RESEARCH REPORT

NOWCASTING

Solar PV Nowcasting Using Deep Learning

Can We Help the National Grid Electricity System Operator to Manage the Electricity Grid by Creating Better Nowcasts of Solar Electricity Generation?

OPEN CLIMATE FIX

Prepared by	Open Climate Fix Limited
Reference	NIA21-WP1-1
Version	v1.0.0
Part of	Network Innovation Allowance Project “Solar PV Nowcasting”
Date of Issue	December 31, 2021
Project Number	NIA2_NGESO002

Acronyms

API	Application Programming Interface
AWS	Amazon Web Services
CNN	Convolutional Neural Network
CEDA	Centre for Environmental Data Analysis
CPU	Central Processing Unit
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GAN	Generative Adversarial Network
GB	Gigabyte
GPU	Graphics Processing Unit
GSP	Grid Supply Point
GRU	Gated Recurrent Unit
HRV	High Resolution Visible
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MW	Megawatt
NASA	National Aeronautics and Space Administration
NG-ESO	National Grid Electricity System Operator
NMAE	Normalised Mean Absolute Error
NWP	Numerical Weather Prediction
OCF	Open Climate Fix
PV	Photovoltaic
PEF	Platform for Energy Forecasting
UKV	United Kingdom Variable (the UK Met Office's high-resolution deterministic model for the UK)
WP	Work Package (for example, "WP1" stands for "Work Package 1")

Table 1. Table of common acronyms used in this document.

Executive Summary

Our main task in Work Package 1 was to develop a novel solar photovoltaic (PV) forecasting model and to test how it performs compared to National Grid Electricity System Operator's (NG-ESO) existing PV forecasts.

At the end of Work Package 1, Open Climate Fix's (OCF) national solar generation forecast is 2.8 times better than NG-ESO's PV forecast (for forecasts up to two hours ahead). NG-ESO's existing national solar PV forecasts have a mean absolute error (MAE) of 650 MW. OCF's best national PV forecasts to date have a MAE of 233 MW and quantify uncertainty in its predictions, and we intend to keep working hard next year to further reduce the error.

This has been achieved by running over one thousand machine learning experiments during Work Package 1. Each experiment benefits from tens of terabytes of data, processed by a custom-built data pipeline, combined with cutting edge machine learning models based on papers released by Google Research and DeepMind over 2020 and 2021. Data inputs include 5-minutely satellite data, numerical weather predictions (NWP), and solar PV data from individual PV systems.

It is frequently said that machine learning is 90% data preparation. Through Work Package 1 we built the foundations in the data feeds and pipeline which will enable future machine learning research and a prototype implementation. We performed small-scale machine learning research through 2021. Once we built our data pipeline, this has allowed us to commence running machine learning experiments on the full-scale pre-prepared datasets since November. We have lots of ideas for how to continue to improve PV forecasting skill in 2022, and we are in a great place to do so!

Background and Related Work

Traditional solar PV forecasting methodologies have almost universally only used numerical weather predictions and PV readings to forecast solar PV output. OCF is looking to add satellite imagery to the input data set, and we briefly review the state-of-the-art in this new domain.

Google Research's "MetNet" paper (Sønderby *et al.,* 2020) describes a deep neural network designed to predict precipitation up to eight hours into the future. MetNet performs better than state-of-the-art numerical weather prediction models and produces probabilistic outputs. The model does not use any numerical weather predictions in its inputs: instead, it takes as input a high-resolution image of the area of interest, as well as a larger, lower resolution, 1,024 km x 1,024 km context image that is large enough to capture any clouds that might cross the area of interest, from both precipitation radar and satellite imagery. Topographic maps are also included as extra channels in these images. This is then passed through a convolutional LSTM model, followed by a few axial attention layers to allow the model to learn what parts of the images are the most important.

On 15 November 2021, Google Research announced MetNet-2, which extends the prediction horizon to 12 hours by using input data with a significantly larger spatial extent (2,048 km x 2,048 km). This is a huge model and runs across 128 Google Cloud TPU v3 cores. To get the model to fit, the authors abandoned the axial-attention module that was present in MetNet-1.

Google DeepMind has also been working on precipitation nowcasting. In 2021, DeepMind released their "skilful precipitation nowcasting" paper (Ravuri *et al.,* 2021) which uses a generative adversarial network (GAN) to create realistic-looking precipitation nowcasts. Like the MetNet papers, the "skilful nowcasting" model does not look at numerical weather predictions. This model takes as input the last 20 minutes of imagery and outputs the next hour and a half of future imagery. It does this by creating four context stacks of images at different spatial resolutions, and passing them through a set of convolutional gated recurrent unit (GRU) layers that combine the context stacks at each level, starting from the smallest set of images, and working up to the largest images. These convolutional GRU layers are used to predict each timestep one at a time. As part of this process, a random vector is drawn from a uniform distribution and used as the initial hidden state of the bottom GRU layer, ensuring that each prediction is slightly different even with the same inputs, creating its probabilistic forecasts. Unlike the MetNet papers, the DeepMind authors put a lot of thought into figuring out if expert human weather forecasters "believe" the GAN's forecasts. The conclusion is that, yes, the DeepMind GAN produces predictions which humans find very believable.

OCF has implemented both MetNet and the model in the skilful nowcasting paper. Please see the "Architecture of OCF's ML models" section for more details.

In the second half of 2021, DeepMind released two related papers, which have been highly influential on OCF's approach to PV nowcasting: The "Perceiver" paper (Jaegle et al., 2021), and the "Perceiver IO" paper (Jaegle et al., 2021b, and Figure 1).

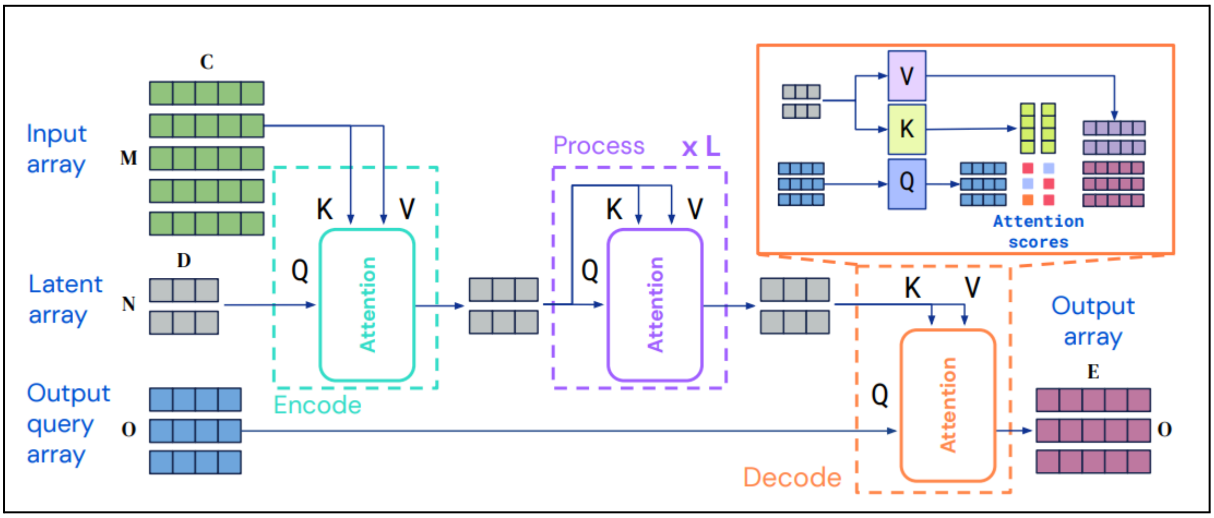

Figure 1. The Perceiver IO architecture. Taken from Jaegle et al., 2021b.

The Perceiver is based on self-attention, which has proven to be an extremely powerful model for many domains. Recent breakthroughs such as OpenAI's GPT-3 and DeepMind's AlphaFold-2 use self-attention extensively.

Conventional self-attention models suffer because their computational complexity goes up with the square of the length of the input (in this case the length of the input is linked to the number of pixels in the satellite and/or NWP data). This makes them intractable for working on inputs such as large images. The Perceiver introduces a beautifully simple way to limit the computational complexity of self-attention, and hence allow these models to be applied to images and even videos.

We are particularly excited about the Perceiver because it excels at "multimodal" tasks. That is, it excels at being able to take multiple different types of input (satellite imagery, numerical weather predictions, etc.), which may be on different spatial or time grids, and combine them; and also to perform multiple tasks (such as predicting GSP-level PV; or PV for single PV systems; or predicting future satellite imagery). Even better, the input modalities do not need to be perfectly aligned in space and time. So, for example, we can take hourly numerical weather predictions on a 2 km grid and 5-minutely satellite images on a 2-6 km grid and input them natively into the model without preprocessing or interpolating the data.

Methodology

Overview

OCF's data pipeline combines data from many sources including satellite imagery, numerical weather predictions, and PV data from individual PV systems. In total, tens of terabytes of data have been collected.

The data is pre-processed into many thousands of "training examples" for our machine learning models.

Our machine learning models are trained to predict the half-hourly total PV generation for the next two hours, for a single GSP region at a time.

Data

Data Sources

So far, we have downloaded and processed the following sources of data:

Satellite Imagery

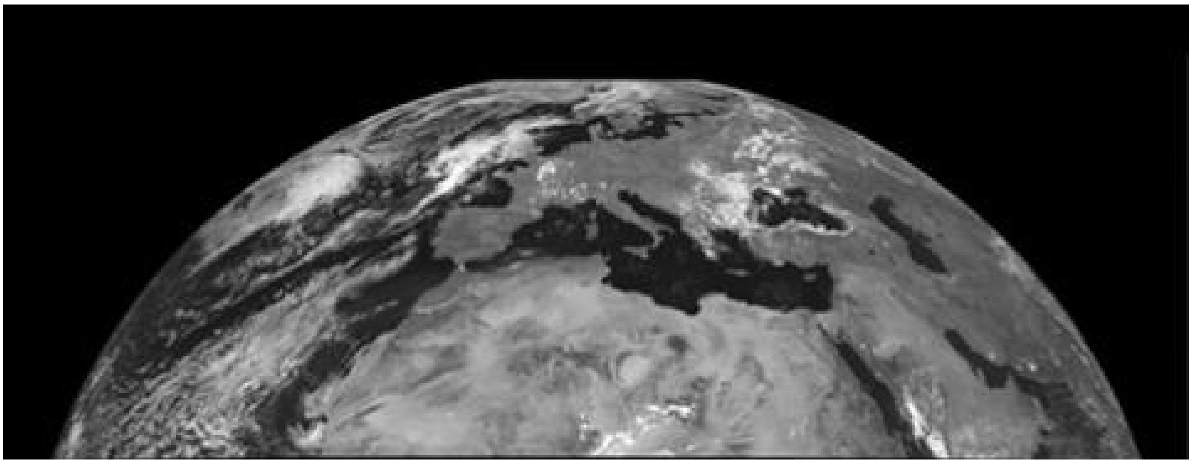

Figure 2. Example EUMETSAT SEVIRI RSS HRV image.

Satellite imagery from EUMETSAT's Spinning Enhanced Visible and InfraRed Imager (SEVIRI) Rapid Scanning Service (RSS). This satellite instrument collects data across 12 spectral channels every five minutes from geostationary orbit. One of these channels (the "high resolution visible" (HRV) channel) has twice the spatial resolution of the other channels. The spatial resolution of the HRV channel is about 2 km to 3 km. The spatial resolution of the other channels is about 4 km to 6 km (the spatial resolution decreases from south to north, as can be seen in Figure 2). Real-time data is available with a latency of just a few minutes.

Photovoltaic Generation Data

For model training and prediction, we use Solar PV data from individual solar PV systems from Passiv Systems and PVOutput.org. The Passiv Systems dataset includes data from a total of about 25,000 PV systems, of which about 1,000 provide near-real-time data at a temporal resolution of two minutes. We have focused our experiments on data from these 1,000 "2-minutely" PassivSystems PV systems. The data was provided to OCF by Sheffield Solar (with agreement from PassivSystems).

Figure 3 shows the location of these Passiv PV systems.

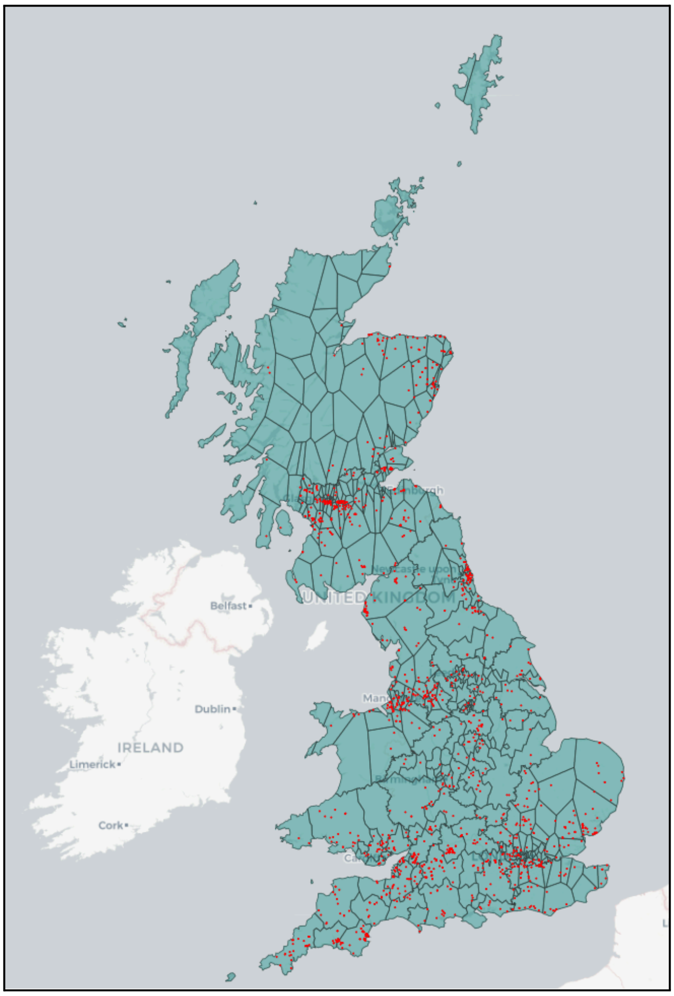

Figure 3. Map of UK showing GSP regions (green) and numerous PV systems (red dots).

For assessing model accuracy at the national and GSP level, we use the outturns from Sheffield Solar's PVLive service. This service models the PV outturn in Great Britain and is used by NG-ESO as the official PV data. It is the closest we have to ground source truth for these regional values. We use the latest PVLive values available, which may include corrections. It should be noted that these corrected values are the most accurate values available, but may be different from those outturn numbers available to NG-ESO intra-day.

Installed PV capacity numbers are useful for normalising results to produce values comparable across regions. We use the installed capacity values produced by Sheffield Solar. Again, these are the official values used by NG-ESO.

Numerical Weather Predictions

Numerical weather predictions from the UK Met Office's "UKV" model, the UK Met Office's high-resolution deterministic model for the UK. We use the raw, "gridded" NWPs, which provide predictions at a horizontal spatial resolution of 2 km. These provide hourly predictions, updated eight times per day.

Grid Supply Point Power Generation

Estimates for total solar PV power generation per GSP provided by Sheffield Solar's PV Live Regional API. The actual solar generation is unknown but these are the best estimates available for the UK. Note that the intra-day PV Live estimate (computed using the ~1,000 PV systems which report in near-real-time) is less accurate than the "day-behind" update (computed using the ~25,000 PV systems which report once per day). Only the "day-behind" data is archived and therefore we do not feed GSP Live data into our models as an input. This is because the "updated" data we have access to is of higher quality than the intra-day data that would be available at the time of doing the forecasts. Figure 3 shows the locations of the GSP regions.

Sun

The position of the Sun with the elevation angle and azimuth computed using PVLib.

Topological

Elevation maps from NASA's Shuttle Radar Topography Mission are available for the whole world at a resolution of roughly 30 meters. We downsample the topographic maps to a 2 km resolution.

Date Range

Data is available from all these data sources back to at least 2016. For Work Package 1, we focused on data from 2020 and 2021. In Work Package 2, we will extend our dataset back to 2016. We have used data from 2020 for training and validation in our ML models and data from 2021 is used for model evaluation.

Data Preparation

The cliché is that data preparation is often more than 90% of the engineering work when doing machine learning research and that was certainly true for Work Package 1.

To train our machine learning models, we need all the data sources to be:

- Cropped into "training examples" covering similar periods and geospatial areas
- Cleaned of obvious errors
- Pre-prepared into "batches" of data suitable for ML training. When we train our models, we need to load about 2.5 GB (about the same amount of data as one hour of a high-definition movie) per second off disk, so the data must be in a structure that is very fast to read from the disk.

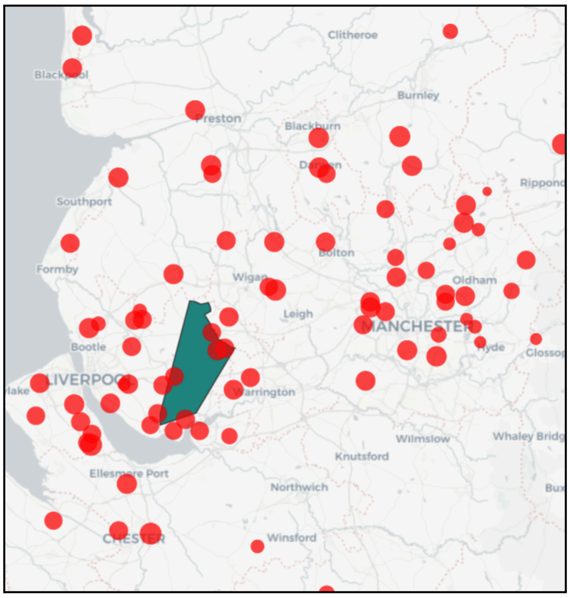

Figure 4. Map of data showing one GSP region (green) and PV systems (red) for one training example.

The majority of our software engineering effort throughout Work Package 1 has gone into engineering a flexible and fast data pipeline that can process tens of terabytes (TB) of data from multiple data sources. The input to this pipeline is the raw data. The output is thousands of pre-prepared "batches" of machine learning data, ready to be fed into our models (one example from a batch is shown in Figure 4). To prepare the data in the shortest possible time frame, we spent a lot of effort ensuring the code can use all CPU cores at once, and concurrently from each data source.

One of the great benefits of ML models based on self-attention is that the data sources do not need to be perfectly aligned on the exact same "grid". As such, our data pipeline does not spatially reproject data, which helps the model accuracy because reprojection introduces artefacts.

Lessons Learnt

At the start of Work Package 1, we tried to load the raw data on the fly during ML training. Unfortunately, this was just not fast enough. So we re-wrote our data pipeline to save pre-prepared ML batches to disk ahead of time, which took a few days in data preparation and then trained our models on the pre-prepared batches.

Pipeline Components

The data pipeline is made up of several components (all of which are openly available on GitHub):

- nowcasting_dataset: This is the main component of the data pipeline. This consumes data from intermediate "OCF formats" and outputs pre-prepared batches.
- satip: Downloads satellite imagery from the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and converts them into an intermediate Zarr format.
- nwp: Downloads numerical weather predictions from the Centre for Environmental Data Analysis (CEDA) and converts them into an intermediate Zarr format.

Each "batch" contains 32 training examples. Each example is specified by a t_0 (the time "now": the time of the most recent observation); and the geospatial location of the centre of the region of interest. The "training target" is the half-hourly GSP PV yield. During pre-processing, different amounts of "history" and "future" are selected for each data source (relative to t_0), which is shown in Table 2.

	size of the region of interest	history length (minutes)	future length (minutes)	number of channels
Satellite	24 x 24 pixels (~96 x 96 km)	30		11
HRV satellite	64 x 64 pixels (~128 x 128 km)	30		1
PV	256 x 256 km	30		up to 128 PV systems per example
NWP	64 x 64 pixels (~128 x 128 km)	60	180	10
GSP-level PV	1 GSP		120	1

Table 2. Table to show the size of different data sources we have considered.

The data pipeline is illustrated in Figure 5.

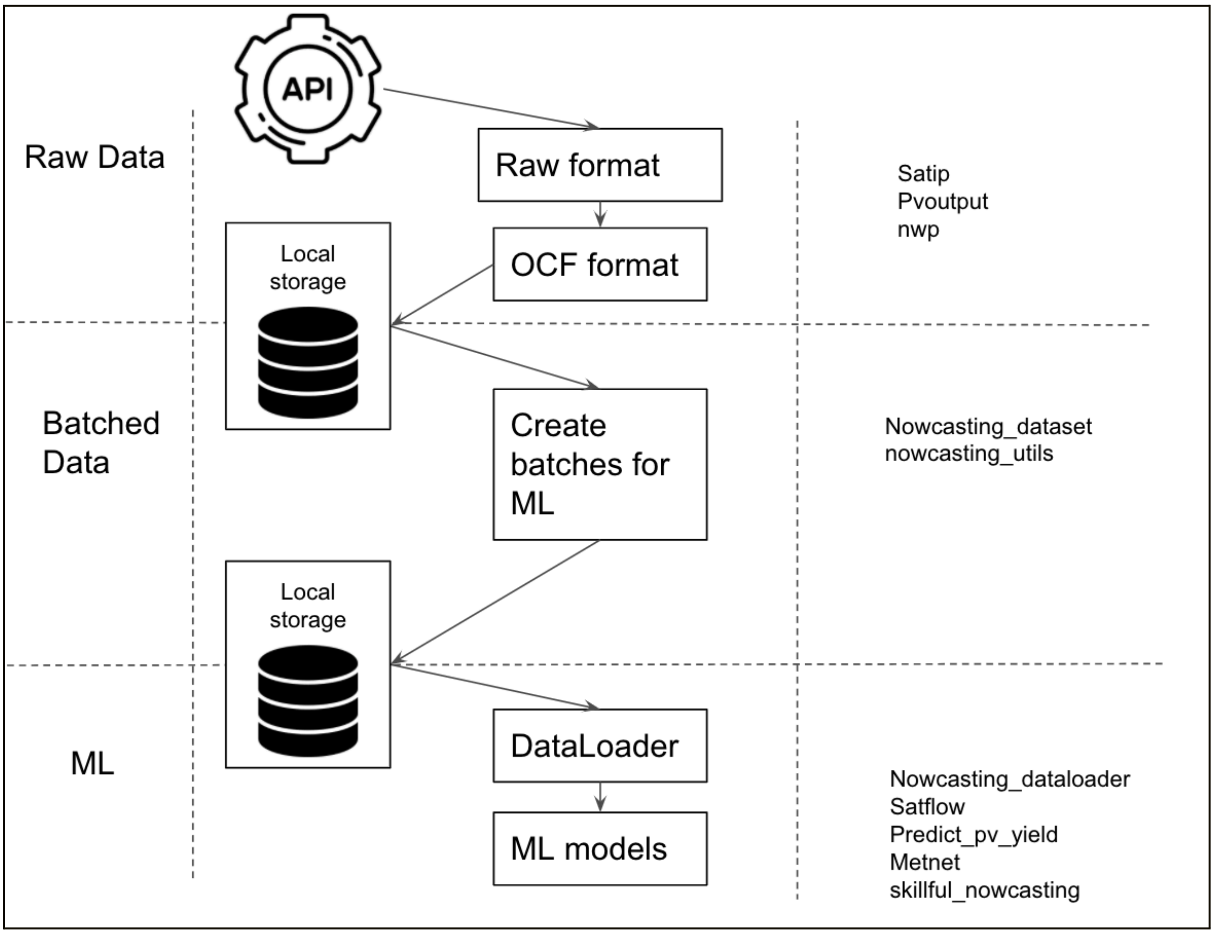

Figure 5. Illustration of data pipeline from the raw API call to ML models.

Data Validation

We have used a variety of methods to ensure that the data is correct. First, all data sources are visualised and checked by domain experiments (see Figures 6, 7 and 8). Second, data sources are visualised together to confirm system behaviour, for example when a cloud causes a dip in solar generation. Third, automatic checks are used to check for data quality, for example, to check there is no missing data or to check that there are no invalid values.

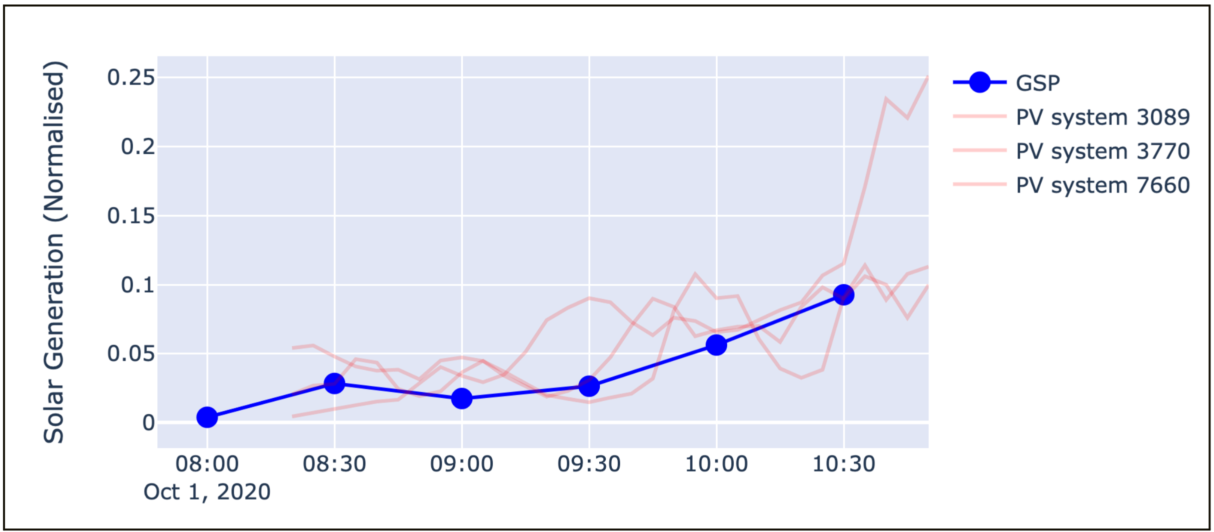

Figure 6. This figure shows the correlation between GSP solar generation (blue) and nearby PV systems (red).

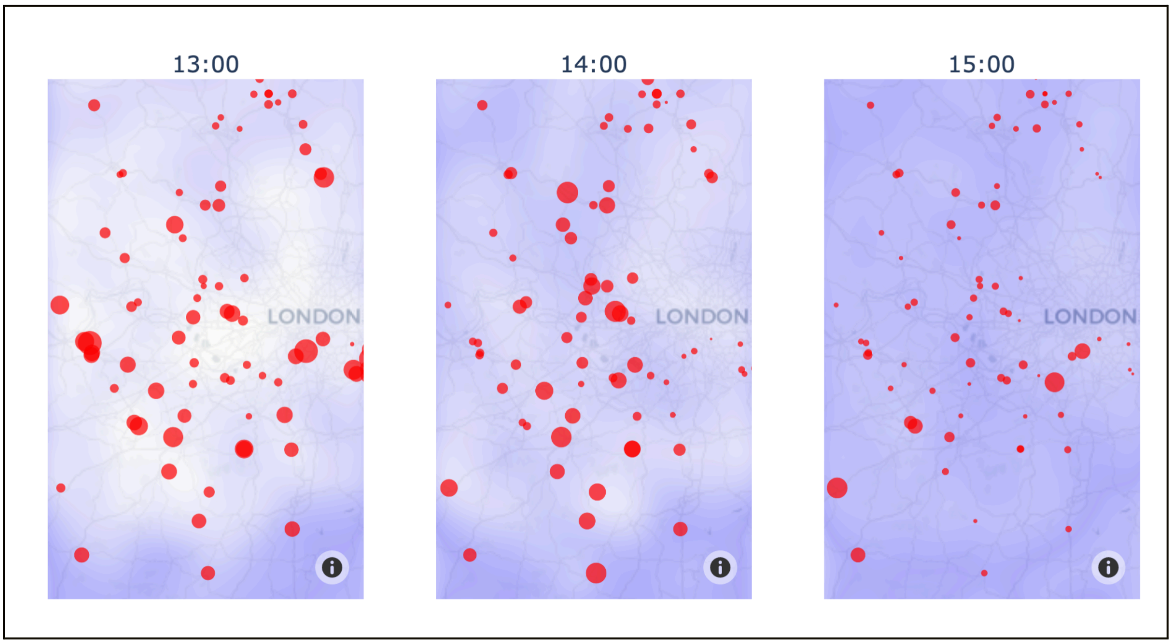

Figure 7. Three maps around London, on 15th October 2020, showing satellite data in blue and PV solar generation in red. The magnitude of the PV solar generation is proportional to the size of the red dots. We can see that the PV solar generation decreases as clouds develop over time.

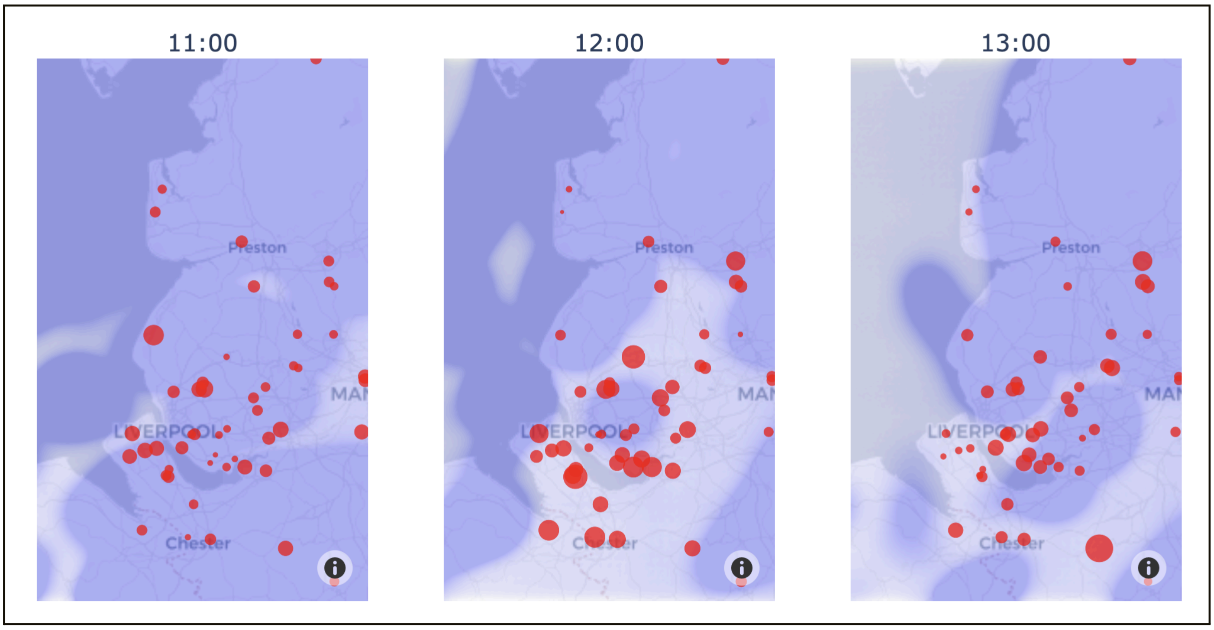

Figure 8. Three maps around Liverpool, on 27th July 2020, showing NWP precipitation data in blue and PV solar generation in red. The magnitude of the PV solar generation is proportional to the size of the red dots. We can see that solar PV generation increases as the rain clears over time.

Data Sizes

Table 3 shows the size of the different data sources we have used.

Data Source	Raw [GB]	Processed [GB]
Satellite	14,000	395.0
NWP	21,000	653.0
Individual PV systems	~	0.72
GSP-level PVLive	~	0.12

Table 3. Showing the size of each data source, divided up into raw and processed data.

The Architecture of OCF's Machine Learning Models

We have implemented six main models, which are described below. For each model, we have conducted hundreds of experiments so far and will conduct hundreds more in 2022.

OptiFlow

The OptiFlow model (see Figure 9) uses optical flow to predict satellite images, across all satellite channels, for the next two hours, at half-hourly temporal resolution. Each optical flow prediction is fed into a Perceiver model (Jaegle *et al.* 2021). The Perceiver sees a single timestep of optical flow imagery at a time. The output of the Perceiver is fed into a fully connected network which, in turn, outputs the predicted PV yield for the target GSP for that timestep. The output of the network is a mixture density network that parameterises a mixture of Gaussian distribution (with two Gaussians). This model also receives four hours of NWPs and the last half an hour of PV yield from individual PV systems in the region of interest, for up to 128 PV systems. Temporal and spatial encodings are concatenated to all elements. The geospatial encoding includes an embedding of the GSP ID (for example, each element of PV data includes an embedding of the GSP region ID in which that PV system exists in the real world). The query into the Perceiver is mostly learnt, but also includes an embedding of the GSP ID and the Fourier features of the target DateTime.

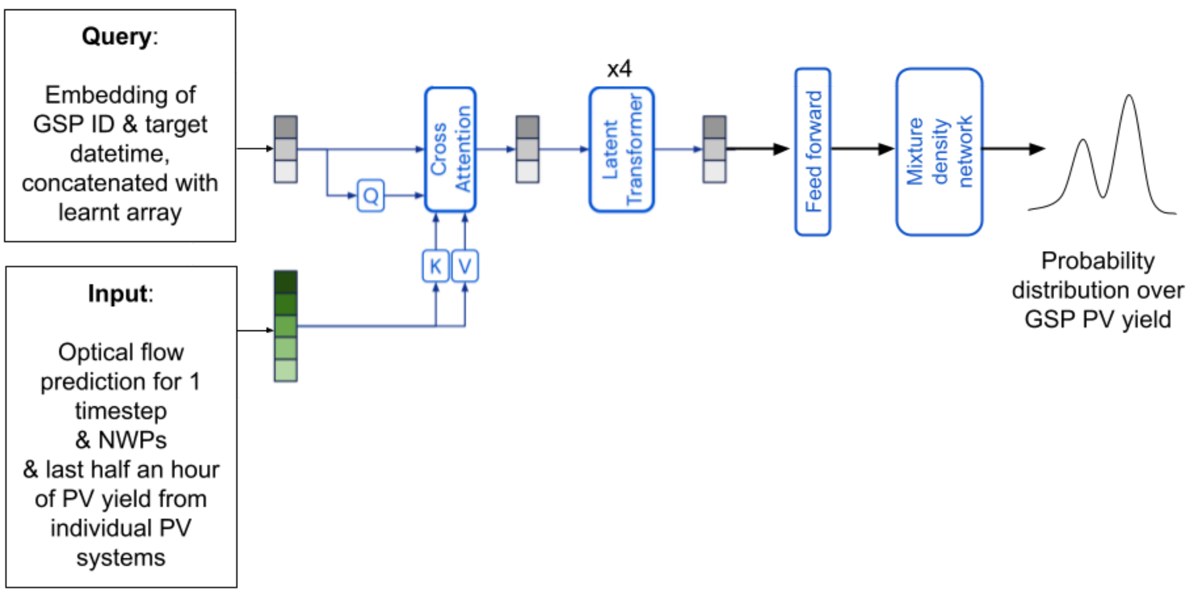

Figure 9. OptiFlow architecture (adapted from Jaegle et al. 2021 with OCF's modifications).

Conv3d

This model takes both satellite and NWP video data and puts them through separate 3d convolution networks. These are then connected with a few fully connected layers, joined with some simple input data like historic PV data (see Figure 10).

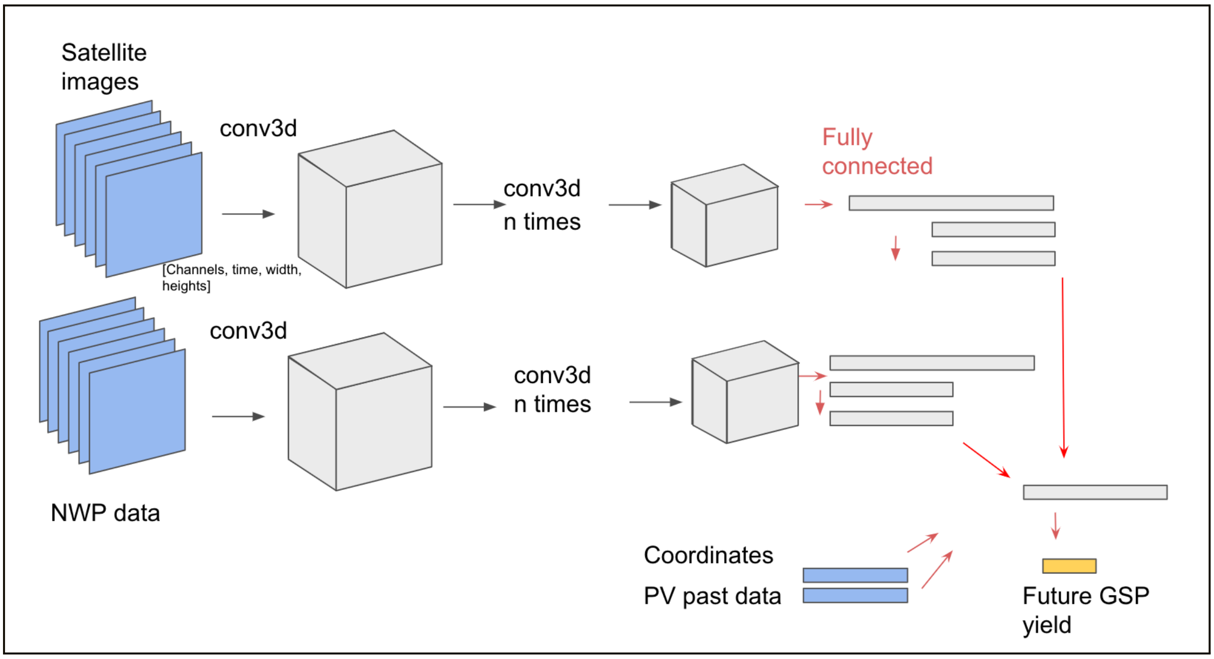

Figure 10. Illustration of convolution 3D network. Satellite and NWP are passed through several convolution neural networks layers and then connected with some fully connected layers. Additional inputs like the PV past data are then added to finally predict the GSP level solar generation.

Perceiver

Both Satellite and NWP are first fed through some 3d convolution layers and then into the Perceiver network. The network is sensitive to Satellite and NWP being the same size, which is not always the case. The Perceiver network model is based on a DeepMind model that works by encoding multiple different types of inputs into a latent space, and then running a self-attention model over that latent space to create an output vector (please see the "Background" section for more information).

PerceiverIO

This model is based on a DeepMind model that works by encoding multiple different types of inputs into a latent space and then running a self-attention model over that latent space to create an output vector. That vector can then be queried with different queries for different outputs, such as future satellite imagery, and GSP power. This means that we can provide the model with lots of different inputs, such as satellite imagery, NWPs, PV historical data, topographical maps, and more without having to change the data from their native formats and introducing artefacts. The model learns where the different data is in time and space through a position encoding that is common across the geographical and temporal extent of the training data, allowing it to learn how the different input data relates to each other, and the future GSP solar generation yield.

MetNet

We modified the original model architecture to support more satellite channels, and to remove the precipitation radar images. So far, we have used this model to attempt to predict future satellite imagery, but not future GSP yield yet.

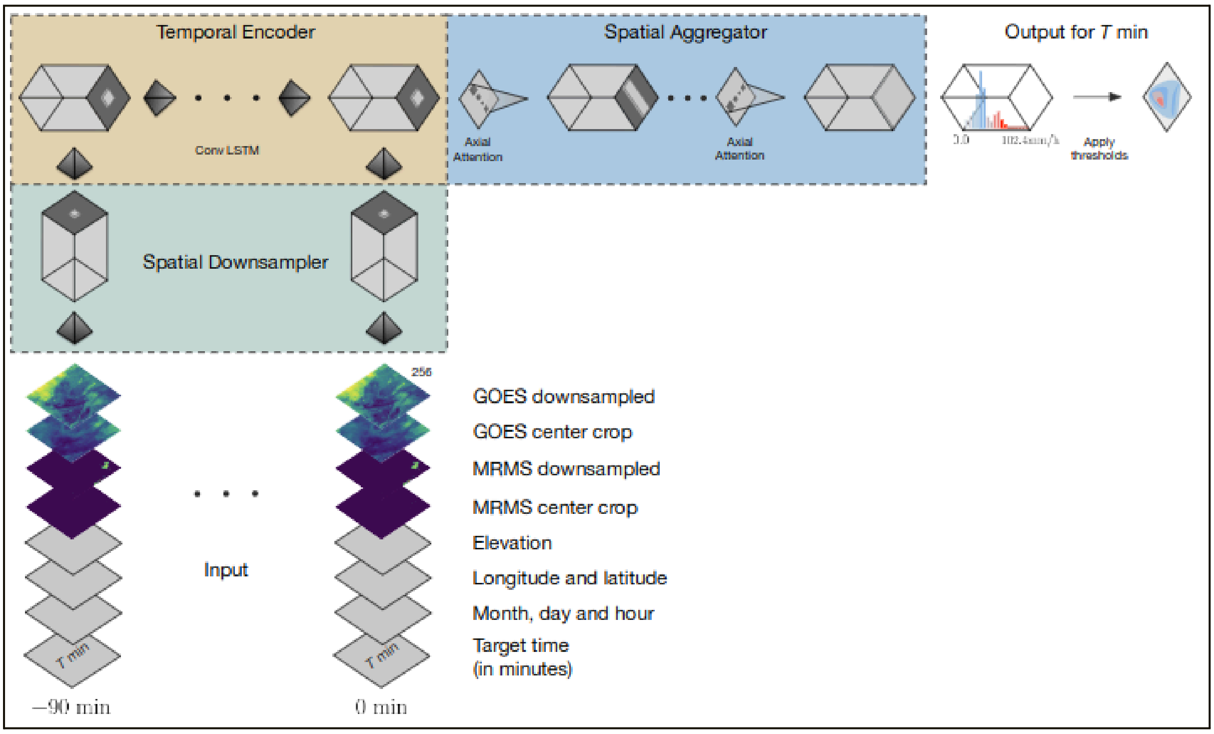

Figure 11. MetNet model diagram, showing the different inputs. For our research, the GOES images are replaced with EUMETSAT SEVIRI RSS images, and the MRMS images are not included. The figure is taken from [this paper](#).

Skilful Nowcasting GAN (Deep Generative Model of Radar)

For our purposes, the original model was modified to take multi-channel images and be able to output multi-channel images for predicting satellite imagery. We have not tried using this model to predict future GSP yield yet, only future satellite imagery.

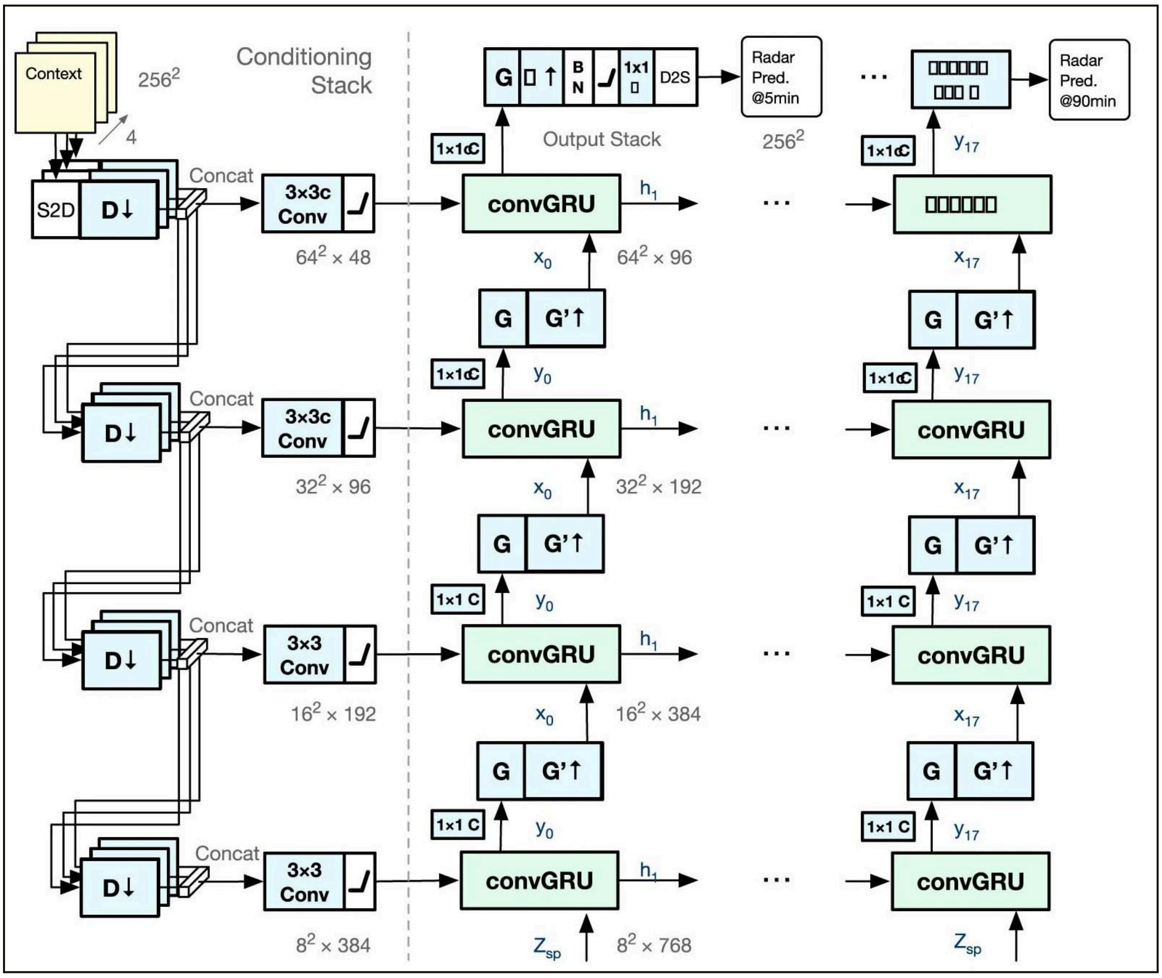

Figure 12. Diagram of the generator in the Skillful Nowcasting GAN, showing the context images, convolutional GRU layers, and outputs. The figure is taken from [this paper](#).

How Our Machine Learning Models Are Trained

Our models are trained on fast graphics processing units (GPUs). OCF has six NVIDIA [RTX-A6000](#) GPUs on-premise, and also uses GPUs in Amazon Web Services (AWS) and Google Cloud. Each model takes up to 20 hours to train on a single GPU. We use [PyTorch Lightning](#) to train our models.

Modelling Results

Introduction

We have conducted numerous experiments with different models and in this section, we will present the results.

Evaluation Data

The results are based on an evaluation dataset that is taken from the 1st of January, 2021 to the 1st of September, 2021. This data was not given to the ML models during training. 1,000 random timestamps were chosen at all the GSP locations, allowing us to evaluate the model's national solar forecasts. The total number of data points in the test set is approximately 300,000.

Metrics

To evaluate the models we will use a variety of metrics, but for this report, we will focus on MAE - Mean Absolute Error, measured in units of MW. MAE is calculated by taking the absolute difference between the forecast value and the actual PV outturn value (for the GSP, national or site-level figures). If a model has a MAE of 7.6, then the average absolute error is 7.6 MW.

It is often useful to consider normalised data to compare areas with different installed PV capacities. This is called Normalised Mean Absolute Error (NMAE). For the National Forecast, we simply normalise by the installed capacity at that time. For GSP results we normalise the individual GSP results by the installed capacity of that GSP and then average NMAE across GSPs. This is done so that the GSPs with the largest installed capacity do not dominate the overall results.

Equations for MAE and NMAE can be found in the appendix.

Example of Predictions

As noted before, the models are trying to predict solar PV generation for an entire GSP region. Figure 13 shows an example of the predictions and the truth values. There is a good variety across the different examples of small forecasting errors and large ones.

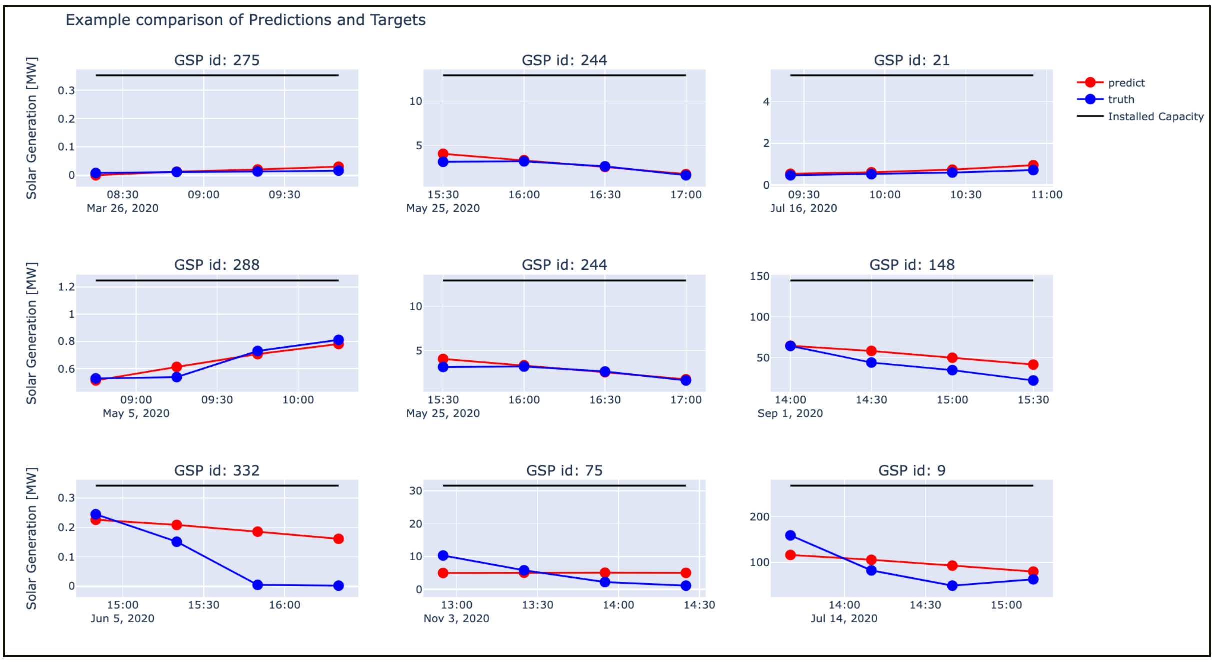

Figure 13. Nine plots comparing predictions and targets of GSP solar generation. The predictions in the top three have little errors, the middle three have medium errors and the bottom three have large errors.

Model Results

Table 4 summarises the results for the different models. The results are both on a National and GSP level. This allows us to easily compare the baseline model, NG-ESO results, and OCF results.

We can see that the four different OCF models are significantly outperforming the NG-ESO model. Of the four different OCF models, OptiFlow has achieved the best National MAE results of 232.6 MW.

Model Name	National		GSP
	Mean Absolute Error [MW]	Normalised Mean Absolute Error [%]	Normalised Mean Absolute Error [%]
Baseline (yesterday)	1135.5	8.68	14.82
NG-ESO	649.8	4.98	9.96
OCF PerceiverIO	308.2	2.36	7.07
OCF Perceiver	296.3	2.29	7.01
OCF CNN	276.4	2.12	7.03
OCF OptiFlow	**232.6**	**1.78**	**6.34**

Table 4. For several different models, the national and GSP level summary metrics are shown. Both MAE and NMAE are shown. Note the baseline results are made by using the outcome solar generation outcome from one day before.

Comparison with NG-ESO's PV Forecasts

By comparing the OCF models with NG-ESO we can get an impression of the improvement offered by the OCF models.

Looking at Table 4, we can see that the OCF OptiFlow model's MAE is 36% of NG-ESO's PV forecast. It is also interesting to look at the GSP NMAE error and compare the OCF models with NG-ESO forecasts. Figure 14 shows the OCF forecast has a lower mode and a shorter distribution tail compared to the NG-ESO forecasts.

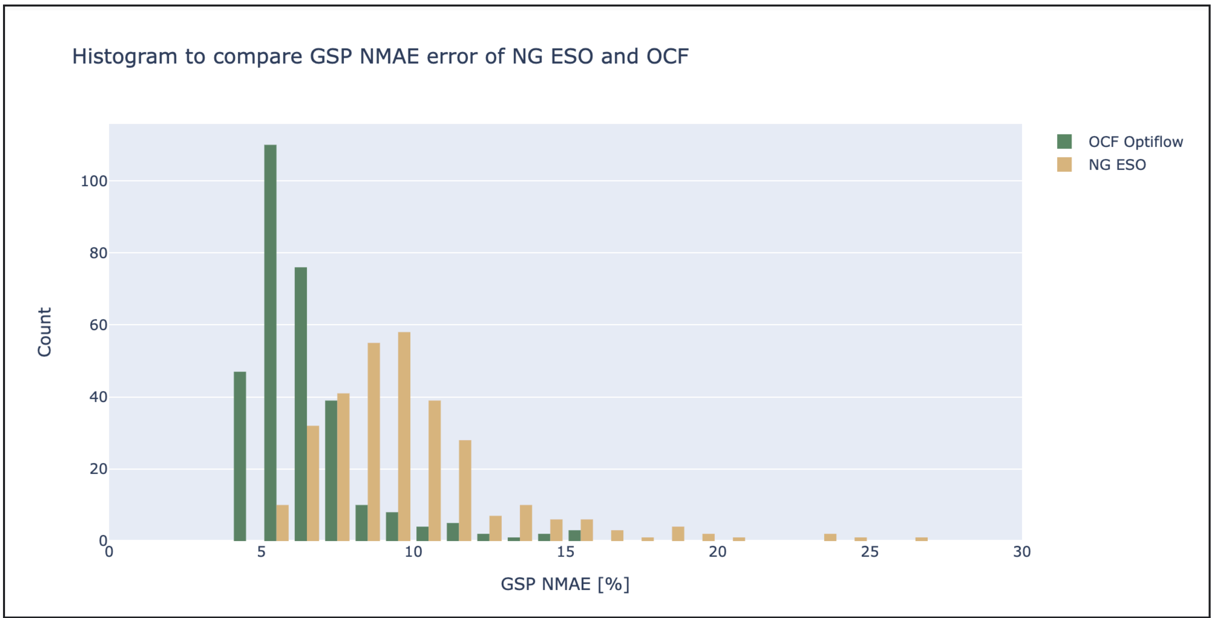

Figure 14. Two histograms to compare the distribution of GSP NMAE errors. OCF's Optiflow model is shown in green and the NG-ESO model is shown in yellow.

The takeaway result is that the best OCF model has a National Forecast MAE of 233 MW compared to 650 MW from the NG-ESO results. This is already a significant improvement at this early stage in the project.

Probabilistic Predictions

Our OptiFlow model quantifies the uncertainty of its predictions of PV power generation by outputting a probability distribution over PV power generation. Figure 15 shows two predictions:

The top row shows data from the 3rd of June, 2021 when the UK was covered in intermittent cloud cover. In this weather scenario, solar PV power production bounces up and down rapidly as the small clouds move overhead. This can be seen in the "rats nest" of thin grey lines on the left of the top time series plot, where each thin grey line shows the actual power generation from a single PV system. The ML model knows that it is hard to predict PV generation accurately when there are many small clouds, so it correctly produces a dispersed probability distribution (the probability distribution is illustrated by the grey "blur" towards the right of the time series plot).

In contrast, the bottom row of Figure 15 shows PV predictions when the UK was covered in multiple layers of dense clouds on the 20th of May, 2021. In this weather scenario, the model understands that it can confidently predict that there will be very little solar PV generation (the probability distribution is very "sharp")!

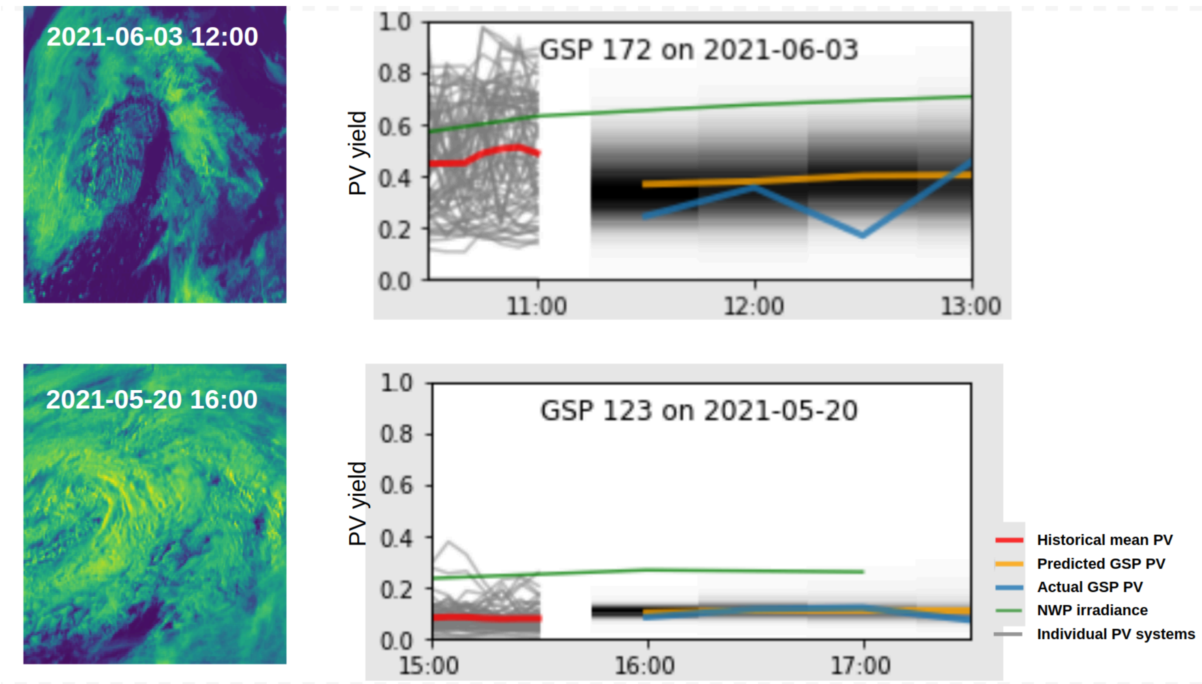

Figure 15. Probabilistic predictions of GSP PV yield. The left column shows two satellite images over the UK. The right column shows two time-series plots. The top row represents 2021-06-03. The bottom row represents 2021-05-20. The "rats nest" of grey lines on the left of the time series plots shows the last half hour of actual PV power from individual PV systems. The grey "blur" on the time-series plot shows the probability distribution of the predictions.

Ablation Results

An ablation study was conducted using the OptiFlow model. This involves systematically adding data sources to the model training and exploring the prediction results of including these data sources. This shows us which data types provide the most predictive information to the model. Table 5 summarises these results based on the input types as explained in the model description above. The results show that using PV and Optical flow data significantly improve the prediction accuracy, validating the value of these inputs.

Experiment	NWP	PV	Optical Flow	Normalised Mean Absolute Error [%]
Experiment 1	✓	x	x	9.27
Experiment 2	✓	✓	x	7.74
Experiment 3	✓	✓	✓	6.34

Table 5. Results from Ablation study using OptiFlow model. Normalised Mean Absolute Error (NMAE) starts at 9.27% with just NWP and decreases to 6.34% when all data sources are included.

Forecast Horizons

It is interesting to look at the results across different forecast horizons. The different forecast horizons we have looked at are 30 minutes, 60 minutes, 90 minutes and 120 minutes.

The results in Figure 16 are typical across all the models and shows the NMAE is approximately 2% for a thirty-minute horizon and the NMAE is 2.7% for a two-hour horizon.

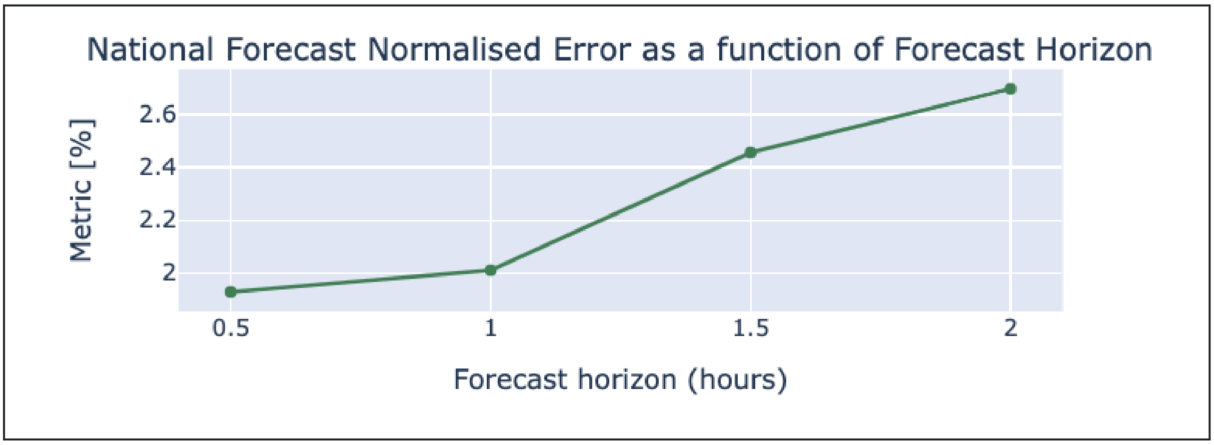

Figure 16. Plot to show Normalised Mean Absolute Error as a function of the forecast horizon, for a national forecast. This is for the OCF CNN model.

Future Work

Throughout Work Package 1, we laid solid foundations and have demonstrated in a research setting that our ML approach performs well compared to NG-ESO's historical PV forecasts. We are excited to be diving into Work Package 2.

From a technical perspective, Work Package 2 will consist of two main themes:
1. Build an operational, prototype solar PV nowcasting service satisfying the requirements of the NG-ESO control room.
2. Continue to research how to improve the skill of the solar PV nowcasts.

Build an Operational, Prototype PV Nowcasting Service for NG-ESO

By the end of Work Package 2, we will have delivered a prototype operational service to the NG-ESO control room. Solar PV nowcasts will be delivered through an API (so NG-ESO can integrate the nowcasts into its Platform for Energy Forecasting (PEF) and other downstream systems), and through a web-based user interface that OCF will build. The service will deliver real-time solar PV nowcasts, updated frequently.

We will conduct in-depth user interviews to understand the users' needs and requirements and translate these needs into machine learning modelling requirements.

Continue to Research How to Improve the Skill of the Solar PV Nowcasts

We still have many ideas for how to improve the skill of the solar PV nowcasts, and we are excited to conduct hundreds more ML experiments to explore if we can continue to improve the skill of the forecasts.

In particular, we plan to:

Extend the Dataset

- Extend the dataset back to 2016 for all data sources
- Use NWPs at multiple altitudes

Explore Further Research Ideas

A brief overview of some specific ML research ideas:
- It looks like satellite imagery is great for telling our models that "a big dark cloud is coming", but satellite images do not contain enough information to tell our models precisely how much solar energy will get through that cloud. Luckily, we can use the recent history of data from individual PV systems. A major focus will be on teaching our ML models to associate clouds in the recent

satellite images with dips in recent PV generation. This should help the models predict ramps.
- Explicitly predict future satellite images (as well as predicting PV yield). This will allow us to "pre-train" our ML models on the entire geographical extent of the satellite imagery. A major message from other domains (such as natural language processing and image processing) is that pre-training often significantly improves performance.

Access to Additional GPUs to Train ML Models

- OCF plans to build at least one more on-premise GPU server to cost-effectively accelerate our ML research.
- Work with Lancium, who run data centres that turn on and off to help balance the grid.
- Use a month of free Tensor Processing Units time on Google Cloud.

Analyse Model Strengths

- Additional analysis on model results to determine where and when the model performs well and gives the optimum benefit, and where it provides less predictive power.
- Continue testing which data inputs improve performance the most.

Probabilistic Forecasts

- Experiment with various ways to allow the model to output a probability distribution over the predicted PV yield (such as "mixture density networks")
- Experiment with using different queries to generate an ensemble of outputs.

Appendix

Metrics

Mean Absolute Error (MAE) and Normalised Mean Absolute Error (NMAE) are calculated using the following equations:

$$MAE = \frac{1}{n}\sum_{i=0}^{n}\left|x_i - y_i\right|$$

$$NMAE = \frac{1}{n}\sum_{i=0}^{n}\frac{\left|x_i - y_i\right|}{c_i}$$

where x_i is the prediction, y_i is the truth value, and c_i is the installed capacity. Note that these formulae can be used for national forecasts and GSP level forecasts.