Solar power nowcasting to days-ahead post-processing for extreme events: a modular approach including synthetic data

DESTINATION EARTH

Petrina Papazek, Pascal Gfäller, Irene Schicker
GeoSphere Austria – Analysis and Model Development/Post-Processing, Hohe Warte 38, Vienna

correspondence: Petrina.papazek@geosphere.at

Funded by the European Union Destination Earth implemented by ECMWF esa EUMETSAT

Introduction and data

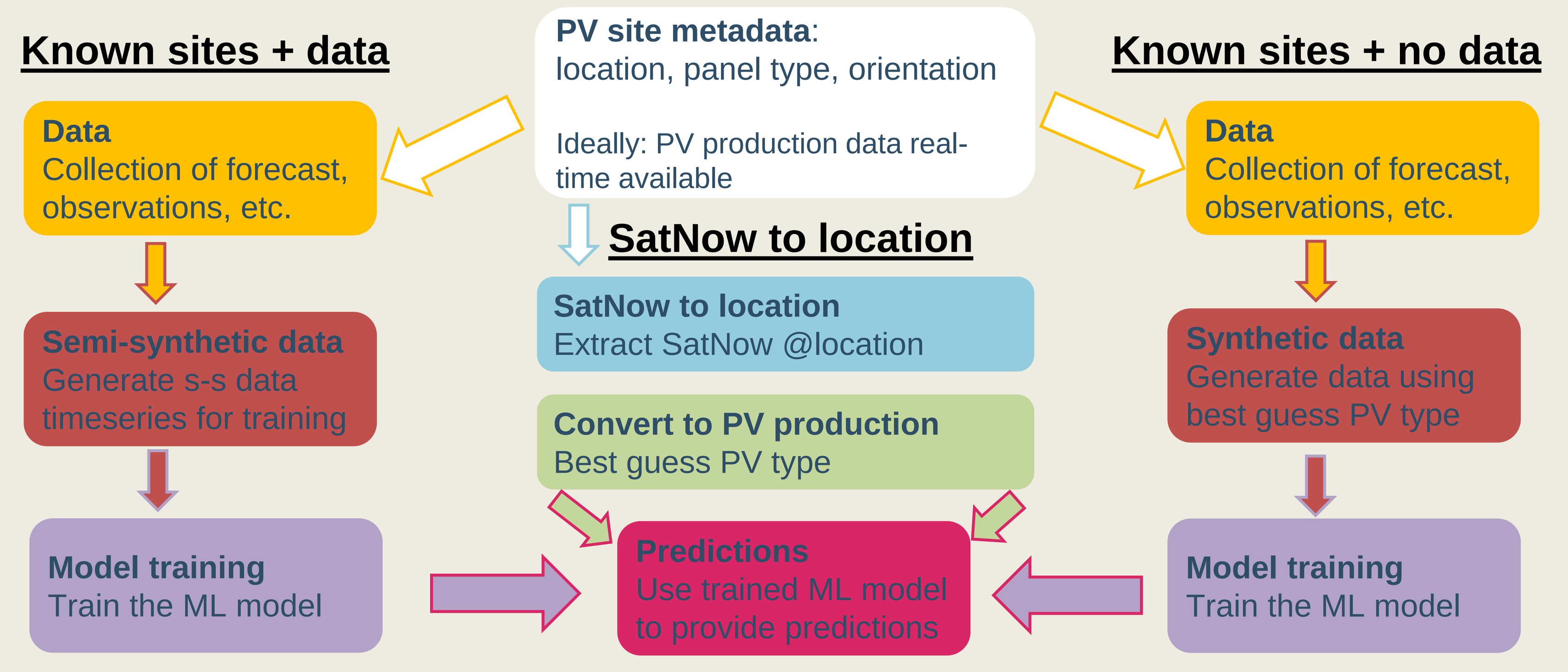
Forecasting solar irradiance, nowcasting to short-range, is crucial in managing solar power. With fast GPUs, deep learning techniques may learn complex structures across time-series in heterogenous datasets. We study deep learning for irradiance/ghi using, here as an example, the BSRN network as well as MSG LSA SAF data.

The focus of our current method is to provide location optimized PV forcasts providing daily hourly from 6:00-16:00 with a 15 minute update frequency in the nowcasting to short-range. Also we aim for spatial irradiance nowcasts.

Data sources are:

- **BSRN**: Baseline Surface Radiation Network
- **ERA5-land**: hourly reanalysis in high resolution of surface variables
- **CAMS**: radiation time series of the ADS data store
- **AROME Austria**: high resolution NWP for the alpine region, hourly forecasts
- **Observations**: observations obtained at a nearby weather observation/synoptic site (quality controlled), e.g.: Sonnblick Observatory
- **MSG**: for spatial GHI nowcasting, the LSA SAF DSSF_TOT product
- **PV production:** of selected Austrian and German sites

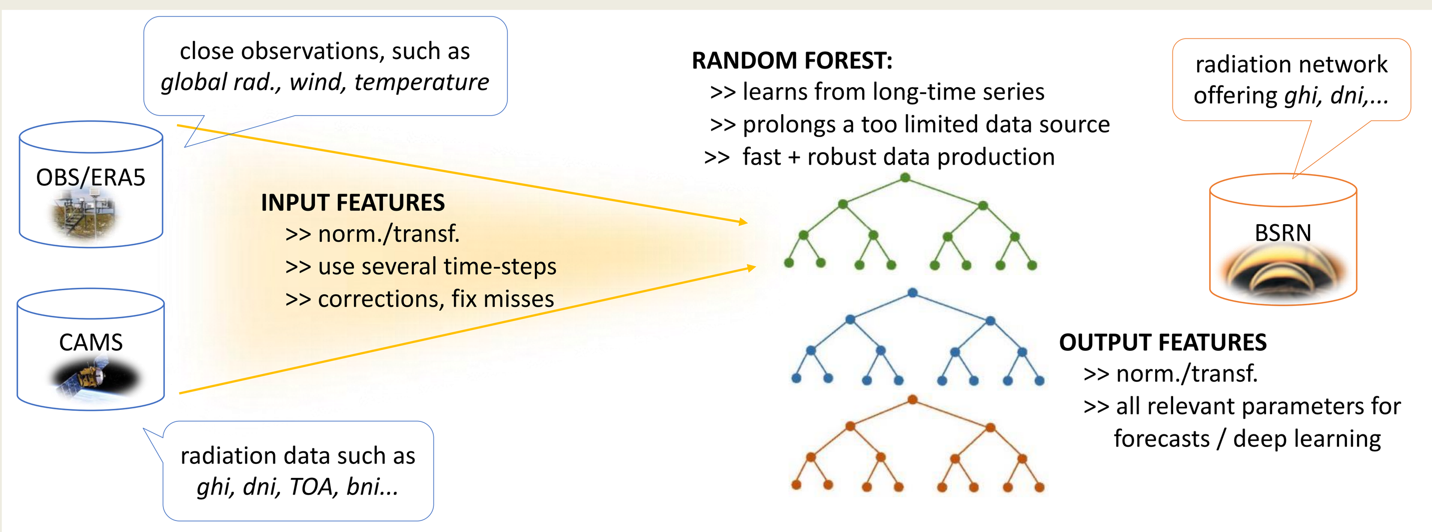
Workflow



Deep learning methods

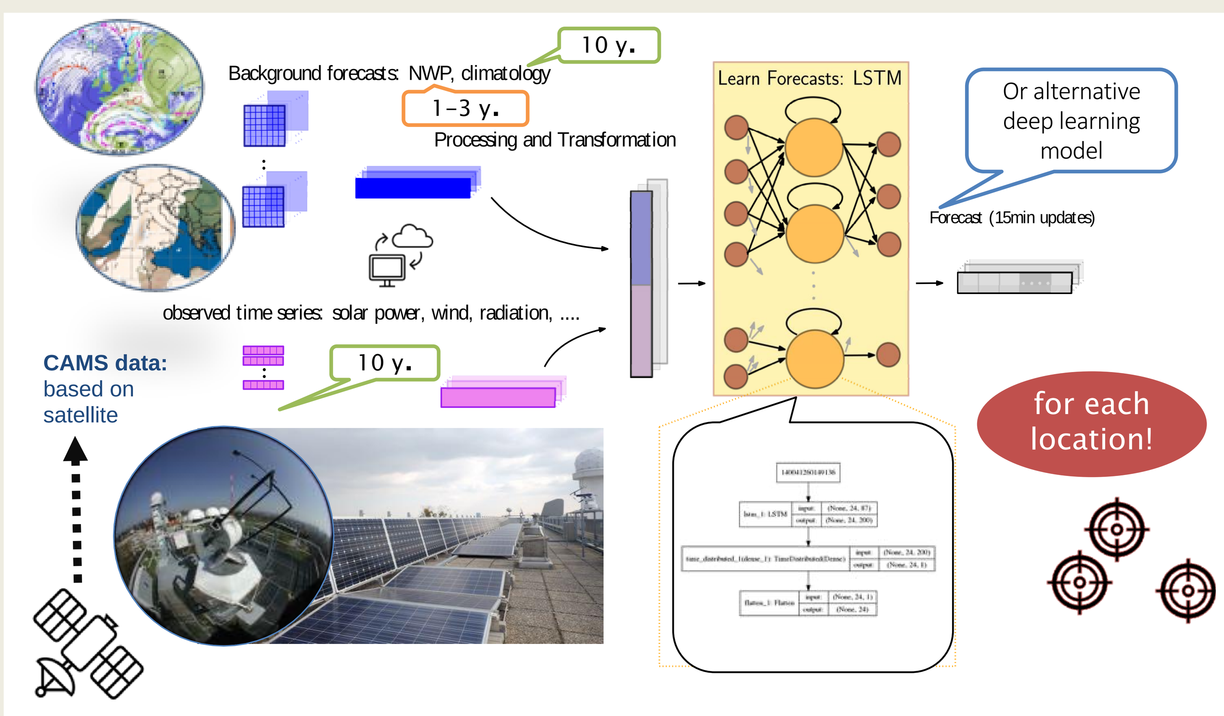
(i) Semi-synthetic Data Generator

As often observation time series are short/incomplete, and deep learning methods or any other statistical method need training data to fit a model for a specific task, semi-synthetic data are generated.
A random forest model using observation time-series of similar data sources is used to predict data of a limited data source. Needs some information of the PV farm/panel types.
Predicted parameters are: PV production and/or GHI (depending on site data). Used input data are: GHI, DNI, DHI, LWD, T2M, RelHum, pressure as well as model (AROME here) parameters.



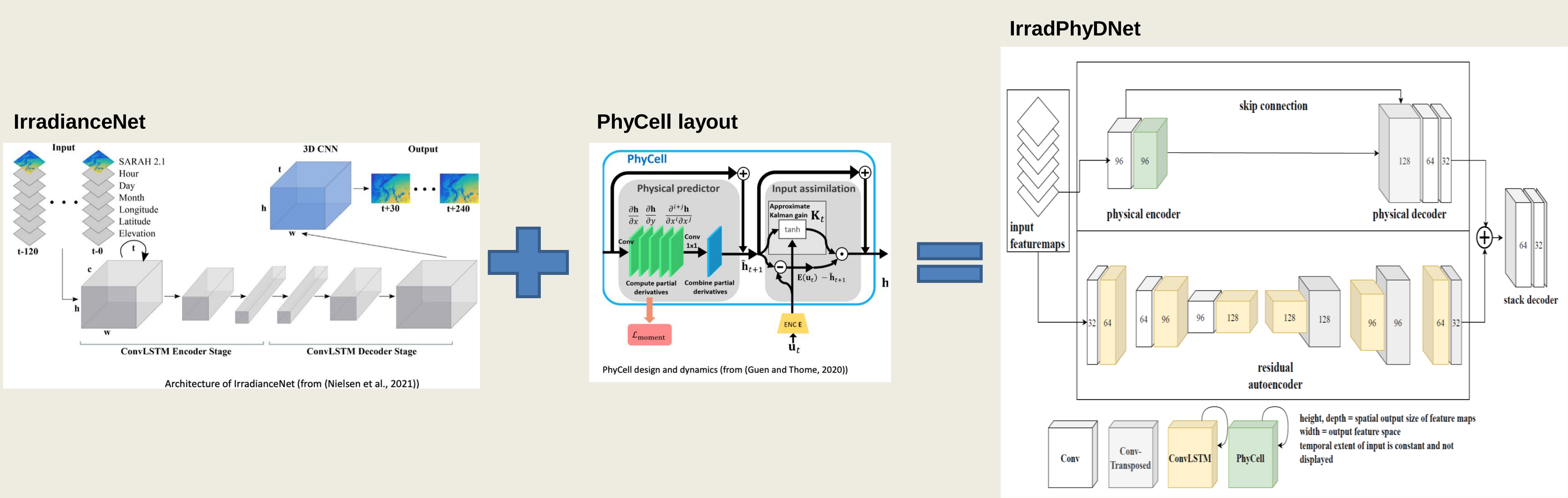
(ii) Location-based predictions

Using a sequence-to-sequence LSTM (long short-term memory) deep learning model, pre-trained with the semi-synthetic data and re-trained with in-situ data. Additional features include a background forecast model (LSTM, climatology) and historic time-series data (observations, model). Both nowcasting and day-ahead mode possible.



(iii) Spatial nowcasting with satellite data (SatNow)

Spatial, data-driven nowcasts using the LSA SAF DSSF_TOT parameter. A combination of two deep learning models (IrradianceNet, PhyDNet) including adaptations and robustness for missing time steps in both training and prediction (i.e. missing satellite data for an observation time step). Includes a ConvLSTM-autoencoder and physical disentangling concept by PhyCell design.
Currently implemented operational for Austria, adaptations to other DE_330 regions are ongoing.



Case studies

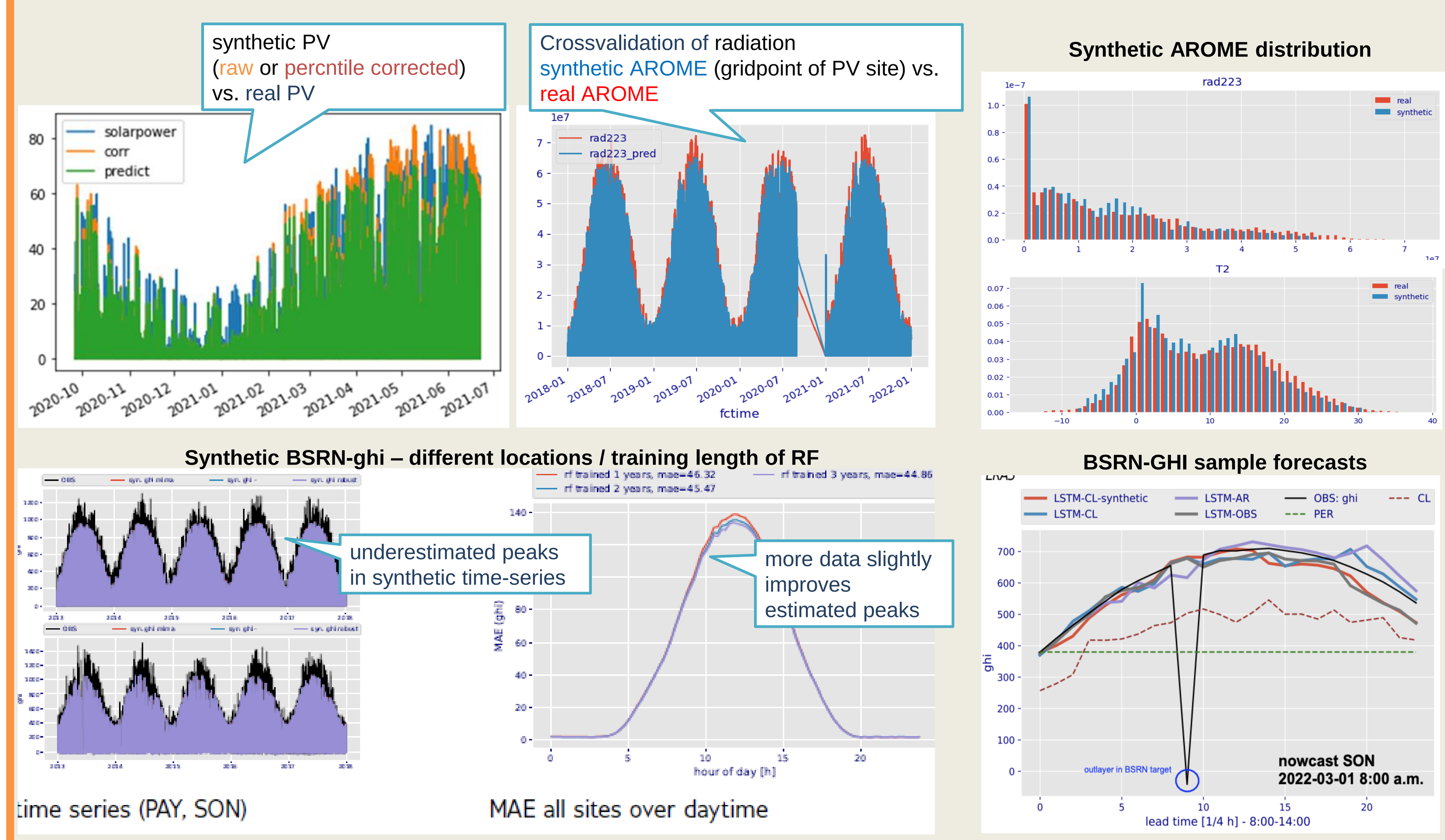
Case studies presented here include hourly issued nowcasts (2021/2022) for up to +6 hours ahead for location based nowcasts and up to +3 hours ahead for spatial nowcasts.

Location-based cases

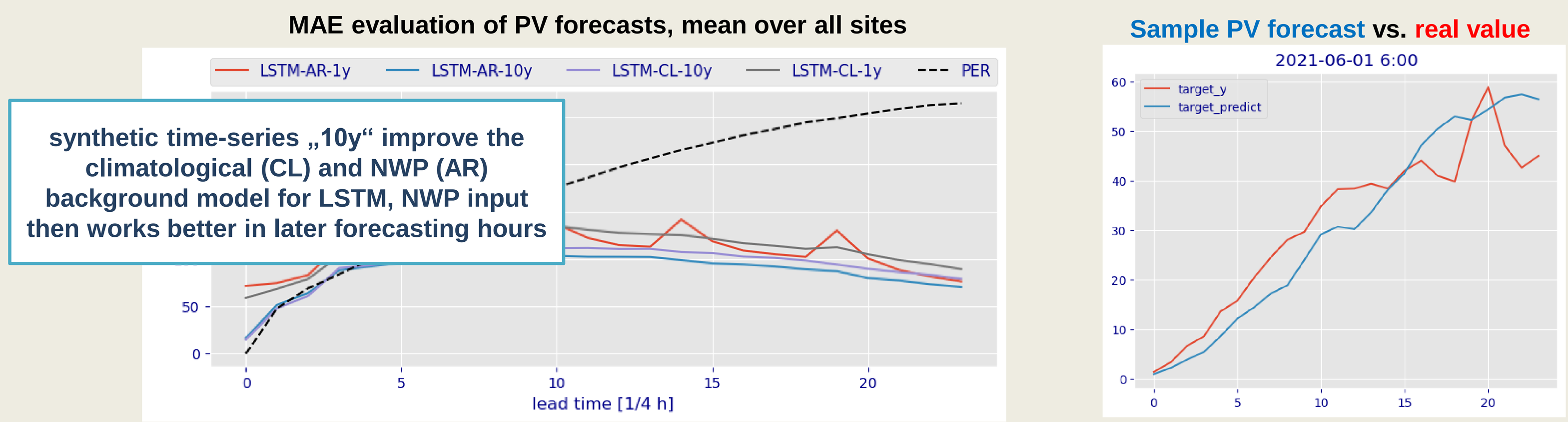
CASE STUDIES 2021/22

- 6:00-16:00 UTC in 2021/22
- +0:15 to +6:00 ahead (i.e., 15 min. Updates) frequency).
- Austrian + German PV sites, european BSRN, spatial forecasts
- Training: combination of 10 years real and synthetic data

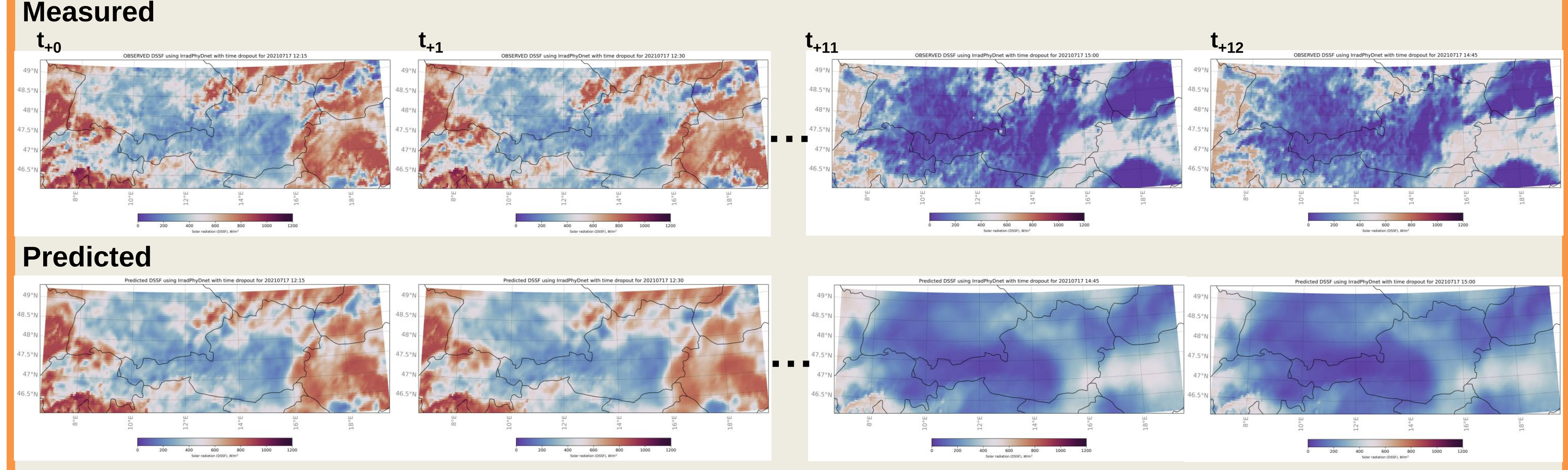
(i) Semi-synthetic data results

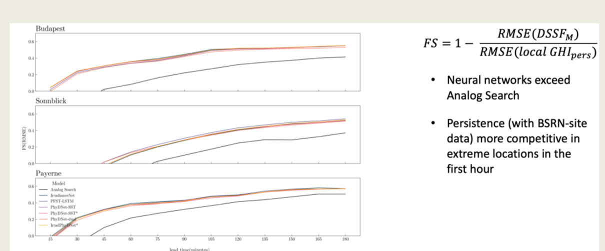


(ii) Location-based predictions



(iii) Spatial nowcasts 3-hours ahead, issued every 15 minutes (SatNow)





Summarizing

- Semi-synthetic data generator beneficial for deep learning
- Synthetic data necessary for no-data sites but might lack accuracy
- Semi-synthetic + location based predictions deliver good results
- In complex terrain semi-synthetic data generator very benecial for deep learning
- Random forest as a synthetic data generator best choice as fast, able to robustly process many parameters with short training horizons
- IrradPhyDNet very promising and very good results
- Under-representation of extremes/outliers needs more investigation

Next steps

- Implementation of IrradPhyDNet for Denmark/Netherlands domain
- Transfer learning approach with IrradPhyDNet using several different domains in training
- Combine location-based and IrradPhyDNet in prediction
- Implementation of transformer ensemble models for location based prediction
- User integration

Driemel, A.et al.: Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017), Earth Syst. Sci. Data, 10, 1491-1501, doi:10.5194/essd-10-1491-2018, 2018