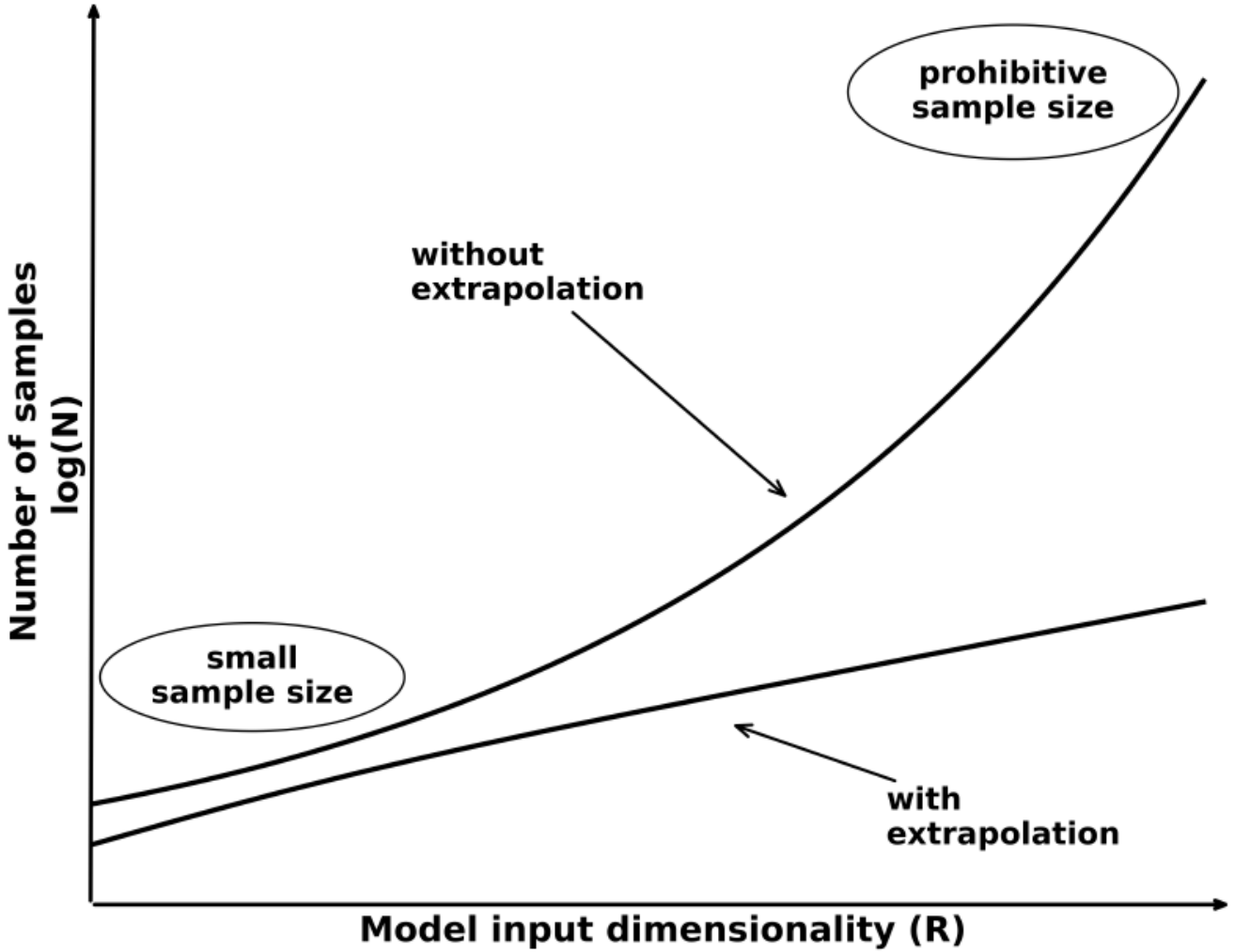
3d: 3^3 regions

1000d: hopeless

As dimensionality grows: fewer samples per region.

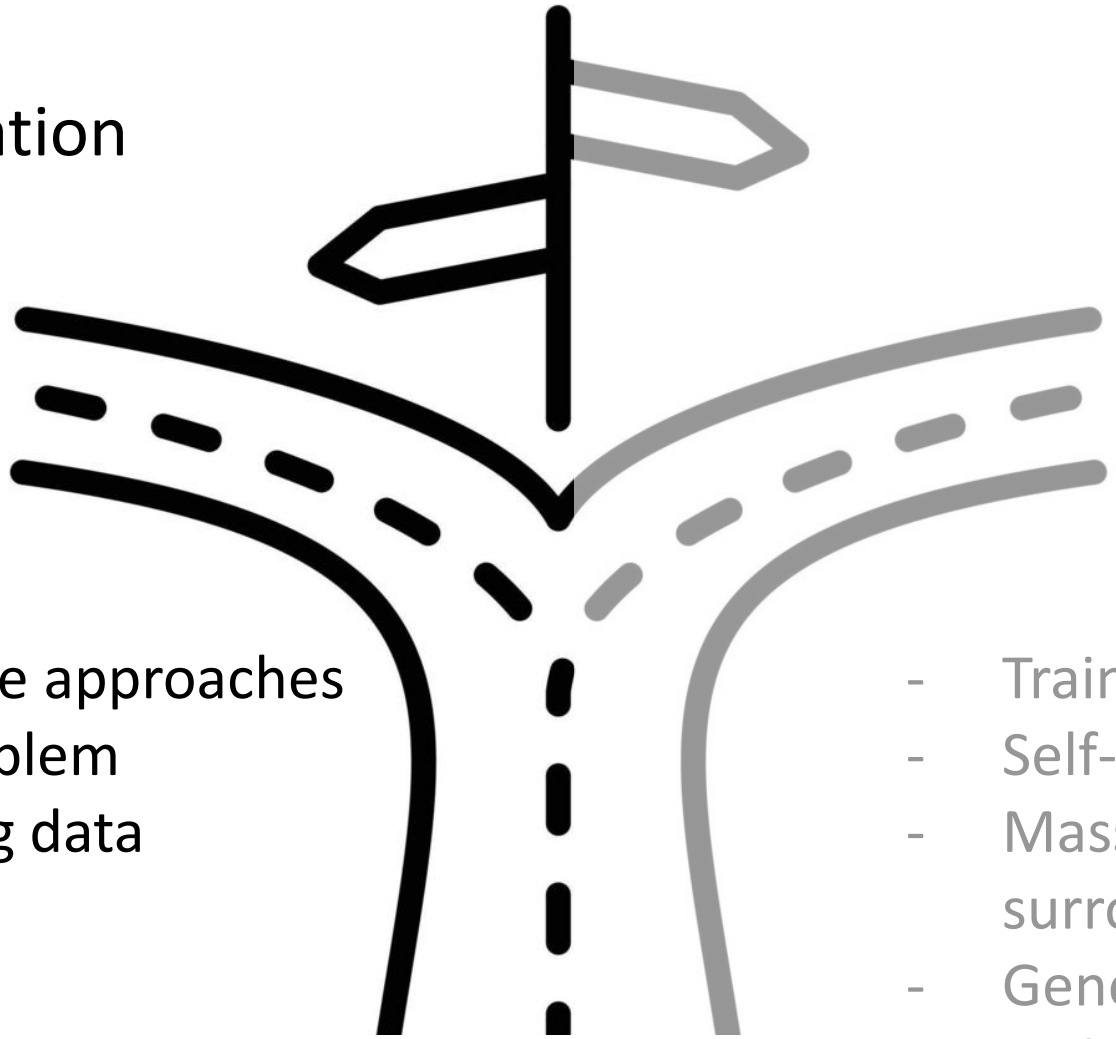
[12] James, Gareth, et al. *An introduction to statistical learning*. Vol. 112. No. 1. New York: springer, 2013.

Priority: think about extrapolation



Power systems ML research at a crossroad

Task specialisation



- Supervised-type approaches
- Exploit the problem
- Limited training data
- Hybrids…

Foundation models

- Trained for broad use
- Self-supervised learning
- Massive amount of ‘cheap’ surrogate data
- Generalisable model architecture…

Security constrained optimal power flow (SCOPF)

Objective:	minimize cost
Constraints:	In = out
	Generator limits
	Line flow limits
Contingency Constraints:	Line flow limits

$$\min_{n \in \Omega^G} \Sigma \, c_n P_{G_n}$$

$$\boldsymbol{B \cdot \delta = P_G - P_D}$$

$$P_{G_n}^{min} < P_{G_n} < P_{G_n}^{max} \; \forall n \in \Omega^G$$

$$F_l^{min} < F_l < F_l^{max} \; \forall l \in \Omega^L$$

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

Combinatorial complexity

Conventional approaches

Outaged line

Line flow changes

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

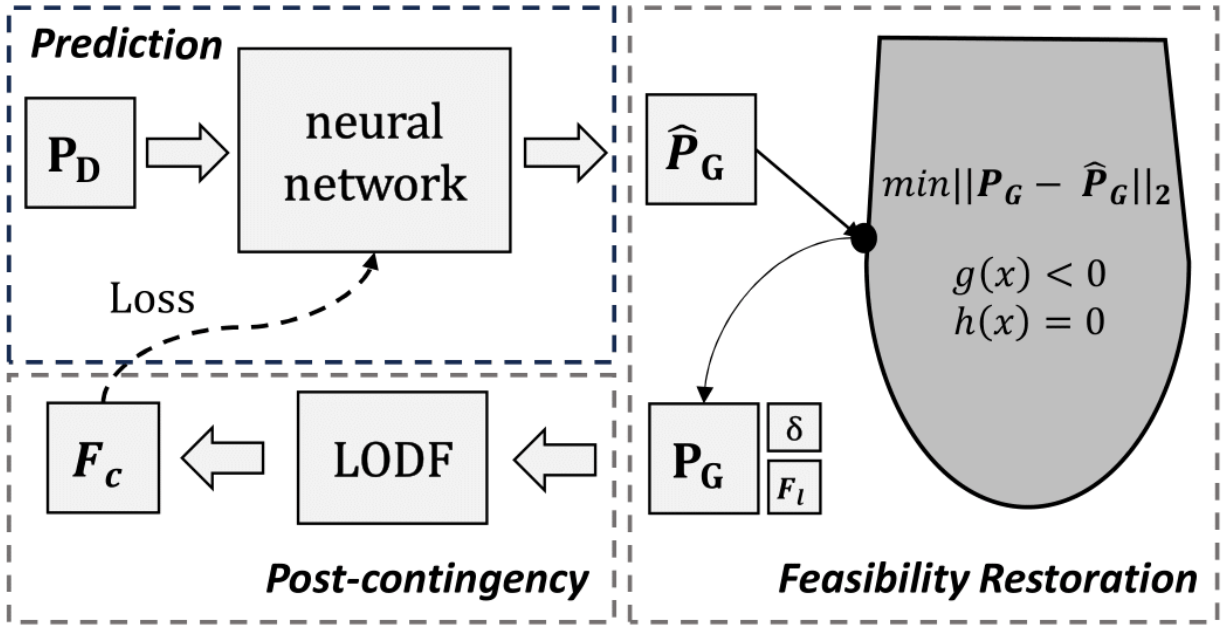
- Benders decomposition
- Column and constraint generation algorithm with robust optimization
- Line outage distribution factors (**LODF**)

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

- Intractable for increasing k

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

Proposed constraint-driven approach



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

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

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

4) Power imbalance $\lambda_2 \left\| \sum \widehat{\boldsymbol{P}}_\boldsymbol{G} - \sum \boldsymbol{P_D} \right\|_1$

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

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

[14] J. L. Cremer, "Polynomial Line Outage Distribution Factors for Estimating Expected Congestion and Security," *in IEEE Transactions on Power Systems,* vol. 40, no. 3, pp. 2532-2544, May 2025

Performance

Two baselines

- Contingency screening (CS) baseline: the N-k SCOPF optimization with iterative CS. At each iteration, the 20 most critical contingency cases are added to the SCOPF. We repeat the CS for 3 iterations.
- Heuristic (H) baseline: a set of critical contingency set is created offline on a separate dataset. The N-k SCOPF optimization is solved considering only this set of critical contingencies.

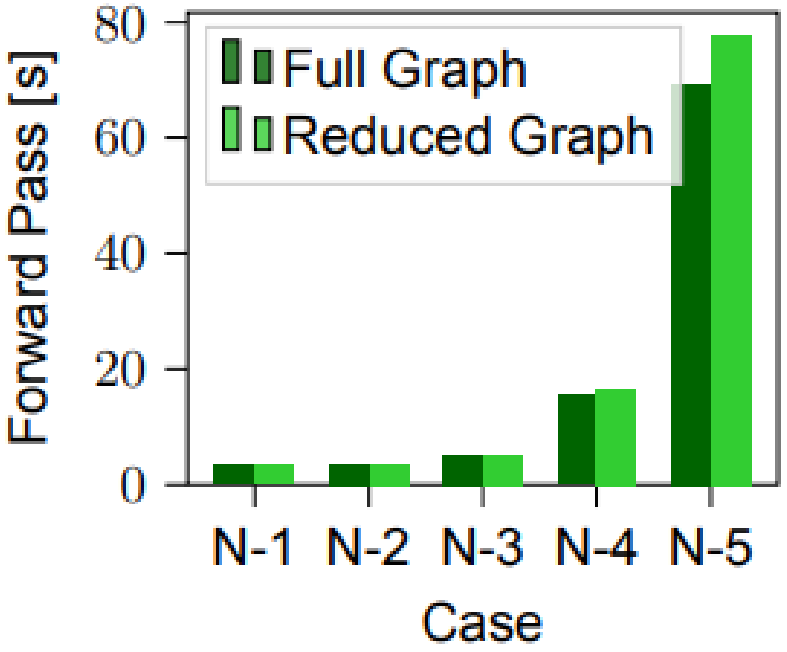
		39-bus		118-bus	
		Cost [%]	Speedup	Cost [%]	Speedup
CS	N-1	+1.44	12×	+2.51	158×
	N-2	+1.17	49×	−6.13	173×
	N-3	+0.51	41×	−5.29	27×
H	N-1	+1.45	15×	+2.88	76×
	N-2	+1.22	21×	−6.41	165×
	N-3	+0.10	21×	−4.92	15×

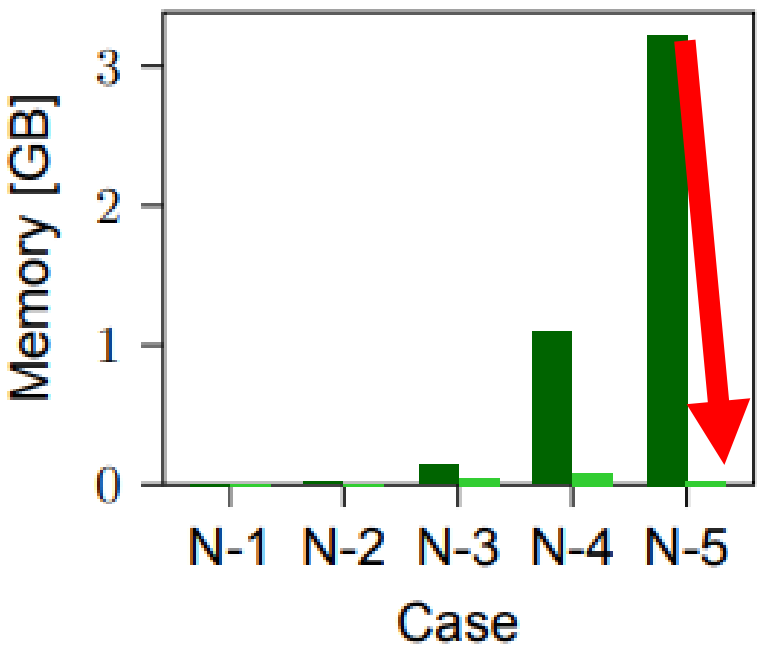
Post-contingency

System	Model	#violations [%]		
		N-1	N-2	N-3
39-bus	Proposed approach	0.31	1.33	2.16
	CS	0.06	0.74	3.30
118-bus	Proposed approach	1.08	1.70	2.64
	CS	0.91	1.03	5.88

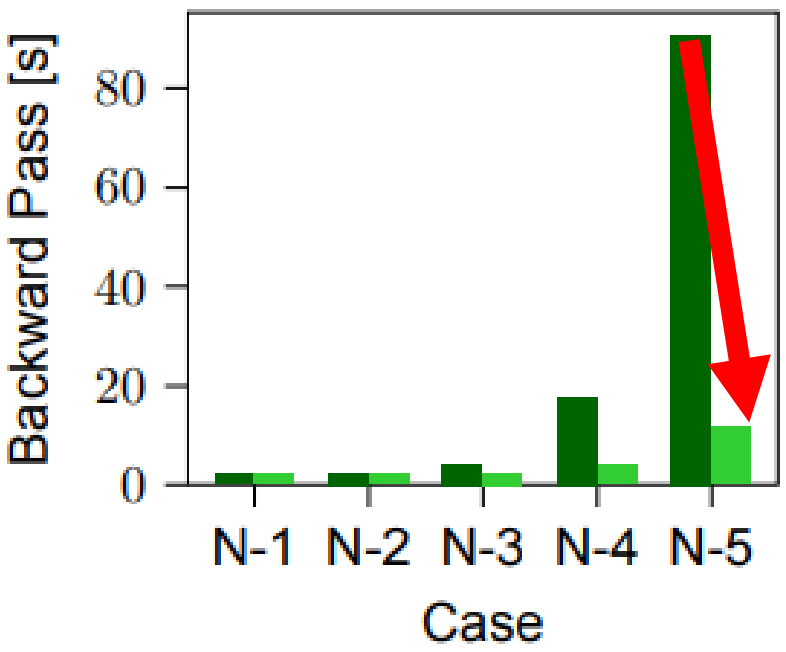
[13] Giraud, Bastien N., Ali Rajaei, and Jochen L. Cremer. "Constraint-driven deep learning for n-k security constrained optimal power flow." *Electric Power Systems Research* 235 (2024): 110692.

Reducing computational graph





Reduction in memory



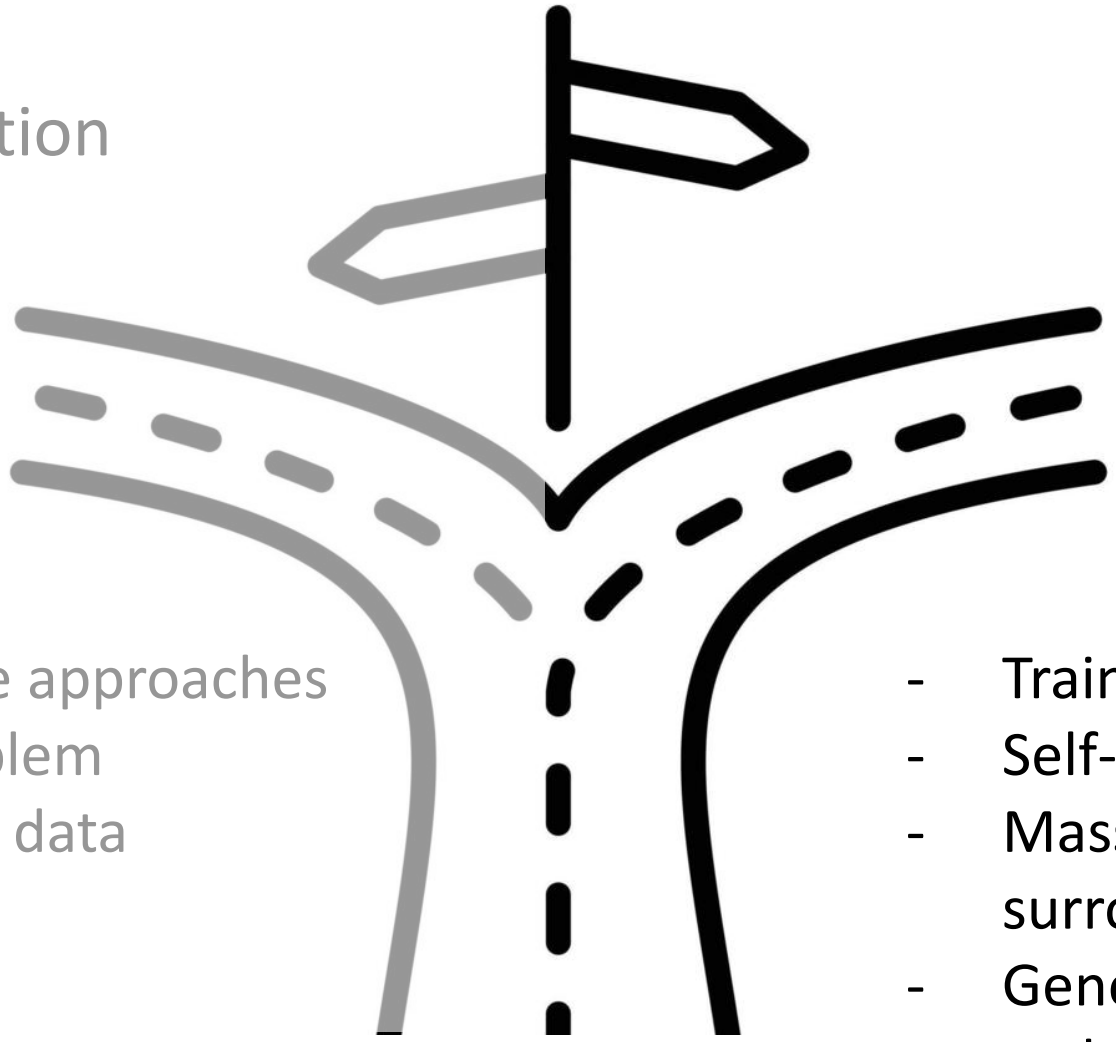
Reduction in computation time

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

Power systems ML research at a crossroad

Task specialisation

- Supervised-type approaches
- Exploit the problem
- Limited training data
- Hybrids…



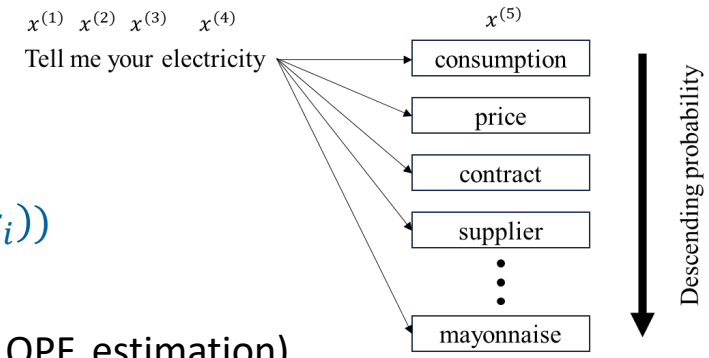
Foundation models

- Trained for broad use
- Self-supervised learning
- Massive amount of ‘cheap’ surrogate data
- Generalisable model architecture…

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)

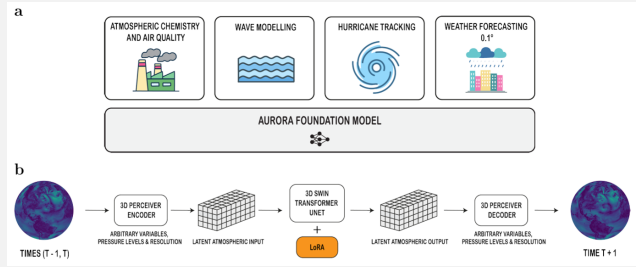
Benefits: Good initialisation when little data, good transfer to downstream tasks

Challenges

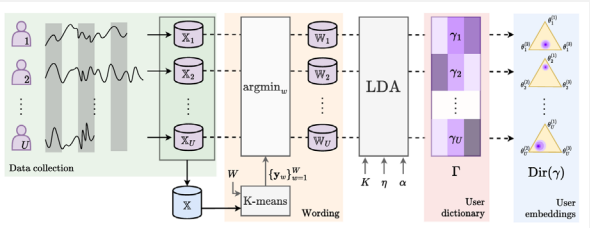
- Design pretext loss and model architectures with broad set of tasks, grid conditions, topologies
- Generate large data sets
- …

Applications

- Natural Language Processing
- Weather foundational models
- Earth system foundational models [13]



Load forecasting of users [14]



Grid foundation models (GFM) [15]

[15] 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*.
[16] Bölat, K., & Tindemans, S. (2024). GUIDE-VAE: Advancing data generation with user information and pattern dictionaries. arXiv preprint arXiv:2411.03936.
[17] 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.

We got a long way in other domains like images…

Prompt: “Take the picture and turn into a summer vibe with great swimming pool and a slide”, ChatGPT5.2, 11-02-2026

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

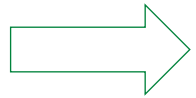
Those are the network parameters to be learned

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

Filter 1 Matrix

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

Filter 2 Matrix

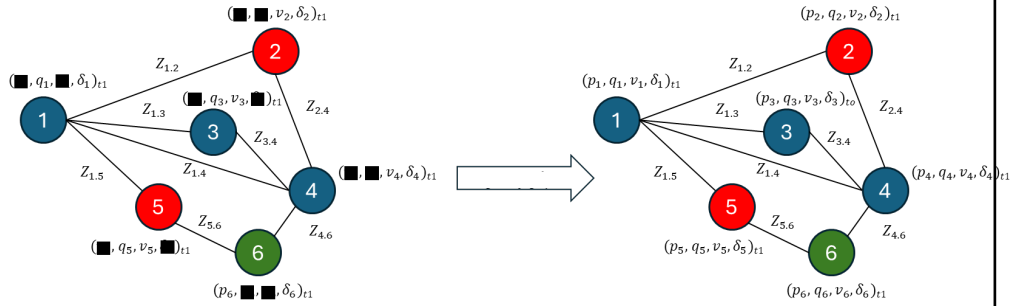
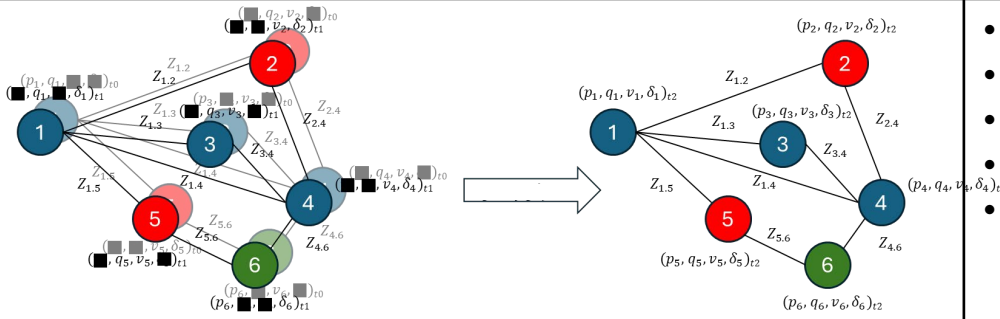
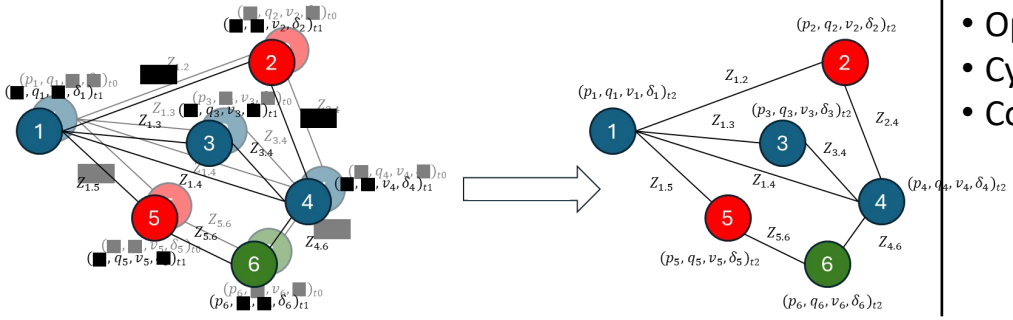


3	-1	-3	-1
-2	1	-1	-3
-2	-4	0	1
-1	0	-2	-1

Each filter detects a small pattern (3 x 3)

[18] LeCun et al. (1990) Handwritten digit recognition with a back-propagation network. In Advances in neural information processing systems.

Self-supervision

#	Pre-training	Related Applications
1	**Bus value Reconstruction** 	• Load flow • State estimation • (N-k) Contingency Analysis with k>1 • Expansion scenarios • (Optimal Power Flow)
2	**+ Temporal Reconstruction** 	• Load forecasting • Renewable energy forecasting • Look-ahead power flow • Look-ahead state transition • (Transient stability analysis)
3	**++ Edge Reconstruction** 	• Optimal expansion planning • Cybersecurity • Control operations

** IBM, Research Foundation Model Workshop Argonne, February 2025,*

Power Flow is at the heart of many power system tasks
How should we formulate a foundational PF?

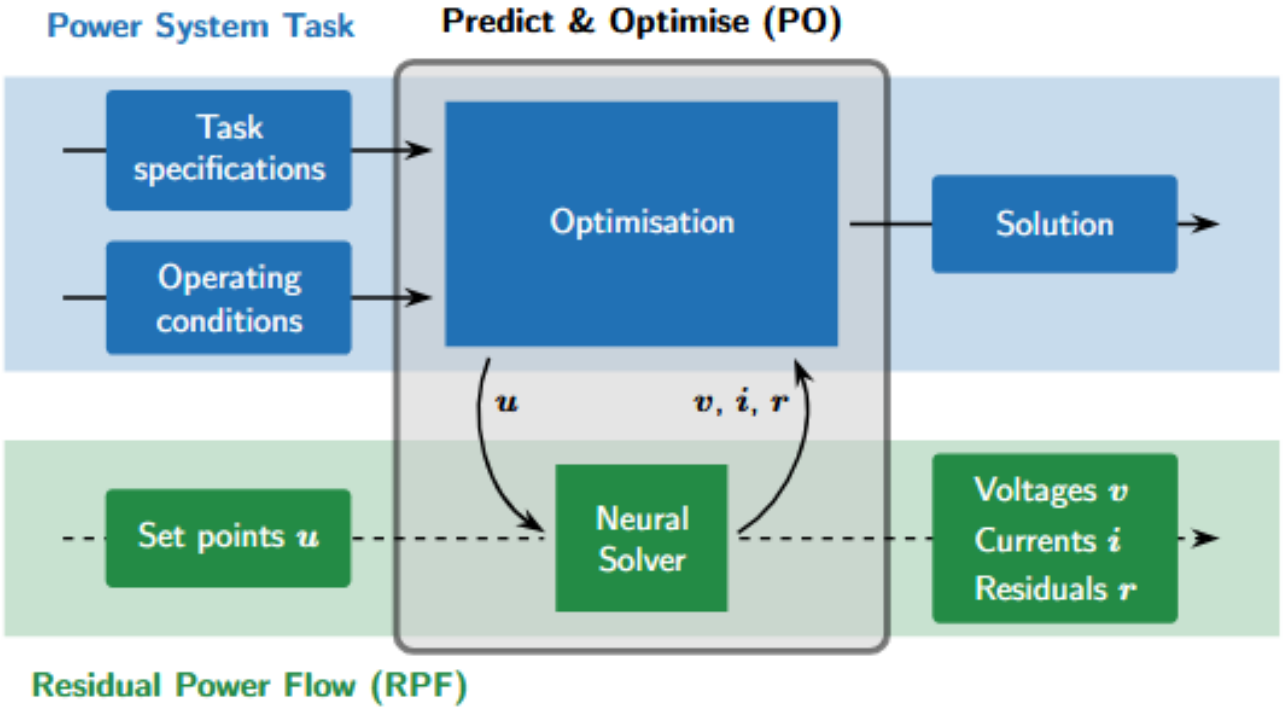
Jochen Stiasny

Residual power flow (RPF)

- RPF quantifies infeasibility
- Simpler formulation for neural solvers

Predict-and-Optimise (PO)

- Flexible handling of constraints and objectives
- While minimising infeasibility



[24] 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.
[25] Stiasny, Jochen, and Jochen Cremer. "Residual Power Flow for Neural Solvers." arXiv preprint arXiv:2601.09533 (2026).

Conclusions

- All models are approximations: usefulness depends on structure and aligning with task
- Purely supervised surrogates struggle with system shift and dimensionality
- Injecting structure (physics, constraints, topology) improves generalisation
- Constraint-driven learning may enable scalable security-constrained optimization
- Foundation-style approaches are promising, however, power systems require domain bias
  - -> The “Power System Neural Network” still needs to be invented…
  - -> Power flow may be the right abstraction for building grid foundation models

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

Graph Neural Networks as architecture?

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

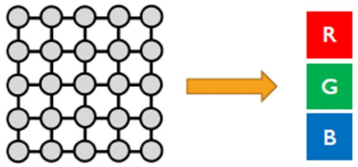
Idea: embedding graph topology directly into the model architecture as bias

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'

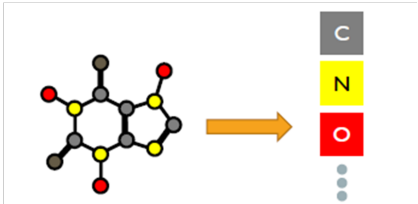
Benefits: Data efficient, generalisation to changes in topologies

Example: $p \times p$ RGB image



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

Example: molecular graph

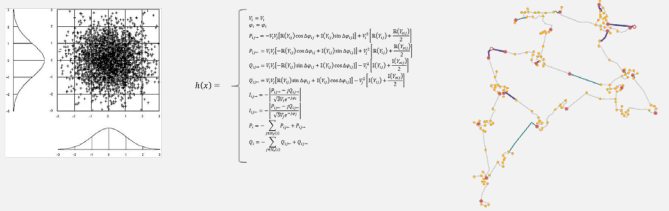


$\Omega = \{1, .. n\}$ $\quad \mathcal{C} = \mathbb{R}^c$

Applications

- Graph neural solvers [19] for ACOPF [20,21], powerflow [22]
- Distribution system state estimation [23]

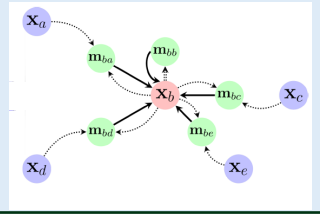
Noisy measurements | Power flow equations | Topology



Challenges

- Model uncertainty $s \neq m$
- 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



[19] 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.
[20] 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*.
[21] Olayiwola Arowolo, Jochen L Cremer, "Towards Generalization of Graph Neural Networks for AC Optimal Power Flow" , https://arxiv.org/pdf/2510.06860
[22] Lin, Nan, Stavros Orfanoudakis, Nathan Ordonez Cardenas, Juan S. Giraldo, and Pedro P. Vergara. "PowerFlowNet: Power flow approximation using message passing Graph Neural Networks." International Journal of Electrical Power & Energy Systems 160 (2024): 110112.
[23] 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.