Notes on Completion: Please refer to the appropriate NIA Governance Document to assist in the completion of this form. The full completed submission should not exceed 6 pages in total. Network Licensees must publish the required Project Progress information on the Smarter Networks Portal by 31st July 2014 and each year thereafter. The Network Licensee(s) must publish Project Progress information for each NIA Project that has developed new learning in the preceding relevant year.

NIA Project Close Down Report Document

Date of Submission

Jan 2026

Project Reference Number

NIA2_NGESO002

Project Progress

Project Title

Solar PV Nowcasting

Project Reference Number

NIA2_NGESO002

Funding Licensee(s)

NESO - National Energy System Operator

Project Start Date

September 2021

Project Duration

3 years and 9 months

Nominated Project Contact(s)

Innovation@neso.energy

Scope

The project is split into three different work packages:

WP1 - Design
Research & develop the use of machine learning & satellite images to nowcast PV power generation at GSP-level. Research will be conducted in close collaboration with academia & industry.

WP1 will include the following:

- Use historical data to train machine learning models.
- Evaluate forecast skill using historical data.
- Compare against ESO's current approach; and ESO's approach + gridded NWPs; and CM SAF.
- ML model output: Probabilistic predictions for total solar PV power generation for each GSP at 5-minute intervals.
- The system will also estimate PV outturn 'now' (situational awareness).
- Calibrate forecasts in near-real-time using live PV power data.
- Static designs for user-interface (UI)

WP2 – Prototype Development

Research & develop a prototype of real-time PV nowcasting system. Research how to present nowcasts to Control Room users via an interactive web user-interface. Build a prototype of API & web UI & nowcasting engine capable of running in real-time.

WP2 will include:

- Develop operational requirements with Control Room engineers
- Research ways to run the nowcasting ML models every 5 minutes for all of the 1 million PV systems in the UK.
- Develop a suite of PV nowcast performance validation metrics
- Build a functional prototype of nowcasting system & web user interface.
- Expose PV nowcasts via the API for PEF implemented in work package 1.
- The UI will give an overview of the country's UK PV fleet and also allow the user to drill into details.

WP3 – Prototype Demonstration
Evolve nowcasting system through multiple rounds of user feedback. Quantify impact on grid balancing.

WP3 will include:

- Feedback, evolve, develop with Control Room engineers to learn what they do and don't like about the prototype PV nowcasting system.
- Quantify the effect of PV nowcasting on balancing costs & CO_2 emissions (building on the Control REACT NIA project).
- Explore using probabilistic PV nowcasts to dynamically set reserves (working with the Dynamic Reserve NIA project).
- Explore ways to measure users' interactions with the PV nowcasting system.
- Research ways to blend PV Live with PEF's half-hourly PV forecasts with OCF's 5-minutely PV nowcasts.

Objectives

The project proposed to complete the following objectives:

1. Research report comparing the performance of our ‘Deep Learning’ nowcasting system against other forecasting techniques (including ESO's current approach).
2. Cost estimates for running business-as-usual (BaU).
3. Prototype web UI & PV nowcasting service running in real-time. Including a suite of validation metrics.
4. Report on the feasibility of running ML-powered PV nowcasts in real-time.
5. Developed and socialized agile/CI/CD methodology.
6. Workshop on implementing satellite-powered nowcasts BaU.
7. Workshop on agile/CI/CD in critical energy applications.
8. Joint proposal with ESO for candidate adjacent project with planning on cadence and target level of agile/CI/CD.

Success Criteria

Project success will be defined by:

1. Demonstration of improved PV nowcasting skill compared to ESO's existing PV forecasting approach.
2. Delivery of a prototype of PV nowcasting service.
3. Measured change in balancing costs & CO_2 emissions. User satisfaction. Identified next steps to apply agile/CI/CD to a future NGESO project.

Performance Compared to the Original Project Aims, Objectives and Success Criteria

The project successfully met its primary objectives and key success criteria. The project leveraged machine learning and satellite imagery to nowcast PV power generation at the national level. An interactive web user interface was also developed to enable Control Room users to access and review Solar PV forecasts.

WP1 – Design
Throughout WP1, we conducted Machine Learning (ML) research, involving thousands of experiments, and tested four different models which have all outperformed existing NESO models in 2021. At the end of WP1, the national solar generation forecast was 2.8 times better than the NESO PV forecast (for forecasts up to two hours ahead). NESO's existing national solar PV forecasts model in 2021 had a Mean Absolute Error (MAE) of 650 MW. The best national PV forecasts developed in this WP had a MAE of 233 MW an improvement of 64%.
The guiding principle of the model design is to use microservices with clear interfaces and deploy the infrastructure on Amazon Web Services (AWS) to make the system easily scalable and robust.
The top four components of system design are for data ingestion, training, and predictive components are for Machine Learning modelling and results. The results are then presented through the Application Programming Interface (API) and User Interface (UI) components.

Data Sources:
NWP (Numerical Weather Predictions): This service will collect relevant numerical weather predictions from the UK Met Office and save them to an AWS S3 bucket.
Satellite: This service will collect relevant satellite images from European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and save the images to an AWS S3 bucket.
PV: This service collects live PV output data from 'pvoutput.org' and saves this data to an AWS Relational Database Service (RDS).
GSP (Grid Supply Points). This service collects live GSP output data from 'PVlive' and saves this data to an AWS RDS.

WP2 – Prototype Development:
During WP2, we continued Machine Learning (ML) research, started User Experience (UX) & UI architecture development as well as open-source community engagement to deliver a prototype of the PV Nowcasting service. Early in WP2, we engaged with key external stakeholders and end users of the forecasts to make sure the service meets specific NESO requirements. A session was held in June 2022 with relevant NESO stakeholders to demonstrate the first UI prototype of the service. As a result of further ML research in WP2, a better performing model was developed that has high potential to outperform the current NESO PV forecast. More research in this area was conducted in WP3.

WP1 allowed us to research the fundamental ML models, gather the basic data we required, and to create a template for the deployment infrastructure. In WP2, we have continued the work on the ML modeling.

During WP2, we worked on two models: a relatively low-risk model named "PVNet" which is an evolution of the Convolutional Neural Network (CNN) we developed in WP1. And a high ambition but more experimental model we named "Power Perceiver". The PVNet model is running in production now.

WP2 had two key deliverables:
1. Feasibility of Running ML-powered PV Nowcasts in Real-Time
2. Access to a Nowcasting Service Prototype web UI & PV Nowcasting service running in real-time.

Data Used to Train and Evaluate ML Models:
We used the same training and evaluation data as the data we used in WP1, to allow us to compare the WP2 models with the WP1 models. The models were trained on data from 2020 and evaluated on data from 2021. Our best model at the end of WP2 halves the error for GSP-level forecasts compared to OCF's best model in WP1 and reduced the national forecast error to about 80% of the error in our best model from WP1 (from 233 MW to 190 MW).

WP3 – Prototype Demonstration
In WP3, we started working on creating national forecasting models that should significantly improve the accuracy of our forecast. We also continued work on productionising the service. Due to PVLive updating its GSP regions, we spent a reasonable amount of time making the GSP data updates. Further work on developing a UI and engaging with users resulted in the successful deployment of the first version of the forecast for use by the control room.
In this WP, we also started working on creating national forecasting model based on Extreme Gradient Boosting (XGBoost). This model works by predicting the aggregate national demand directly in contrast to the current model (PVNet) that was predicting each grid supply point and then summing up the generation afterwards.

One of the largest investments we made in WP3 was improving the data pipeline to make the data pipeline components more reusable, reduce errors, and allowed for faster feedback loops.

We also improved the API by adding security which allowed us to give secure API access to users and prevent unauthorised users accessing the service.

WP1 (extension) - Added probabilistic forecasts:
In WP1 of the project extension, we completed our work on the national model and delivered a user-consumable probabilistic forecast. We back tested the probabilistic forecast with a train and test set to check the results are not over fit, and the live results will have the same accuracy. We allowed the user to access the probabilistic results through the API and the UI we have made.
In this WP, we developed PVNet 2.0 through the extension, which was an upgrade from PVNet 1.0. The model used an increased number of Numerical Weather Prediction (NWP) variables, PVNet summation, European Centre for Medium-Range Weather Forecasts (ECWMF) data and ensembling for model improvements providing a 29% decrease in MAE. It also used a new, more flexible data pipeline which allowed faster iteration when training the model. This allowed us to conduct more model experiments. All these iterations resulted in improvements in the 4-8 hours forecast by approximately 30% for the GSP and National forecasts.

Other developments in this WP included:
1. Improvements to the user interface, including the "dashboard view" and probabilistic display for the Control Room.
2. Extended National Forecast from 8 hours to 36 hours.
3. Probabilistic forecasts for all horizons
4. Backtest runs for Dynamic Reserve Setting (DRS) project showing almost 100 MW reduction in reserve settings
5. UI speedup, with query times from 20s down to <1 second

The API was deployed on Amazon Web Services and can now be accessed by anyone. Please see api.nowcasting.io/docs for further information.

WP4 – Ensemble Model Development:
The goal of this WP was to integrate NWP ensembles into the PV Nowcasting pipeline and explore methods for generating probabilistic outputs that account for both weather-related uncertainty and machine learning model uncertainty.

In WP4, we conducted a literature review of the approaches to combining ensemble weather forecasts to generate solar PV forecasts. We created a pipeline for ensemble weather data download and transformation into a format that can further be used in the existing ML pipeline, as well as the creation of an archive. We adjusted the existing ML model setup for work with ensemble weather data and running inference on the collected data to obtain initial results and produce solar PV forecasts at Grid Supply Point (GSP) level and national level using ensemble weather data.

The use of NWP ensembles demonstrated measurable improvements in forecast performance. Results show that incorporating NWP ensemble improved deterministic forecast accuracy by 5% and reduced large forecasting errors by up to 12%.

Required Modifications to the Planned Approach During the Course of the Project

As an R&D project, some modifications to how we carry out ML research were envisaged from the start. See the 'Lessons Learnt' section below for details.

The project was extended by 15 months to deliver the model's probabilistic forecast results, developing the machine research further, and focusing on model precision through adding more features to the forecast. The project extension targeted a further 20% improvement in the MAE across horizons up to 4 hours, and 10% for day ahead horizons. This extension included an additional £355,000 in the project costs.

The project was extended to assess the impact of adding improved forecasts to the Dynamic Reserve Setting (DRS) model (this model was developed through the Probabilistic Machine Learning Solution for Dynamic Reserve Setting project NIA2_NEGESO003)

This extension did not affect the project timeline but added an additional £80,000 to the project costs. The project was then extended further to investigate the use of NWP ensemble to comprehensively account for both weather and ML model uncertainty in probabilistic outputs. This further extension added 7 months to the project timeline and an additional £103,500 to the project costs. During the course of the latest extension, the initial timeline was extended by 5 months due to delays in acquiring weather ensemble data from external sources. The original plan involved exploring weather data from two different sources into the modeling framework. However, in light of time constraints and challenges encountered in data acquisition, the approach was revised to utilise only one weather data source. This approach did not affect project deliverables as we still explored the methods of ensemble-based forecasting. No modifications were made to the project's scope or key objectives.

Lessons Learnt for Future Projects

In the course of W1 and WP2, the project identified the following lessons:

1. Never underestimate the importance of cleaning up and checking data in advance: Several approaches to loading data were tried, from on-the-fly to pre-preparing, and instituted automatic and visual tests of the data to ensure the project was always lining up the various data sources correctly. The more effective approach has been to pre-prepare thousands of batches and save each batch as a separate file on disk. This allowed us for rapid loading of these files during training, resulting in a 12x increase in training speed compared to loading data on-the-fly.

2. Having infrastructure as code allows the main production service to run uninterrupted: Having code to easily instantiate infrastructure is very useful to the efficient management of environments to ensure the project could bring the algorithm into productive use. The Terraform software tool was used which made spinning up (and down) environments very easy and repeatable. Being able to spin up new environments allowed the project to test new features in development environments while allowing the main production to keep running uninterrupted.

3. Using Microservices to “start simple and iterate” accelerates development: Using a microservice architecture allowed the project to upgrade individual components as we saw benefits in improving them, independently of changing other components’ behaviour. This was very useful when building out the prototype service, as it allowed the project team to start with a simple architecture - even a trivial forecast model - and iteratively improve the functionality in the components. For example, first starting out with one PV provider of data allowed the project to get a prototype working, and in WP3 we will expand onboard an additional PV provider.

4. Data processing may take longer than expected: While it was initially planned to extend our dataset back to 2016 for all data sources during WP2, it turned out that data processing took much longer than expected. This did not have a direct impact on project deliverables but is something to consider in further ML research.

5. Data validation is important: For both ML training and inference, using clear and simple data validation builds trust in the data. This helps build a reliable production system and keeps software bugs at a minimum.

6. Engaging specialist UX/UI skills is important: By acknowledging that UX and UI design is a specialised area and incorporating those skills, a UI was developed which will be easier to use and convey information effectively. This was validated over WP3 through working with the end users (control room engineers).

7. Building our own hardware demonstrates value for money but may pose other challenges for a small team: Two computers have been built during the project with a total of six GPUs and it was estimated that using on-premises hardware instead of the cloud for data-heavy & GPU heavy machine learning R&D can significantly reduce the direct costs. However, the time it would require for a small team to put together all the components was significant (approx. 25 days for one person in total). While the total costs would still be lower, appropriate resource planning should be considered if planning hardware upgrades in the future.

In the course of WP3 and WP1 (extension), the project identified the following lessons:

1. Merging the code right away when performing frontend testing is of upmost importance: Merging the code after frontend testing proved to be time-consuming, and it was important to consider when performing tests.

2. Large Machine Learning models are harder to productionise: Large Machine Learning models proved to be difficult to productionise due to their substantial requirements for compute, memory, latency and operational overhead which increase deployment cost and system complexity. The size of these models introduces inference latency, hardware dependency, and scalable contraints, making them difficult to integrate into real-time and cost-sensitive production environments.. Ongoing advances in model efficiency, compression techniques, specialised hardware, and Machine Learning Operations (MLOps) tooling are expected to significantly reduce these barriers, enabling more practical and scalable deployment of these large models in the future.

3. Machine Learning training always takes longer than expected: Even with an already made model, making data pipelines work correctly took time. It was important to always have enough time allocated when planning ML training activities.

4. Security and authentication is hard: Ensuring robust authentication/security measures are in place was harder than we envisaged. It may be easier to implement packages already built or contract third party providers to support the process.

5. Separate National Forecast Model: PVLive estimate of National Solar generation does not equal the sum of PV Live’s GSP generation estimate. This motivated us to build a separate National forecast, compared to adding our GSP forecast.

6. Investment is needed to take open-source contributions to the next level: Time and resources are needed to engage with open-source contributors and develop an active community. We may want to consider hiring an additional resource to support this activity.

Lessons from WP2 (extension) were as follows:

1. Expensive cloud machines storage disk left idle: We used some GCP/AWS machines for R&D, and we often paused the machine when they are not in immediate use. This was because some GPU machines could cost significant amounts per hour. It was discovered that costs still accrue for the disk (storage) of the paused machines. Balancing the pausing of the machines with the ability to start them up quickly versus starting a cloud machine from scratch each time has no golden rule, but it was useful to be aware of.

2. Challenging to maintain active communication with NESO: Particularly high turnover at NESO forecasting team affected communication on the project. This evolved over the duration of the project, and more active and easier communication was observed in the latter phase.

3. Reproducibility on cloud vs local servers: When results differ between the cloud and local services, it can be tricky to determine the cause. Verbose logging, saving intermediate data, and maintaining consistent package versions and setup on both machines helped. One particular bug involved results differing when multiple CPUs were used locally, but only one CPU was used in the cloud.

4. Protection of production data: Two environments in the cloud, “development” and “production,” are maintained to protect the “production” data. This setup allowed developers to access the “development” environment, where changes do not affect the live service. Although maintaining two environments increases costs, it was considered worthwhile.

5. Probabilistic forecasts: Some unreleased open-source packages were used to implement one of the probabilistic forecasts. The advantage of using this code before its release was noted, but it also means more thorough checks for bugs are required, which can take more time.

6. Leap year - Clock changes all good due to using UTC. But tiny bits of code broke on leap year. Needed more testing and checking.

Further lessons from WP (extension) of NWP ensemble model development were as follows:

1. Clear project goals are important from the outset: Clarifying project goals early, even in research projects with high uncertainty, prevents scope creep and wasted effort. This ensured everyone is aligned with what success looks like from the outset.

2. Schedule enough time for obtaining data: Obtaining ECMWF data proved challenging due to its substantial size (over 100GB per day) and the considerable time it takes to download, which can range from over 60 hours for a month's archive to as much as 120 hours depending on the archival system's workload. It is therefore important to schedule enough time to obtain data in the early stages of project planning.

3. Optimising Data for Efficiency: Streamlining the number of variables used in the model can significantly cut down on data download times and storage requirements. A study on feature importance pinpointed 6 out of 12 variables that collectively hold most of the information PVNet needed for forecasting, enabling a halving of both download time and storage by providing regular deterministic data for the remaining variables during inference.

4. Importance of Visualisation for Ensemble Data: The multifaceted nature of ensemble data makes it difficult to quickly grasp. Various visualisation techniques are essential for comprehending general trends, spread, and areas of uncertainty.

5. National ensemble-based forecasts are achievable within reasonable timeframes: Running inference for 50 forecast versions for a single Grid Supply Point (GSP) took approximately 5 seconds, following an initial setup period of under 6 minutes. This suggested that a national ensemble-based forecast could be produced in around 25 minutes.

6. Investing in foundational tools streamlined the entire research and development process: Ensuring enough time is allocated to develop the right tools, such as ocf-data-sampler or new training models, accelerates future development and improved overall efficiency.

Note: The following sections are only required for those projects which have been completed since 1st April 2013, or since the previous Project Progress information was reported.

The Outcomes of the Project

Fully operational PV Nowcasting service ran on two ML models:

1. PVNet for 0-6 hours
2. Blend of PVNet and National_xg for 6-8 hours, and National_xq from 8 hours and beyond.

Accuracy improvement over the previous OCF model by approximately 30% for the GSP and National forecasts (4-8 hours), resulted in forecasts approximately 40% more accurate than the Balancing Mechanism Reporting Service (BMRS) model and over 40% against the PEF forecast (for 0-8 hours).

Probabilistic forecasts for all horizons:
1. Backtest runs for DRS project
2. UI including a new Delta view, dashboard view and probabilistic display
3. UI speedup, with query times from 20s down to <1 second

The most significant of these was the achievement of the target set by NESO of a 20% reduction in MAE error. This was extremely large in renewable forecasting and was the result of numerous machine learning improvements.

Literature Review of NWP ensemble to identify common trends and the methodologies for Solar PV forecasts creation using ensemble weather data. Research showed that incorporating NWP ensemble improved deterministic forecast accuracy by 5% and reduced large forecasting errors by up to 12%. Integration of NWP ensemble into existing PV Nowcasting service running was not done yet as the extension was focused only on research and development of methodology on how to combine ensemble forecasts to create a probabilistic forecast.

Lastly, the forecast from Open Climate Fix was delivered completely open and documented. The resilience was significantly increased over the project duration, resulting in over 99.5% availability. This resilience was implemented by NESO, with all the infrastructure constructed in code to allow replicability.

Data Access

Details on how network or consumption data arising in the course of NIA funded projects can be requested by interested parties, and the terms on which such data will be made available by NESO can be found in our publicly available "Data sharing policy related to NIA projects (and formerly NIC)" and Innovation | National Energy System Operator.

National Energy System Operator already publishes much of the data arising from our NIA projects at www.smarternetworks.org. You may wish to check this website before making an application under this policy, in case the data which you are seeking has already been published.

Foreground IPR

NIA2_NGESO002_Feasibility of Running ML-powered PV Nowcasts in Real Time
NIA2_NGESO002_Research Report: PV Nowcasting Using Deep Learning_WP1-1
NIA2_NGESO002_Open-Source Research Code and Results_WP1-2
NIA2_NGESO002_Progress Report_WP1
NIA2_NGESO002_Progress Report_WP2
NIA2_NGESO002_Progress Report_WP3
NIA2_NGESO002_Progress Report_WP1 (extension)
NIA2_NGESO002_Progress Report_WP2 (extension)
NIA2_NGESO002_Progress Report_WP4 (ensemble_extension)

Planned Implementation

The research work on NWP ensemble extension demonstrated significant improvements in solar PV forecasts. We also explored the application of weighted ensembles for solar PV forecasts. While the original project output is already in use within the Control Room, the NWP extension will require additional research to determine optimal methods for selecting ensemble members and assigning weights to enhance forecast performance. Further research and development are required to demonstrate that the ensemble model can be effectively integrated into existing PV Nowcasting service.

The other next step in innovation could be exploring cloudcasting which has potential to further reduce error by ~5%.

Net Benefit Statement

The National PV Nowcast (Quartz forecast) was improved through many iterations of the project deliverables. The forecast at the end of Work Package 3 showed significantly lower errors relative to the PEF and BMRS forecasts used, with errors 38% and 39% lower respectively by the end of Work Package 3, with further improvements since then. Importantly, large errors were hugely reduced with 1 GW errors being recorded on 23% of days using PEF, and 5% of days using Quartz.

A web User Interface was developed to deliver the PV forecasts. These forecasts were displayed on one of the large screens in the control room for over 2 years and were consulted daily by the control room Engineers.

Using a rule of thumb method from [ESO: Forward Plan Technical Annexe], reducing the daily demand forecast by 10 MW resulted in around a £20m reduction in annual costs. Using Quartz forecast was found to reduce the demand forecast error using the NESO model by 8%, or many tens of MW, or over £40M per year.

Secondly, through the additional Dynamic Reserve Setting savings of up to 96 MW per settlement period, NESO will benefit from up to £30 millionpounds and 100,000s of Tonnes of CO2 per year in savings.

Lastly, Control Room Engineers are making better decisions every day with this improved solar forecast, which will have multiple benefits in carbon and cost which are hard to quantify.

The work is having an impact outside of NESO, as well as being a BAU service for NESO. The Quartz forecast is being used by 8 energy traders and smart home operators to improve their operations, further improving GB market efficiency.

Other Comments

The Project outcomes and results contain confidential information and intellectual property rights that cannot be disclosed in this Report due to their proprietary nature. Should the viewer of this Report ("Viewer") require further details this may be provided on a case by case basis following consultation of all Publishers. In the event such further information is provided each and any Publisher that owns such confidential information or intellectual property rights shall be entitled to request the Viewer enter into terms that govern the sharing of such confidential information and/ or intellectual property rights including where appropriate formal licence terms or confidentiality provisions. Dependent upon the nature of such request the Publishers may be entitled to request a fee from the Viewer in respect of such confidential information or intellectual property rights.

Standards Documents