info@csem.ch www.csem.ch

DIGERATI: a dynamic graph machine learning solution for short-term PV forecasting

Rafael E. Carrillo, Baptiste Schubnel, Renaud Langou and Pierre-Jean Alet

CSEM SA, Switzerland

Introduction

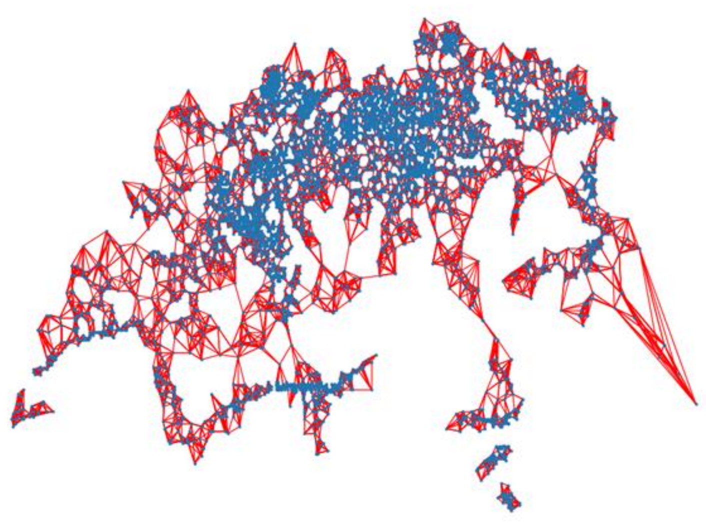
- CSEM has already demonstrated that graph neural networks (GNN) can outperform the state of the art in forecasting photovoltaic production
- Up to now such approaches have only been able to handle data from a fixed network of sensors (nodes) to produce the desired forecasts
- However, in real life, nodes can be frequently added or removed as new customers sign up or physical assets change
- Also, in real situations input data is imperfect affected by measurement and transmission errors
- The DIGERATI[1] project developed a forecasting solution based on dynamic graph machine learning to overcome these problems
- An online demonstrator has been made to showcase the capabilities of the solution

[1]This research was co-financed by Innosuisse – Swiss Innovation Agency – project 57765.1 IP-EE: DIGERATI.

Graph-based multi-site PV forecasting

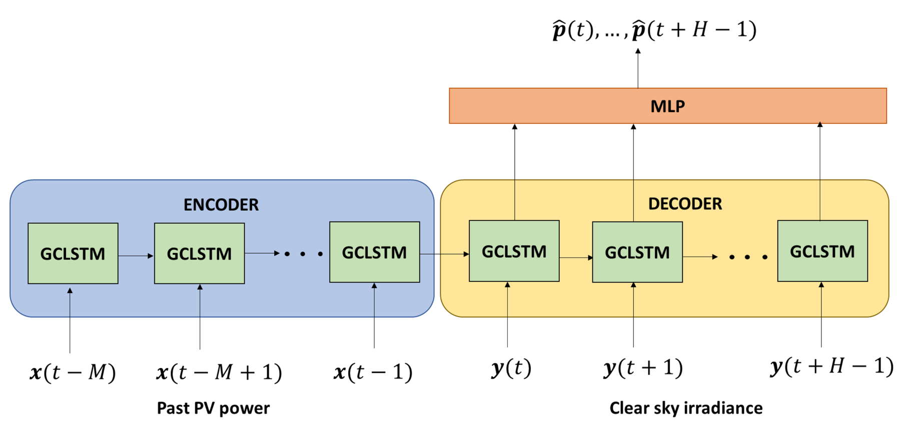
Intuition

- CSEM's data-driven solution relies entirely on measured data
- PV stations can be used as a network of virtual weather stations
- By exploiting the spatio-temporal relations of the power production data, cloud movements can be forecasted



Core graph forecasting model

- The forecast model is an encoder-decoder architecture with graph-convolutional Long-Short-Term-Memory (GCLSTM) cells[2]
- Graph convolutions capture spatial relations and propagate information in the spatial domain
- LSTM models the temporal dynamics in a similar fashion to a (non-linear) state space model

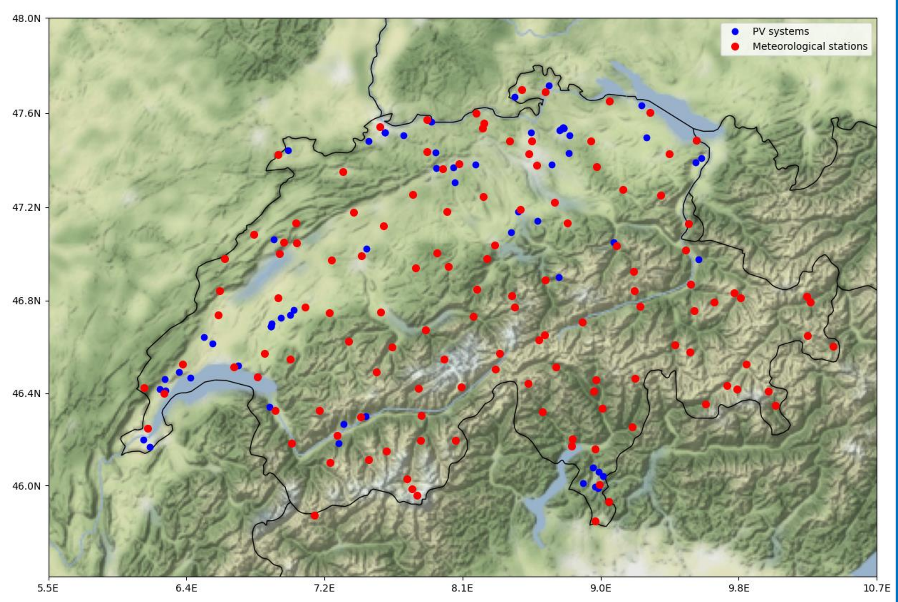


[2]J. Simeunović, B. Schubnel, P. -J. Alet and R. E. Carrillo, "Spatio-Temporal Graph Neural Networks for Multi-Site PV Power Forecasting," in IEEE Transactions on Sustainable Energy, vol. 13, no. 2, pp. 1210-1220, April 2022.

The DIGERATI solution

Main characteristics of DIGERATI:

- Global approach to make localized forecasts
- Fusion of heterogeneous data sources (PV power and meteorological measurements)
- Robust to real-life data
  - Automatic monitoring of data quality for all sources (missing data)
  - Dynamic graph approach adds robustness to missing data and changing sensor network
- Probabilistic forecasts of the 5%, 50% and 95% quantiles
  - 50% quantile (median) estimate taken as the forecast value
- Irradiance forecast for any "unseen" location



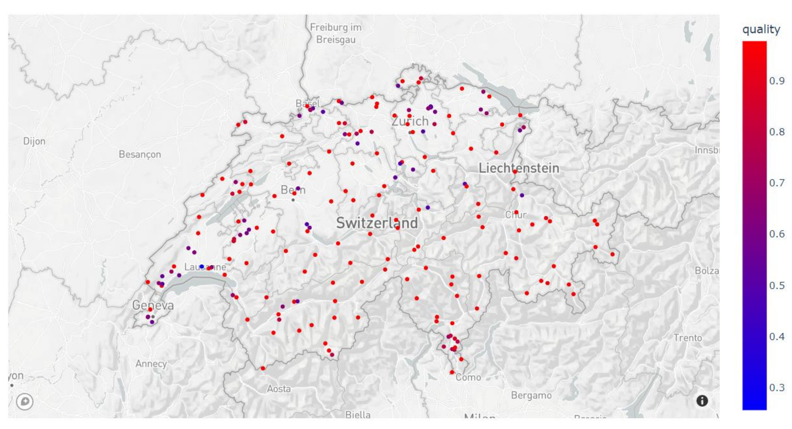
Input data:

- Past 3 hours of measured data
- Power at PV nodes
- GHI, temperature, wind speed and wind direction at meteorological nodes

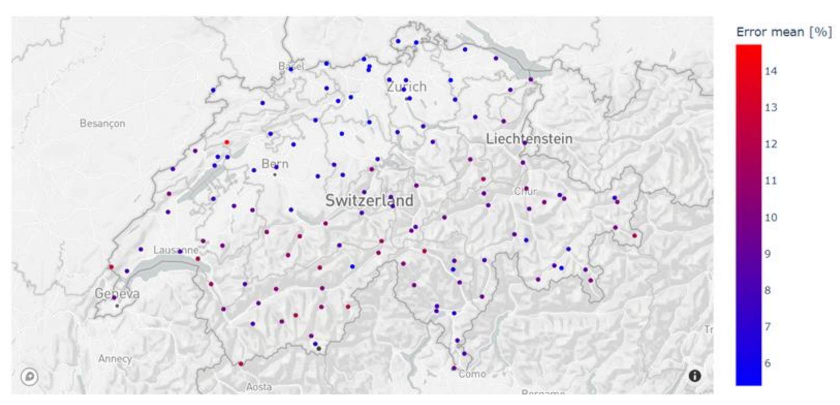
Demonstrator and results

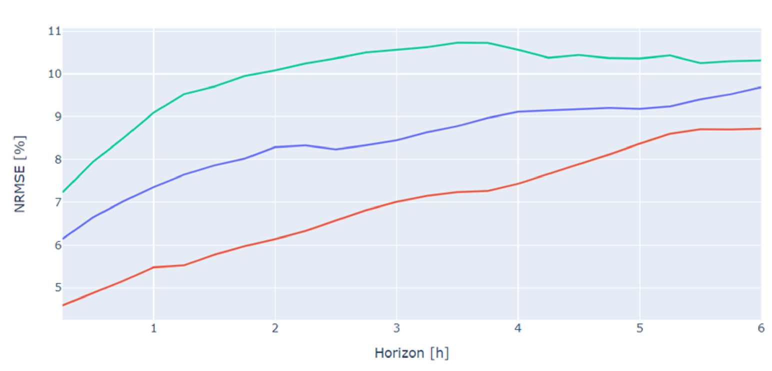
Live demonstrator

- Current web demonstrator uses data from 192 nodes as inputs:
  - 64 PV stations
  - 128 meteorological stations
- Constant forecast of irradiance for the 128 meteorological stations with a horizon of up to six hours ahead with 15 minutes temporal resolution
  - Capable of forecasting irradiance for any location in Switzerland
- API available to retrieve irradiance forecasts for the desired locations
  - Update frequency of 15 minutes
- Dashboard for **automatic monitoring of data quality** for all nodes
  - Color coding indicates the effective uptime of data sources
  - Daily update and system alerts



- Dashboard for **automatic monitoring of forecast quality**
  - Peak normalized root mean square error (NRMSE) as quality metric
  - Daily update of the NMRSE over the past 7 days for all nodes
  - Spatial visualization of errors
  - Aggregated visualization of the NRMSE as a function of the forecasting horizon
  - Visualization of individual node error is also provided





Quantitative evaluation

- We evaluated the accuracy, reliability and sharpness of the probabilistic forecasts over one year of historical data
  - Training in one year, evaluation in the following year
  - Evaluation in four different seasons
- Accuracy: NRMSE
  - Comparison between complete data (all available nodes) and dynamic changes in the node set
  - Simulated deletion of 10% of the nodes in both training and evaluation
  - Changes in the node set mainly impacts the accuracy of the first hour
- Reliability: prediction interval coverage probability (PICP):
  - It measures the reliability of the confidence intervals (5%-95% quantiles)
  - High coverage of the prediction intervals for the entire horizon
- Sharpness: continuous rank probability score (CRPS):
  - Global metric that describes how well the forecasted probability distributions are
  - CRPS reported as percentage of the peak value for each node
  - Overall, the system reproduces well the empirical distribution of the data