IEEE Powertech 2021
Panel Session: “Big Data and Machine Learning for Power Systems”

Interpretability and verification of neural networks: Removing barriers for power system applications

Spyros Chatzivasileiadis
Associate Professor, DTU

Machine learning is not a calculation.
It is an estimation.

- Although machine learning **is able to, it might never be as accurate** as a model that fully describes a system or process.
  - Why: ML does not calculate a function. It estimates its result.

- Then, **why** shall we apply Machine Learning?

1. **extremely fast**

2. good alternative if we do not have full knowledge of the actual model
   - Handle **very complex systems**
   - **Infer** from **large amounts of data**
   - **Infer** from **incomplete data**

3. For many purposes, an **estimation is good enough**
   - e.g. online control, real-time forecasting, and many others

ML Barriers for Power systems

- **ML can work well for forecasting/predicting**
  - Weather/wind/PV or load forecasting, prediction of electricity prices, prediction of failures
  - **But:** still, many are reluctant, as they see it as a black box → **no interpretability**

- **ML has found no real use for safety-critical applications** (except Decision Trees)

1. Why would we depend on **incomplete data**, when we have developed **detailed physical models** over the past 100 years?

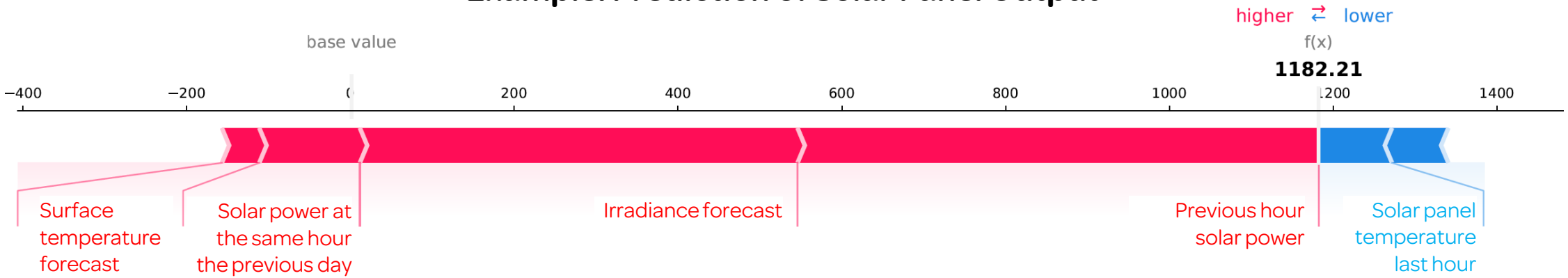
2. Why would we use a "black box" to decide about **a safety-critical application**?
   - Examples: security assessment, power system optimization/optimal power flow

3. Accuracy is a purely statistical ML performance metric. Who guarantees that the Neural Network can handle well previously unseen operating points?

Neural Network Interpretability

- **Rigorous methods** to assess the contribution of each feature to the Neural Network prediction
  - Integrated Gradients, Expected gradients, DeepLIFT

Example: Prediction of Solar Panel Output



Assessment of the contribution of different features:

1. "Previous hour solar power" and "Irradiance forecast" have large contributions to the prediction of the solar power output
2. High "Solar panel temperature of last hour" reduces the solar power output prediction of this hour

Why accuracy is not enough?

- Example: Power System Security Assessment; Classify SAFE or UNSAFE

Total operating points = 1000	Actually Safe (Total = 20)	Actually Unsafe (Total = 980)
Predicted safe	1	30
Predicted Unsafe	19	950

$$\text{Accuracy} = \frac{1+950}{1000} = 95\%$$

- 95% accurate but we have misclassified almost all truly safe points!
- Even with additional metrics e.g. “Recall”, “Specificity”, “Precision”, etc. → all metrics are dependent on the quality of discrete data samples

Why accuracy is not enough?

- Example: Power System Security Assessment; Classify SAFE or UNSAFE

Total operating points = 1000	Actually Safe (Total = 20)	Actually Unsafe (Total = 980)
Predicted safe	1	30
Predicted Unsafe	19	950

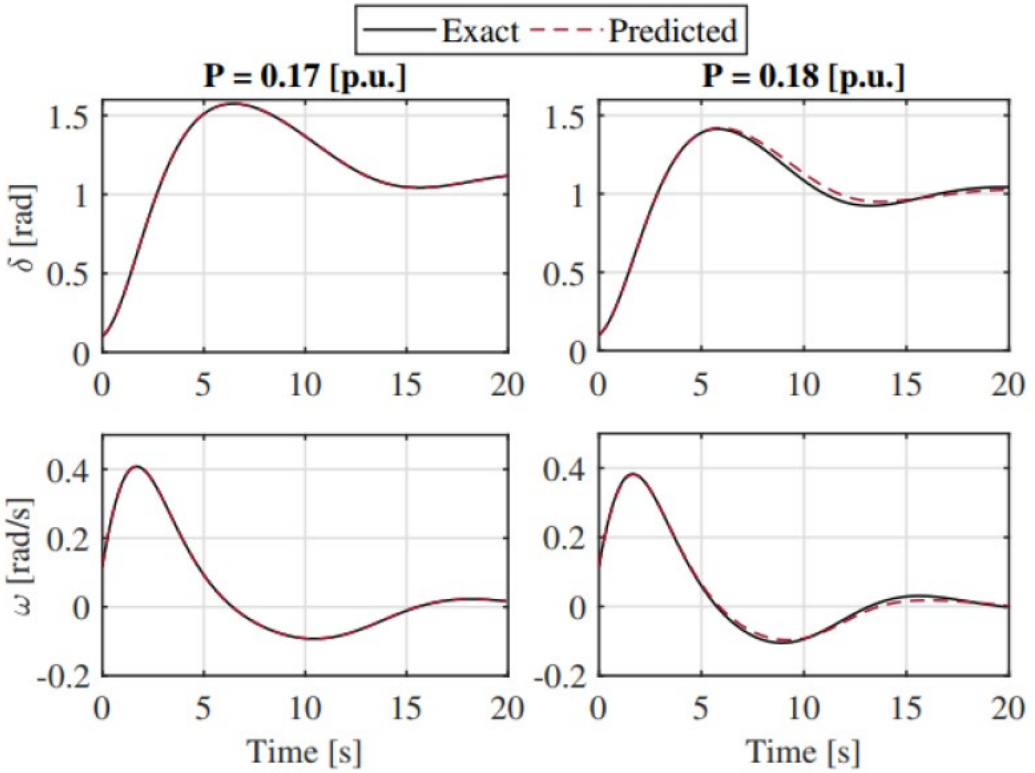
$$\text{Accuracy} = \frac{1+950}{1000} = 95\%$$

- 95% accurate but we have misclassified almost all truly safe points!
- Even with additional metrics e.g. "Recall", "Specificity", "Precision", etc. → all metrics are dependent on the quality of discrete data samples

- **We need Neural Network Verification** for safety-critical applications
  - Formal guarantees about the classification/regression over continuous input regions
  - We no longer rely on statistical performance metrics
  - Systematic identification of adversarial examples

Why use incomplete datasets when we have detailed physical models?

- **Physics-Informed Neural Networks**
  1. Include the physical models inside the NN training → need for less data, probably smaller NN sizes
  2. **Extremely fast:** can potentially replace solvers for systems of differential-algebraic equations
  3. Turn **NN training** from supervised to **unsupervised learning**



Single Machine Infinite Bus Example:
Physics-Informed NN is 87x faster
than ODE solver

Takeaways

1. **Extremely fast** (up to 1'000x faster): Revolutionize power system computation → replace non-linear and differential-algebraic solvers for e.g. OPF, sec. assessment, etc.

2. ML can handle very complex systems or incomplete data

We can now build trust in AI!

3. **Interpretable AI**: if we are able to explain how ML behaves when it estimates an operating point, we can remove the barriers for its application in power systems

4. We need **Verification of Neural Networks** for safety-critical applications:
   - worst-case guarantees for NN behaviour can build the missing trust!

5. **Physics-Informed Machine Learning**: Use the already available models in the NN training

---

6. Before you apply ML/RL on power systems: Think! **ML/RL is an estimation, not a calculation!** (for now)
   - Given infinite time or data ML/RL will converge to the correct decision/optimal strategy. Do you have solid reasons why spending so many resources for ML/RL training will lead to a better performance than a Linear Program or Kalman filter? If not, then better stick to the conventional approach.

Thank you!

Spyros Chatzivasileiadis

Associate Professor, PhD

www.chatziva.com

spchatz@elektro.dtu.dk

A. Venzke, S. Chatzivasileiadis. Verification of Neural Network Behaviour: Formal Guarantees for Power System Applications. Accepted at IEEE Trans. on Smartgrid. 2020.
https://arxiv.org/pdf/1910.01624.pdf

A. Venzke, G. Qu, S. Low, S. Chatzivasileiadis, Learning Optimal Power Flow: Worst-case Guarantees for Neural Networks. **Best Student Paper Award** at IEEE SmartGridComm 2020.[.pdf | slides | video]

G. S. Misyris, A. Venzke, S. Chatzivasileiadis, Physics-Informed Neural Networks for Power Systems. Presented at the **Best Paper Session** of IEEE PES GM 2020. https://arxiv.org/pdf/1911.03737.pdf

J. Stiasny, G. S. Misyris, S. Chatzivasileiadis, Physics-Informed Neural Networks for Non-linear System Identification applied to Power System Dynamics. IEEE Powertech 2021.
https://arxiv.org/pdf/2004.04026.pdf

Y. Lu, I. Murzakhanov, S. Chatzivasileiadis, Neural network interpretability for forecasting of aggregated renewable generation. Submitted. https://arxiv.org/pdf/2106.10476.pdf

Some datasets and code on Machine Learning for Power systems (e.g. Physics-Informed ML, NN verification, etc)

available at:

www.chatziva.com/downloads.html