Solar Power Ensemble Forecaster

Final Report - Public

Project Summary and Findings

Authors:

- IMC: Tim Snell, Sebastian Consani
- CSIRO: Sam West, Matt Amos
- University of South Australia: Sleiman Farah, John Boland
- University of New South Wales: Abhnil Prasad, Merlinde Kay

Release Date: March 2021

Contents

Figures and Tables

Acronyms

AC	Alternating Current
AEMO	Australian Energy Market Operator
AEMC	Australian Energy Market Commission
ASEFS	Australian Solar Energy Forecasting System
AWS	Amazon Web Services
BOM	Bureau of Meteorology
CF	Contribution Factor
CPF	Causer Pays Factors
DC	Direct Current
DI	Dispatch Interval
DUID	Dispatchable Unit Identifier
ECM	Energy Conversion Model
ECMWF	European Centre for Medium-Range Weather Forecasts
EMMS	Electricity Market Management System
FCAS	Frequency Control Ancillary Services
FI	Frequency Indicator
GTI	Global Tilt Irradiance
IP	Intellectual Property
LNEF	Lower Not-Enabled Factors
MAE	Mean Absolute Error
MP5F	Market Participant five-minute Self Forecast
NEM	National Energy Market
NER	National Electricity Rules
NEMDE	National Electricity Market Dispatch Engine
NWP	Numeric Weather Prediction
PCM	Power Conversion Model
PP	SCADA Possible Power
PPA	Power Purchase Agreement
PV	Photovoltaic
rCF	Relative Contribution Factor
RMSE	Root Mean Squared Error
RNEF	Raise Not-Enabled Factors
POE	Probability of Exceedance
SCADA	System Control and Data Acquisition
SDC	Semi Dispatch Cap
SF	Participant’s five-minute ahead Dispatch Self-Forecast
SIFM	Satellite Irradiance Forecasting Model
SPEF	Solar Power Ensemble Forecaster
UIGF	Unconstrained Intermittent Generation Forecasts
UPS	Uninterruptible Power Supply
Web API	Web Application Programming Interface

1. Introduction

The objectives for the "2018/ARP161 – Industrial Monitoring and Control – Skycam and Multi-Model Solar Forecasting" Project (SPEF) were to:

- demonstrate the ability of semi-scheduled generators to submit their own five-minute-ahead Unconstrained Intermittent Generation Forecasts (UIGF) that are more accurate than the Australian Solar Energy Forecasting System (ASEFS);
- improve the accuracy of self-generated five-minute forecasts and test the ability to submit the forecasts into the AEMO market dispatch system;
- increase understanding of the extent to which more accurate self-generated five-minute forecasts can improve the value of solar generation in the NEM;
- determine the value of improved forecasting for solar farm operators and the AEMO based on estimate changes to 'Causer Pays' charges; and
- develop an understanding of the broader cost-benefit analysis of forecasting technology.

The project was carried out in accordance with the requirements outlined in the "2018ARP161 IM&C STF Funding Agreement" document. This report summarises the work completed throughout the project and analyses the technical and economic findings associated with solar forecasting in the Australian National Energy Market.

1.1. Background

The Australian Solar Energy Forecasting System (ASEFS) provides forecasts for solar generation across the NEM. The dispatch of semi-scheduled solar farms depends on the output of ASEFS. The generation forecasts produced by ASEFS use a combination of statistical methods and Numerical Weather Prediction-based models, covering forecasting timeframes from five minutes (Dispatch) to two years (MT PASA).

For a five-minute ahead timeframe, ASEFS produces Unconstrained Intermittent Generation Forecasts (UIGF), which are used to produce a dispatch target for the semi-scheduled solar farm. The National Electricity Market Dispatch Engine (NEMDE) then dispatches the solar farm production based on the bid information provided, and the UIGF provided from ASEFS. The semi-scheduled cap flag may also be set based on constraint limitations or bidding reasons. If this cap is set to false, then the intermittent generator is not required to follow dispatch targets. If the flag is set to true, then the intermittent generator is required to follow the dispatch target only in that its output must not exceed the dispatch target value and will be monitored by the noncompliance monitor.

There are shortcomings in this system that result in high FCAS Causer Pays penalties and reduced revenues for solar farm operators. Improving the accuracy of solar forecasting should help the entire energy system, by improving grid stability and security, helping to better integrate renewable sources into the NEM. As such, semi-scheduled market operators may submit a market participant five-minute ahead self-forecast (MP5F). This project aimed to trial and test MP5F technology at a geographically and operationally diverse set of NEM connected solar farms.

1.2. Trial farms

The SPEF self-forecasting trial submitted MP5F live solar forecasts on behalf of five solar farms. These solar farms have large geographic separation, and include diverse local weather and operational forecasting conditions.

Figure 1 Solar Farm

Figure 2 Images showing the primary camera and weather sensors (left) and the secondary stereo camera (right)

2. Project summary

Fourteen real-time solar power forecasting models were developed, deployed and continuously improved on five solar farms over nine months. Individual models' forecasts were combined into Ensemble models, improving the overall prediction accuracy. Finally, a reimplementation of the Causer Pays procedure was developed to allow financial comparison of the fee reduction resulting from the individual models' forecasts.

Challenges around site integration, project delays, and high solar farm staff turnover in this young industry created deployment difficulties. Technical issues with the varying data quality and precision from the different sites initially caused problems for forecast accuracy and stability, until technical solutions were developed. Significant delays in obtaining data and information about the Causer Pays procedure delayed the financial analysis and limited the development of the planned financial model optimisation algorithms.

Despite these challenges, a high performing ensemble of models is now producing live forecasts used in real-time energy market dispatch and is shown to be generating large fee reductions for the participating solar generators, helping to enable a more stable grid with higher renewable energy penetration, lower energy prices and ultimately lower carbon emissions. The main quantitative findings of this project are summarised below:

1. Significant increases in forecast performance were achieved, as compared to AEMO's incumbent forecasting system. Overall, the best performing models and skill (percentage improvement) vs ASEFS at each site over the last three months were:

Site	Best Model - RMSE	% RMSE Skill vs ASEFS	Best Model - MAE	% MAE Skill vs ASEFS
SF1	EnsembleML	9.19%	Ensemble Median	13.0%
SF2	Ensemble Mean	16.2%	Ensemble Median	18.2%
SF3	Skycam	19.3%	Smart Persistence	16.9%
SF4	Ensemble Mean	2.8%	ASEFS	-
SF5	Skycam ET	21.2%	Skycam ET	16.6%

2. Significant Causer Pays fees savings were achieved from operational forecasts submitted to AEMO at all four sites analysed. Estimated Causer Pays fee reduction over the four-month period from August to November 2020 from the live dispatched ensemble forecast were found to be:

Site	ASEFS Fee	Dispatch Model	Dispatch Ensemble Fee	FCAS Saving vs ASEFS	% FCAS Saving vs ASEFS
SF1	$109,720.54	SF_10	$106,642.18	$3,078.36	3%
SF2	$206,003.92	SF_12	$168,702.56	$37,300.44	18%
SF3	$32,963.14	SF_5	$27,130.32	$5,832.82	18%
			Average Saving	**$15,403.87**	**13%**

(SF4 and SF5 were omitted from FCAS fee modelling due to a shortage of data from commissioning delays)

3. Further savings in Causer Pays fees could have been achieved at all four sites analysed if the best financially performing model had been dispatched over this period. Estimated Causer Pays fee reduction over the four-month period from August to November 2020 period from the best financially performing model at each site were found to be:

Site	Top Performing	FCAS Saving vs ASEFS	% FCAS Saving vs ASEFS
SF1	SmartPersistenceModel	$43,819	40%
SF2	SmartPersistenceModel	$88,983	43%
SF3	SkycamET	$23,702	72%
	Average Saving	**$48,740**	**53%**

4. Fee reduction was found to not be directly associated with improved forecast performance, due to several assumptions in how fees are calculated, such as constant generator dispatch performance from month to month, which appears to be invalid when applied renewable generators.

3. Project performance against outcomes

In general, the outcomes of the project are positive. MP5F for five solar farms are being submitted to AEMO for use in dispatch and on average are beating ASEFS on MAE and RMSE.

3.1. Achieved Milestones

The project has achieved its defined objectives. The key milestones achieved included:

- Equipment
  - Manufacture of six complete sets of stereo sky camera forecasting system hardware (including one spare for development)
- Site installations - Installation of stereo skycam forecasting hardware at all sites
- Internet Cloud infrastructure:
  - Setup of high availability AWS infrastructure,
  - connection to third party data sources including BoM and ECMWF,
  - connection to AEMO MP5F forecasting system
- Submission of MP5F forecasts
  - developed, trialled and evaluated several machine learning, physical and statistical models for short-term five-minute ahead solar farm power forecasting
  - 5-minute ahead solar forecasts were submitted for five solar farms and performed better than the ASEFS forecasting system on required metrics

3.2. Difficulties encountered

This section discusses areas where difficulties were encountered during the project.

3.2.1 Site integration

The project experienced significant delays that were mostly the result of site-specific installation and integration problems. With respect to power generation more broadly, solar generation in Australia is a young industry. This is apparent in dealings with most solar farms. The industry is dynamic, which has both positive and negative side effects. For example, there is a rapid turnover of assets within the industry. When assets change hands, there is also a change of personnel. One solar farm had the technical team change four times throughout the duration of this project. Regardless of the skills and expertise of the individuals involved, high staff turnover and a transient workforce means that there is often a lack of in-depth knowledge regarding the operation of a specific solar farm.

Regarding solar forecasting, this is important as a good technical point of contact will help bring a forecasting system online quickly and help to get the optimal performance from a system. Several farms in the trial had issues with supplying correct data points for training of the forecast models (i.e., SCADA mismatch, incorrectly labelled/documented data). Some farms in the trial are still battling reliability issues several years after commissioning.

The greatest difficulty that arose during the commissioning and integration of the self-forecasting system with the sites was the variability of available data points and SCADA systems. For a solar forecasting technology to be scalable whilst remaining financially viable with minimal technical debt, it is imperative to attempt to standardise the data being gathered from the farms. Although each farm has a similar array of sensors and systems, for example inverters, weather stations, and pyranometers, the implementation details and quality of data varied significantly. Some farms could quickly integrate our self-forecasting system and provide the required data, however other sites had limitations with their existing SCADA systems that then required significant changes on their behalf for our system to receive all necessary data points.

3.2.2 Forecast models

There were several difficulties in the development of the forecast models. These included:

- Differences in each site were found when commissioning PCM models, including:
  - Variations in “transformer loss” de-rate values caused by changes in farm or SCADA configurations. These were presumed to occur during commissioning.
  - Minor discrepancies were found between co-located instrumentation (such as pyranometers), indicating hardware orientation errors.
  - Tracking GTI measurements were found to be not following the same path as tracking modules for the entire day. This was related to the Backtracking algorithm applied to tracking arrays not also being applied to trackers used for irradiance measurements.
  - Unexplained changes to array output occurred either when arrays or parts of an array were disconnected without status flag changes.
  - Small differences in SCADA outputs e.g. DC status flags change format (across and within sites).
  - Differences in GTI measurement on clear sky days at large zenith angles (probably because of intra-site shading)
- Various issues were experienced with missing data or poor data quality, including:
  - Frequent data dropouts or instrumentation output freezes caused issues when training models on historical data. Automated method now detects and removes most of these periods.
  - Low precision values make some statistical methods (e.g. Clear Sky Index calculation) error prone during long, apparently unchanging periods.
  - Live semi-dispatch cap data took a long time to acquire via farm SCADA systems, limiting forecast accuracy because this was required for switching to irradiance-only models during capped periods.
  - Farms generally do not provide information about farm maintenance, which decreases reliability of training data unless manually flagged.
- A variety of difficulties were also encountered while reimplementing AEMO’s causer-pays model
  - To produce four-second and five-minute performance factors as per the Contribution Factors Procedure published by AEMO, Frequency Indicator (FI) and Frequency Deviation values are required. Contrary to the documented advice, the FI values are not publicly available. These data sources are accessible as a market participant. However, AEMO was unable to provide data listings available as part of the subscription service to non-participants to verify their inclusion before purchase.
  - Some important parts of the FCAS procedure are undocumented (e.g. FI values are prematurely rounded to five decimal points in AEMO’s calculations, meaning that intervals where FI is approximately zero, are removed from five-minute factor calculations), and data

which is required for the calculation is in disparate, impractical and poorly documented locations across both NEMWeb and the AEMO website.
  - Further enquiries to AEMO seeking clarification on the exact causer-pays procedure, seeking to correct discrepancies between our implementation based on the documentation and AEMO's published results went without a response for almost four months. This severely hampered our ability to produce the cost-benefit analysis for this report and meant we were unable to investigate planned financial optimisation of the models to further reduce causer-pays charges for the sites in this trial.
  - We suggest that AEMO makes public the full dataset required to calculate the Contribution Factors, and includes full worked examples of these calculations, as these were very helpful when eventually supplied.

3.2.3 Integration into AEMO systems

The forecast submission to AEMO via the web API created several challenges throughout the project that caused significant delays in achieving the project outcome, which included:

- An inability for the self-forecast provider to change the passwords for the EMMS user accounts used as the authentication method when submitting forecasts. This was a challenge as the password must be changed every 90 days. For most of the project, this process could only be performed by the farm's participant administrator or IT team. Managing this across five solar farms, with multiple accounts for each farm (pre-production, production and backup accounts) required a significant amount of time and coordination, which could have been better spent improving other aspects of our forecasting system. This issue was resolved, however, near the completion of the project in mid-October 2020, when AEMO implemented an API endpoint that enabled self-forecast providers to change the account passwords, not only manually, but in a programmatic, automated fashion—a significant improvement to the previous process.
- The AEMO servers often had slow response times to a forecast submission, which would cause the forecast submission to miss the gate closure time. This was a significant issue during the initial assessment period for farms, as the self-forecast was penalised in the reliability assessment due to issues that were the responsibility of AEMO. AEMO would only exclude dispatch intervals from the assessment if all self-forecast providers were affected by an error or outage at the same time. However, all self-forecast providers were affected by similar problems, just not at the same time.
- Similar to the slow response times discussed above, a number of other API errors were common occurrences throughout the project. As of the conclusion of this project, these have largely been resolved. However, they caused significant delays in the self-forecast assessment process. There were also some legitimate errors that occurred, especially during the initial account setup process for a farm. These were often authentication issues. However, the error messages were sometimes cryptic, which made it a challenge to relay the required information back to the participant so they could change the user account in the EMMS system.

3.2.4 AEMO initial assessment

The AEMO initial assessment for five-minute self-forecasting was a requirement to submit forecasts for use in dispatch. The assessment start and end dates for each farm are detailed in Table 1.

Table 1 Self-forecasting initial assessment periods for each site.

DUID	Start Date	End Date	Total Weeks
SF21	2020-03-10	2020-07-07	17
SF11	2020-03-31	2020-05-26	8
SF41	2020-05-05	2020-06-30	8
SF31	N/A	N/A	N/A

SF51	2020-09-29	2020-12-01*	9*

*Expected completion dates.

SF2 was the first farm that began the AEMO assessment process, and as such there were some initial teething issues that arose, resulting in a greatly extended initial assessment. The assessment window is eight weeks, extending up to sixteen weeks, where beyond this the assessment window stays at a rolling sixteen weeks until the forecast passes. In this case, most of the poor reliability that caused the extended assessment occurred in the first week, after which fixes had been applied to the forecasting system. Thus, the penalty for issues that primarily occurred only in one week was quite harsh. If the window defaulted to a rolling eight-week window for assessments that extended beyond the initial eight-week window, then the forecast would have been approved after only nine weeks.

The assessments for three sites, SF1, SF4 and SF5, were completed in the specified eight-week window, with no additional weeks required.

SF3 was approved for self-forecasting primarily using forecasts from another forecast provider as the highest priority, with the IMC-provided forecast being submitted with a lower priority, meaning it was only used when the higher-priority forecast was suppressed or not submitted. As such, the assessment period has been marked not applicable in Table 1.

3.2.5 SIFM Model

The greatest difficulties encountered while working on SIFM included:

- Lack of historical data with enough clear and cloudy days for deriving historical relationships influenced GHI forecasts on a 30-day cycle.
- Satellite images are taken every 10 minutes; thus, SIFM is the only model unable to predict power at fixed forecasts horizon(s). The images are also received approximately 13 minutes after they are captured by the satellite, resulting in a longer forecast horizon than the other models. This made the benchmarking process harder, however, to match with other models, the benchmark period was run with five-minute forecast horizons.
- Lack of DNI data for validation of power conversion using GHI. Note, the power conversion model requires both GHI and DNI for calculation of power generated. Currently, no sites measure DNI directly, so it was harder to diagnose errors resulting from DNI estimates.
- Errors resulting from GHI estimates were amplified after power conversion due to the interaction of errors resulting from the two procedures.

4. Forecast Models

SPEF MP5F forecasts are being submitted to AEMO for all farms that are part of the trial. This project developed, trialled and evaluated several machine learning, physical and statistical models for short-term five-minute ahead solar farm power forecasting. It also evaluated several methods to combine these forecasts into an ensemble model that produces a single forecast value by applying weightings to each individual model. Descriptions of each model are in the sections below.

4.1. SPEFModel

SPEFModel is the parent of most models (other than Skycam and the statistical models which were more complex to integrate) and provides a common training, validation and prediction framework to unify the individual forecast models and EnsembleML. It does not provide forecasts directly, but this common functionality makes it easier to develop new models, maintain existing models, add features and deploy them to production.

4.2. RFAR

RFAR is a Random Forest Auto-Regression model, a simple machine learning model that uses the last 10-20 minutes of exported real power from a solar farm to predict the power at the next dispatch interval.

RFAR was one of the first models implemented and generally provides a small amount of forecast skill relative to ASEFS. It is quite robust when SCADA power measurements from the farm are available.

4.3. Skycam

The Skycam model is a supervised machine learning model which combines data from a single on-site sky camera's image processing pipeline with weather and irradiance data collected by the camera's sensors and provided by the site's SCADA system. The model is trained, tuned and validated on 10-second interval data each night and then run once per dispatch interval during the day to produce a power forecast for the next interval.

Two versions of the Skycam model were tested:

- SkycamET—which used the full set of AC/DC power, camera pixel, onboard weather station and site-provided weather, power and irradiance datasets, and
- Skycam-Simple—which used only AC power, camera pixel, onboard weather station, and site-provided irradiance datasets.

Desktop feature-selection experiments suggested that the full SkycamET model should give improved performance over Skycam-Simple. Running both models live tested whether the desktop results were correct, and whether the simple model would be more robust to SCADA and sensor outages and if the variable quality of the additional tables would affect the forecast reliability.
Both Skycam models generally performed better in RMSE than in MAE because they predict sharp irradiance and power ramps due to observed clouds in intermittent conditions better than in clear-sky or very overcast conditions, which is captured better by the RMSE metric. This is because RMSE penalises infrequent but large errors, such as missing the occasional large ramp, more heavily than frequent but small errors, such as small biases on clear sky periods, which are reflected more effectively by MAE.

4.4. Skycam-stereo

Skycam-stereo is an additional image analysis pipeline that uses two cameras spaced 100-200m apart to estimate cloud height maps using stereography. These cloud heights are used as additional inputs to the standard Skycam models. The hypothesis was that the relationship between cloud height and cloud optical depth should provide extra information to the machine learning models, allowing them to better quantify the ramp rates and depth of power drops caused by clouds at various altitudes.

However, in forecast validation studies performed on the combined stereo and single Skycam datasets, no statistically significant improvement in forecast performance was observed. The stereo analysis functions have been validated at only a single site, and the commissioning procedure for new sites seems to be very sensitive to small tilt and rotation differences in the orientation of the two stereo cameras; we believe this may be causing anomalous height results to be generated. This may be why the addition of the cloud height data did not improve the forecasts. We believe that these issues can be resolved through further software development, but additional testing is needed to improve the analysis pipeline and correct for these rotational errors before this can be verified. Given the insignificant improvement in the forecast and the bandwidth and compute overhead to run the stereo pipeline, the stereo data was not used in live forecasts for this trial.

4.5. Power Conversion Model

The power conversion model (PCM) is a physical model that is used to convert supplied irradiance and temperature values into site power output. The model is based on the underlying physics of the site's photovoltaic (PV) modules combined with descriptive performance functions, which are from manufacturer-supplied datasheets or historical performance data obtained at the site.

The model is broken up into a set of sub-models that estimate the power output for each inverter at the site corresponding to each PV array. This breakdown allows for variation of the performance across the field due to different PV array sizes, PV module types, or current status to be captured. The DC power output of each PV array within the field at any supplied set of conditions is estimated using an extended single diode model that has been configured to match the module performance characteristics and scaled to match the PV array size. This DC value is then converted to an AC power estimate, taking into consideration the performance of the installed inverters. A comparison of the estimated and measured values for each inverter allows a further derating refinement to be applied that can account for other losses in the array (e.g. wiring loss, module performance variation, average soiling level). Variations due to the solar angle of incidence and losses occurring between the inverter output and the site power export connection are accounted for using numerical fitting to historical data.

Development of the model requires a detailed assessment of each site including location, layout, topography, string configuration, module and inverter datasheets, tracking behaviour and site operation. Tuning of the model is performed based on historical site data and requires module global tilt irradiance (GTI), module temperature, ambient temperature inverter power and export power. Currently, the PCM is tuned only once for each site however scope for improvement exists in an adaptive tuning regime where de-rate and correction functions are updated in reasonable intervals to better account for variations caused by seasonal changes, temporary array performance variation (e.g. soiling/cleaning) and long-term degradation.

The PCM differs from most other models in the project in that it does not produce forecasts. Since it only acts as an engine to convert an irradiance and temperature dataset into estimated power values, it requires another mechanism that supplies a forecast of these values in order to make predictions. The achievable accuracy for this model is fundamentally limited by how representative the supplied values are for each component of the field. This is most evident when irradiance variation occurs across the site that is not captured in the supplied irradiance data, resulting in significant errors in power estimates.

4.6. Smart Persistence

The Smart Persistence model uses site irradiance data to calculate a clear sky index representing the current cloud cover conditions. Using the PCM, current and future clear sky power values up to the forecast horizon are generated. Site SCADA power data is used to calculate the current clear sky index as a fraction of the current clear sky power as generated by the PCM. The current clear sky index is multiplied by the future clear sky power to generate the Smart Persistence forecast.

An alternative Smart Persistence model was developed using only the site irradiance data mapped to current power using the PCM in place of the site SCADA power data. This model was developed in an effort to alleviate the forecast degradation in intervals after cap-curtailed (non-assessable) periods where the site SCADA power would drive power forecasts low and cause these power-forecast models to have a noticeable impact on performance.

4.7. EnsembleML

EnsembleML is a supervised machine learning regression model which is trained on all available five-minute interval historical single-model forecasts to predict the actual power value for that interval. It also uses rolling window statistics calculated from recent irradiance measurements, to allow selection of the best weights to apply to each model's forecast based on the current irradiance conditions.

For example, during very intermittent irradiance periods when the Skycam model has historically performed well, EnsembleML may place greater reliance on the Skycam forecast, but then switch back to a heavier weighting of Smart Persistence or other performance models during sunny periods.

4.8. Mean Ensemble

The Mean Ensemble was the first ensemble approach implemented. With this method, the mean (average) of selected individual model forecasts is calculated and submitted as the final forecast. To overcome forecast penalties after a cap-curtailed (non-assessable) period, the mean of selected irradiance-forecast models is used instead of the selected power-forecast models. Models are manually selected, based on historical performance, to include in the Mean Ensemble. As a result, as more model forecasts became available, the Mean Ensemble forecast predictions typically improved.

4.9. Median Ensemble

As an alternative approach to the Mean Ensemble, the Median Ensemble was developed to improve the MAE performance, to which the median is more closely correlated. This model was designed to operate similarly to the Mean Ensemble, including switching to irradiance-informed forecasts after cap-enforced periods and manual selection of power and irradiance models for forecasting.

4.10. Statistics Model

This model was implemented using a combination of Fourier series to describe the seasonality and an autoregressive model for the stochastic component. This was re-trained every five minutes on one-minute SCADA data from the farms. Then the model was run for seven steps ahead, and the last five forecasted instantaneous results were averaged and designated as the instantaneous forecast for the start of the required interval. All milestones have been met except the requirement for analysis over 12 months of data, as we do not have that length of time available. However, the statistics model has been running in real-time on all the partner farms and forms part of the ensemble models.

The results are generally good in that it mostly performs better than the present tool ASEFS. There are two provisos to this. One is that it is trained on the SCADA feed from the farms, but for the ASEFS comparison, we have to compare it to INITIALMW which is a separate feed, and the two are not always the same. The second proviso is also due to this restriction in that the SCADA feed occasionally drops out. This latter situation is the reason that this model sometimes performs worse than ASEFS.

4.11. Satellite Irradiance Forecasting Model

The Satellite Irradiance Forecasting Model (SIFM) was developed and tested using near-real-time Himawari 8/9 satellite images. The downwelling solar irradiance was converted to power forecasts using the power conversion model. The SIFM model performed better in capturing GHI (especially on clear days), however, conversion to power forecasts amplified errors. The model was also included in the blended forecasting system to account for intermittent generation. The model can be improved by better capturing the relationship between cloud index and the clear-sky index in the historical data using machine learning approaches with images observed at all spectral wavelengths (bands) and by considering neighbouring grids. Also, reducing errors resulting from the power conversion would greatly improve predictions. The algorithm included three key processing phases:

- ***Offline processing***: The derivation of fitting functions against cloud index and clear sky index using historical observations.
- ***Image Processing***: The derivation of cloud motion vectors using near real-time satellite imagery.

- ***Online Processing***: Derivation of power ensembles using derived GHI from advected pixels after image processing.

SIFM was run in two modes: benchmarking mode and in real-time. The benchmarking mode was used for pre-evaluation, testing and debugging of the beta-version of the model code. The model was then run in real-time as part of an ensemble forecast providing forecasts in real-time. SIFM was only run at four sites including SF1, SF2, SF3 and SF4.

5. Forecast performance

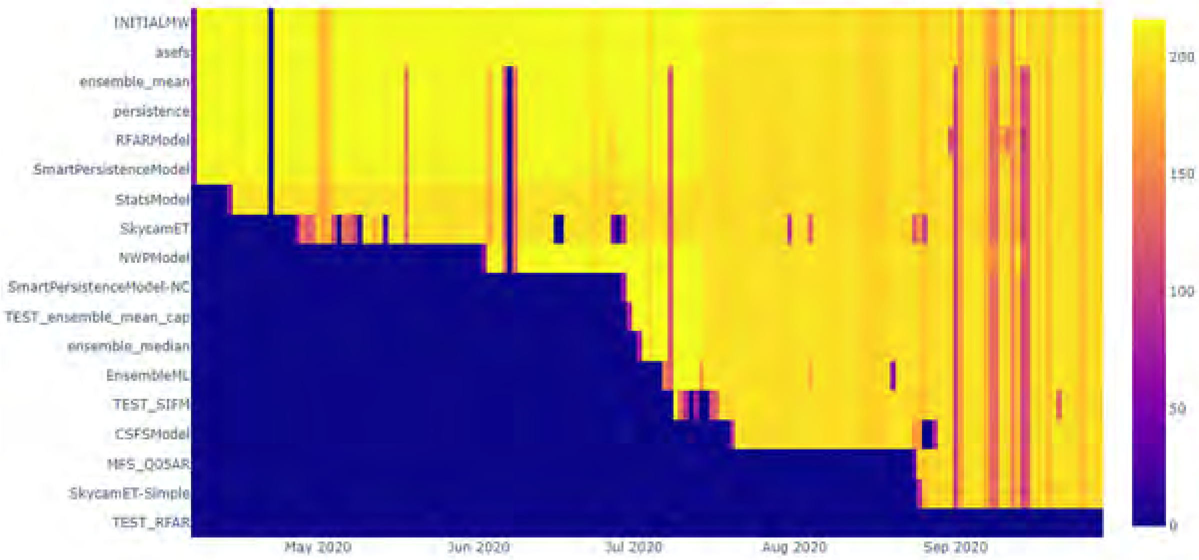
5.1. Methodology

Forecasts were assessed by comparing the same set of intervals for all models to get a valid comparison of model performance for each site. shows heat maps for each site indicating when each model came online. Because not all models came online at the same time, to select the data used for assessment a trade-off had to be made between including as many models as possible and assessing as long a period as possible. Accordingly, we have chosen two periods to assess. The logic for assessment was:

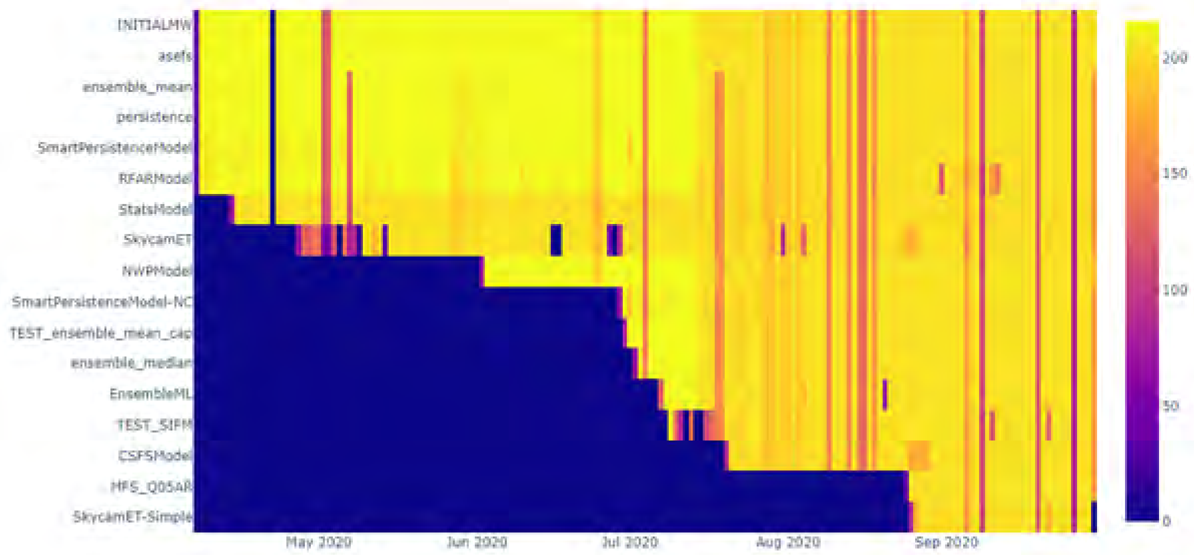
1. For each site, build a table of all available model and ensemble forecasts plus ASEFS forecasts and INITIALMW ground-truth power measurements.
2. Pick two subsets of this data:
   a. **Maximum period**: A period where a subset of the models (the first five models online) were forecasting live for the longest period available. This period was more than six months for most sites.
   b. **Maximum models**: Almost all models were forecasting live. This period was three months for most sites.
3. For each subset, remove any dispatch intervals where we don't have a full set of forecasts from all models. Additionally, any non-assessable intervals (i.e. where the farm output is constrained by the *SEMIDISPATCHCAP* and *TOTALCLEARED* is less than *AVAILABILITY*) were removed from assessment.

All results presented below are split into these two subsets.

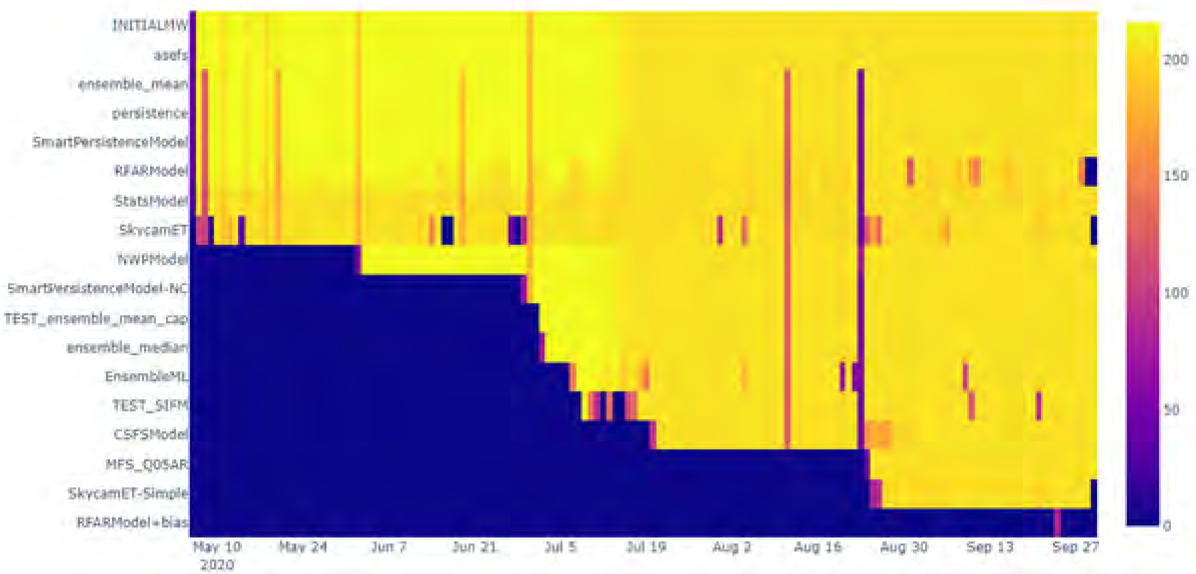
SF1 (SP Start)



SF2 (SP Start)



SF3 (SP Start)



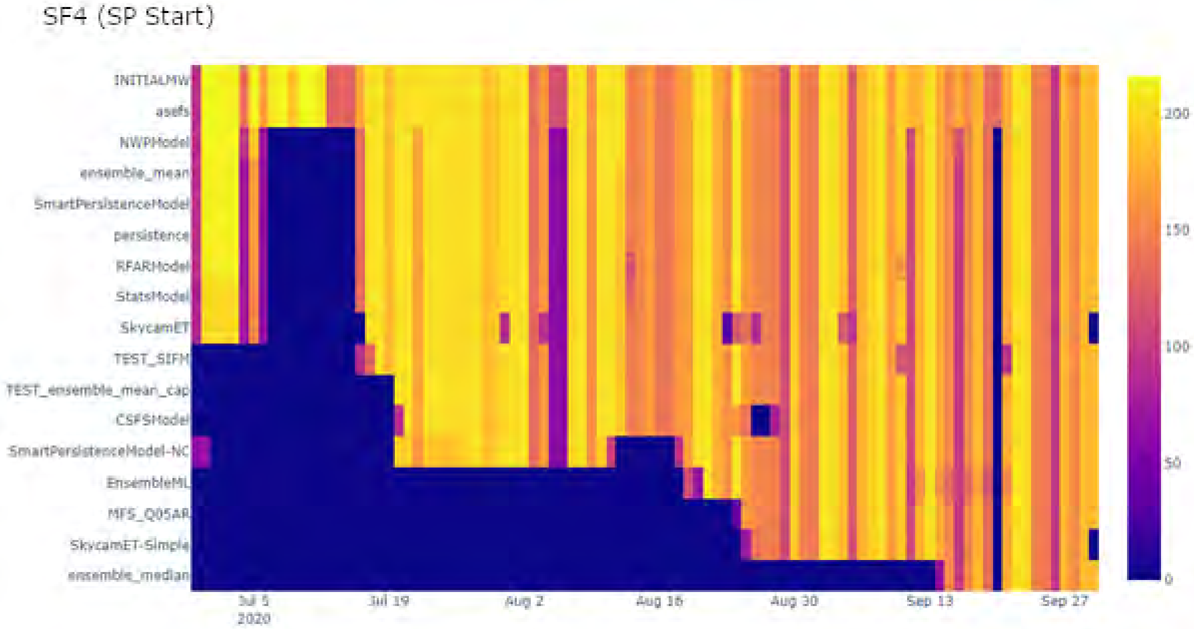


Figure 3 Heatmaps showing when each model produced forecasts per site. Yellow indicates a full set of forecasts for a day, blue and purple shades indicate missing forecasts for a model. Non-assessable periods (where a farm is curtailed because of a semi-dispatch cap) have been removed from this data, most obviously at SF4.

5.2. Forecast performance metrics

We have developed a standardised set of error metrics, described in Table 2 below. Values closer to zero are better for all real-valued numerical metrics in the table below, except R^2 (where closer to 1.0 is better). Table 2 describes the error metrics that were calculated when assessing the forecast models. For brevity, most of these have been omitted from this report, but are supplied in spreadsheets in the supporting material.

Table 2 Forecast error metrics calculated for each model

Field	Description
site	DUID of the generator
model	Name of the model
mean_gen	The mean generated power over the period analysed. All normalised metrics are normalised by this value.
start	First day of the dataset analysed
end	Last day of the dataset analysed
n_days	The total number of days of data analysed, taking into account missing days
n	The number of samples (generally 5min dispatch intervals)
NAs	The number of NAs – i.e. missing values
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
MBE	Mean Bias Error
R2	R2 The coefficient of determination, i.e. correlation between forecasts and actual power
RCF	Relative Contribution Factor – a simple model of Causer-Pays contribution factor based on weighted MAEu and MAEo. This is indicative only and has been superseded by the Financial Analysis shown in Section 5 below.
NRMSE	Normalised RMSE
NMAE	Normalised MAE
RMSEu	RMSE of the under-forecasts only
NRMSEu	NRMSE of the under-forecasts only
MAEu	MAE of the under-forecasts only
NMAEu	NMAE of the under-forecasts only
R2u	R2 of the under-forecasts only
RMSEo	RMSE of the over-forecasts only
NRMSEo	NRMSE of the over-forecasts only
MAEo	MAE of the over-forecasts only
NMAEo	NMAE of the over-forecasts only
R2o	R2 of the over-forecasts only
err_skew	Statistical skewness of the errors
err_kurtosis	Statistical kurtosis of the errors. A measure of the ‘tailedness’ of the error distribution.
err_entropy	Entropy in the errors. A measure of the remaining Information in the errors.
err<1%	% of samples where error forecast is within ±1% of the actual power (as requested by ARENA)
err<5%	% of samples where error forecast is within ±5% of the actual power (as requested by ARENA)
err<10%	% of samples where error forecast is within ±10% of the actual power (as requested by ARENA)

Note that there are some issues with the implementation of the ‘err<*N*%’ metrics, as defined by ARENA, in that they are inflated by values of actual power close to zero. To work around this, we filter out any samples where the actual power is less than 5% of its average.

Generally, RMSE was found to be the most sensitive to forecast errors and is more computationally efficient for model optimisation. RMSE also gives a better indication of infrequent but large differences between forecast and actual export power. Most of our models are optimised for RMSE for these reasons.

From AEMO's published FCAS Causer Pays procedure, MAE appears as though it should relate more closely to FCAS causer-pays fees, because a generator's Contribution Factors are derived from the sum of its absolute under/over errors between dispatch target (which is based on its forecast for renewable generation) and actual exported power. Our modelling disagreed with this assumption, showing that RMSE is slightly more important for calculating fees than MAE is, and fee estimates based on MAE resulted in a linear model with a much worse fit.

5.2.1 Effect of nameplate rating limit

The individual forecast models are generally not used directly for submission to AEMO, rather, they are used as members of the given ensemble forecast. Hence, they operate with slightly different criteria to a forecast suitable for direct use in dispatch. Namely, the forecasted power of the individual models has not been limited to the nameplate rating of the farm, as this allows for more advanced ensembling techniques and outlier rejection that can make use of the slightly more accurate forecasts when a farm's generation exceeds its nameplate power rating.

Conversely, the ensemble forecasts are limited to the nameplate power rating of the farm, as this makes them valid for use in dispatch and submission to AEMO. It must be noted that when a farm is operating at its maximum power output this level will often be slightly above the farm's nameplate power rating. The implication of this is that all forecasts (including the incumbent ASEFS) will have an error that is accumulated when the generated power of a farm exceeds the nameplate rating. This will affect farms that use single-axis tracking more than others, as on a clear-sky day, the generated power curve will have a much wider flat top near to the maximum power output.

5.3. Forecast evaluation & analysis

This section contains plots of MAE and RMSE from all models overall and broken down by various criteria for all sites where sufficient forecasts were available.

5.3.1 Maximum models

The 'maximum models' period was chosen to be 2020-08-01 to 2020-11-11, as this period captured most models for most sites over a reasonable period, allowing a direct comparison of all the models at specific sites and between sites. The only exception is SF5, where this period did not begin until late September when most models were not online there. The Maximum Period section (5.3.2) below shows a comparison over a longer 6 month period, but this was only possible with a smaller number of models, as not all models were running operationally for this whole time.

5.3.1.1 Overall

These plots and table show the overall performance of key metrics (RMSE, MAE, MBE and R^2) for each site over this period. The tables are sorted by ascending MAE.

Analysis:

- While no one model outperforms all others on all sites, one of the three ensemble models was in the top two overall on every site for both MAE and RMSE, suggesting that ensembles are an excellent approach to this type of forecasting.
- On average, NWP, SIFM, RFAR, and SmartPersistence-NC didn't outperform ASEFS on any site during this period. NWP and SIFM provide irradiance-only forecasts during curtailed periods, and operate at

longer time frames than five minutes, so this was expected. RFAR and SmartPersistence-NC had known issues during this period that have since been corrected.

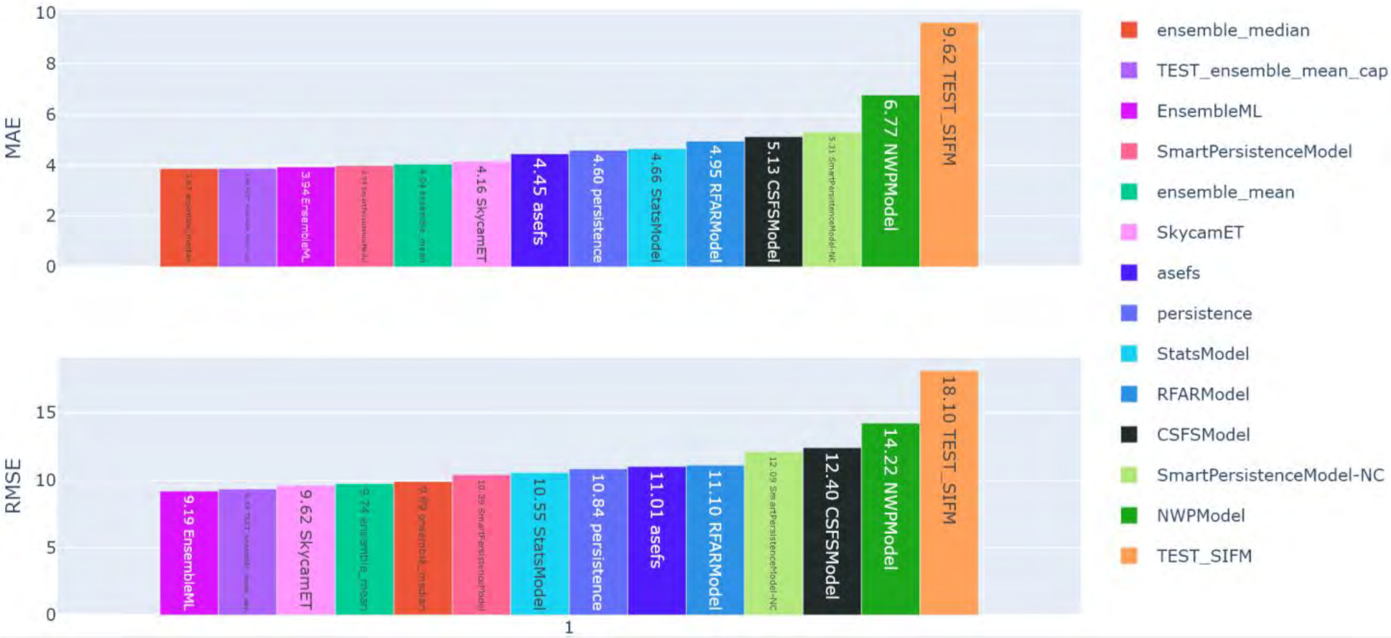
- Overall, the best performing models and skill (percentage improvement) versus ASEFS at each site over the last two months were:

Site	Best Model - RMSE	% RMSE Skill vs ASEFS	Best Model - MAE	% MAE Skill vs ASEFS
SF1	EnsembleML	9.19	Ensemble Median	13.07
SF2	Ensemble Mean	16.23	Ensemble Median	18.19
SF3	Skycam	19.31	Smart Persistence	16.81
SF4	Ensemble Mean	2.84	ASEFS	-
SF5	Skycam ET	21.22	Skycam ET	16.56

SF1

The best model, in both MAE and RMSE at SF1 over this 2-month period was EnsembleML. The ensembles, Skycam and smart persistence models all outperformed ASEFS.

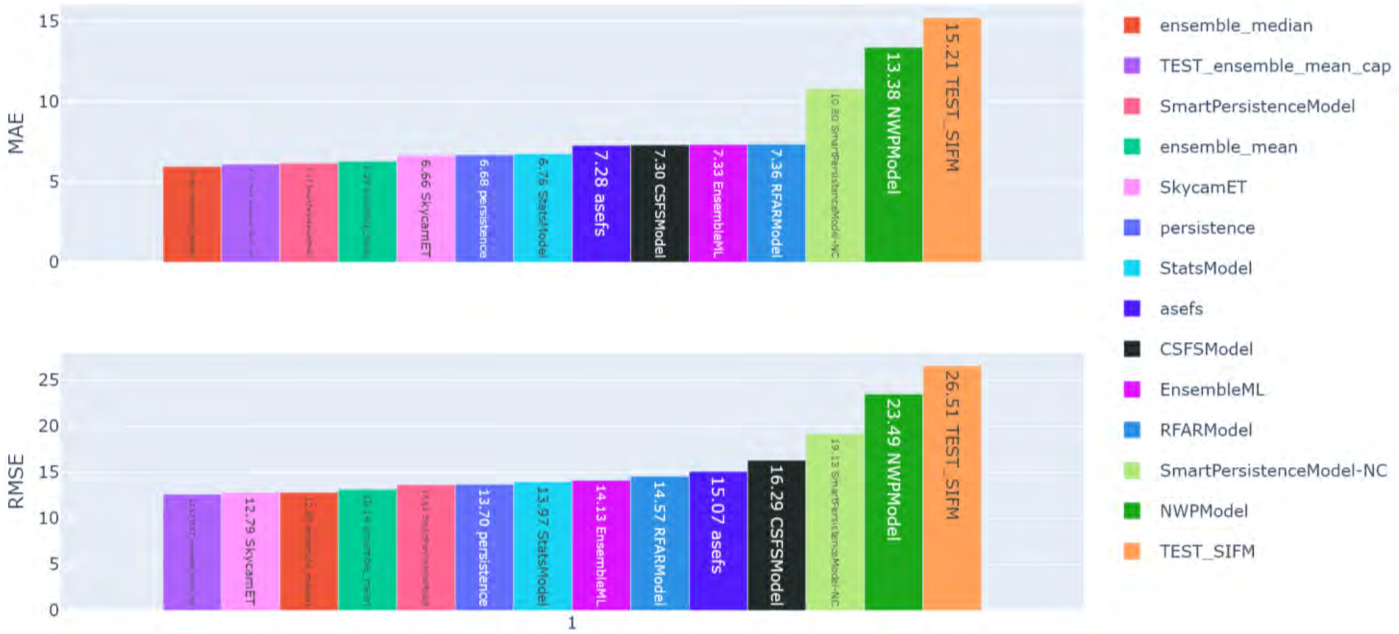
Model	mean_gen	start	end	n_days	n	NAs	RMSE Skill % vs ASEFS	RMSE	MAE Skill % vs ASEFS	MAE	NRMSE	NMAE
ensemble_median	39.25	2020-08-01	2020-11-11	99	17630	0	10.15	9.89	**13.07**	**3.87**	0.25	0.1
TEST_ensemble_mean_cap	39.25	2020-08-01	2020-11-11	99	17630	0	15.16	9.34	12.94	3.88	0.24	0.1
EnsembleML	39.25	2020-08-01	2020-11-11	99	17630	0	**16.54**	**9.19**	11.59	3.94	0.23	0.1
SmartPersistenceModel	39.25	2020-08-01	2020-11-11	99	17630	0	5.59	10.39	10.33	3.99	0.26	0.1
ensemble_mean	39.25	2020-08-01	2020-11-11	99	17630	0	11.53	9.74	9.14	4.04	0.25	0.1
SkycamET	39.25	2020-08-01	2020-11-11	99	17630	0	12.59	9.62	6.57	4.16	0.25	0.11
asefs	39.25	2020-08-01	2020-11-11	99	17630	0	0	11.01	0	4.45	0.28	0.11
persistence	39.25	2020-08-01	2020-11-11	99	17630	0	1.53	10.84	-3.23	4.6	0.28	0.12
StatsModel	39.25	2020-08-01	2020-11-11	99	17630	0	4.15	10.55	-4.63	4.66	0.27	0.12
RFARModel	39.25	2020-08-01	2020-11-11	99	17630	0	-0.87	11.1	-11.24	4.95	0.28	0.13
CSFSModel	39.25	2020-08-01	2020-11-11	99	17630	0	-12.67	12.4	-15.22	5.13	0.32	0.13
SmartPersistenceModel-NC	39.25	2020-08-01	2020-11-11	99	17630	0	-9.88	12.09	-19.25	5.31	0.31	0.14
NWPModel	39.25	2020-08-01	2020-11-11	99	17630	0	-29.17	14.22	-52.08	6.77	0.36	0.17
TEST_SIFM	39.25	2020-08-01	2020-11-11	99	17630	0	-64.46	18.1	-116.04	9.62	0.46	0.25



SF2

Ensemble_median was the best model at SF2 in MAE, while ensemble_mean_cap had the lowest RMSE.

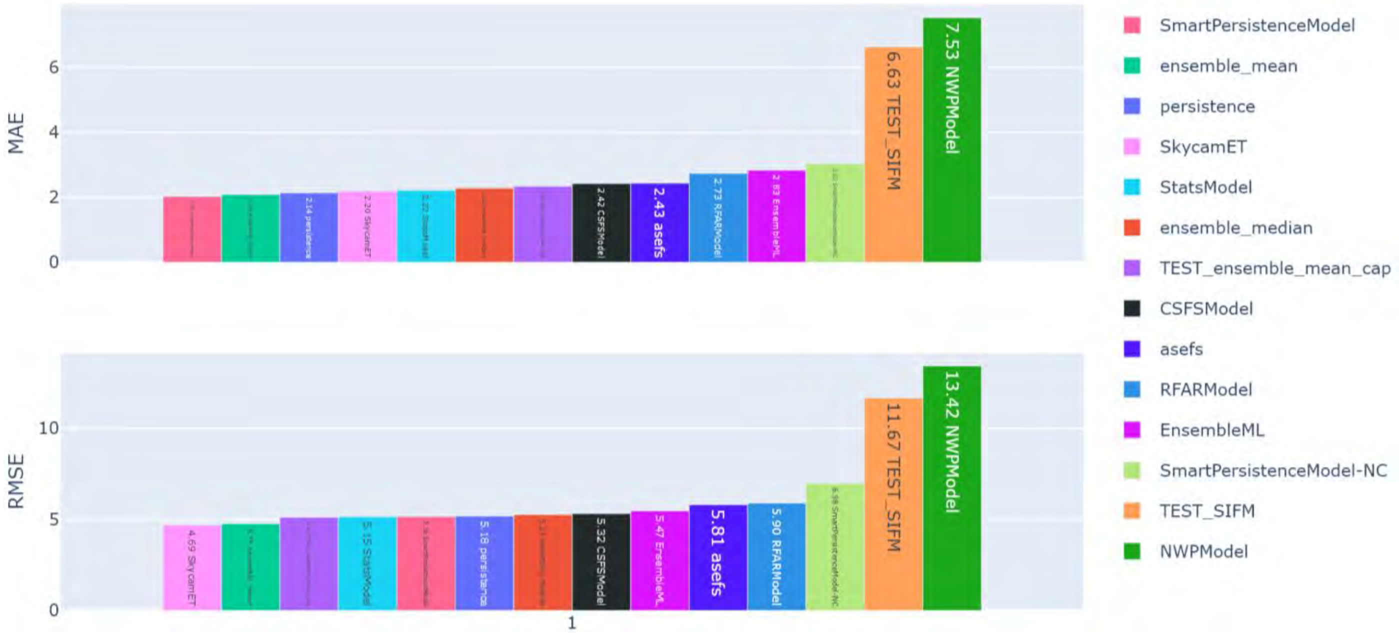
model	mean_gen	start	end	n_days	n	NAs	RMSE Skill % vs ASEFS	RMSE	MAE Skill % vs ASEFS	MAE	NRMSE	NMAE
ensemble_median	59.65	2020-08-01	2020-11-11	102	17310	0	15.03	12.8	**18.19**	**5.96**	0.21	0.1
TEST_ensemble_mean_cap	59.65	2020-08-01	2020-11-11	102	17310	0	**16.23**	**12.62**	16.12	6.11	0.21	0.1
SmartPersistenceModel	59.65	2020-08-01	2020-11-11	102	17310	0	9.5	13.63	15.22	6.17	0.23	0.1
ensemble_mean	59.65	2020-08-01	2020-11-11	102	17310	0	12.8	13.14	13.55	6.29	0.22	0.11
SkycamET	59.65	2020-08-01	2020-11-11	102	17310	0	15.1	12.79	8.47	6.66	0.21	0.11
persistence	59.65	2020-08-01	2020-11-11	102	17310	0	9.1	13.7	8.2	6.68	0.23	0.11
StatsModel	59.65	2020-08-01	2020-11-11	102	17310	0	7.29	13.97	7.15	6.76	0.23	0.11
asefs	59.65	2020-08-01	2020-11-11	102	17310	0	0	15.07	0	7.28	0.25	0.12
CSFSModel	59.65	2020-08-01	2020-11-11	102	17310	0	-8.16	16.29	-0.28	7.3	0.27	0.12
EnsembleML	59.65	2020-08-01	2020-11-11	102	17310	0	6.23	14.13	-0.69	7.33	0.24	0.12
RFARModel	59.65	2020-08-01	2020-11-11	102	17310	0	3.31	14.57	-1.12	7.36	0.24	0.12
SmartPersistenceModel-NC	59.65	2020-08-01	2020-11-11	102	17310	0	-26.96	19.13	-48.36	10.8	0.32	0.18
NWPModel	59.65	2020-08-01	2020-11-11	102	17310	0	-55.95	23.49	-83.77	13.38	0.39	0.22
TEST_SIFM	59.65	2020-08-01	2020-11-11	102	17310	0	-75.95	26.51	-108.98	15.21	0.44	0.26



SF3

Smart Persistence had the lowest MAE at SF3, while Skycam had the lowest RMSE.

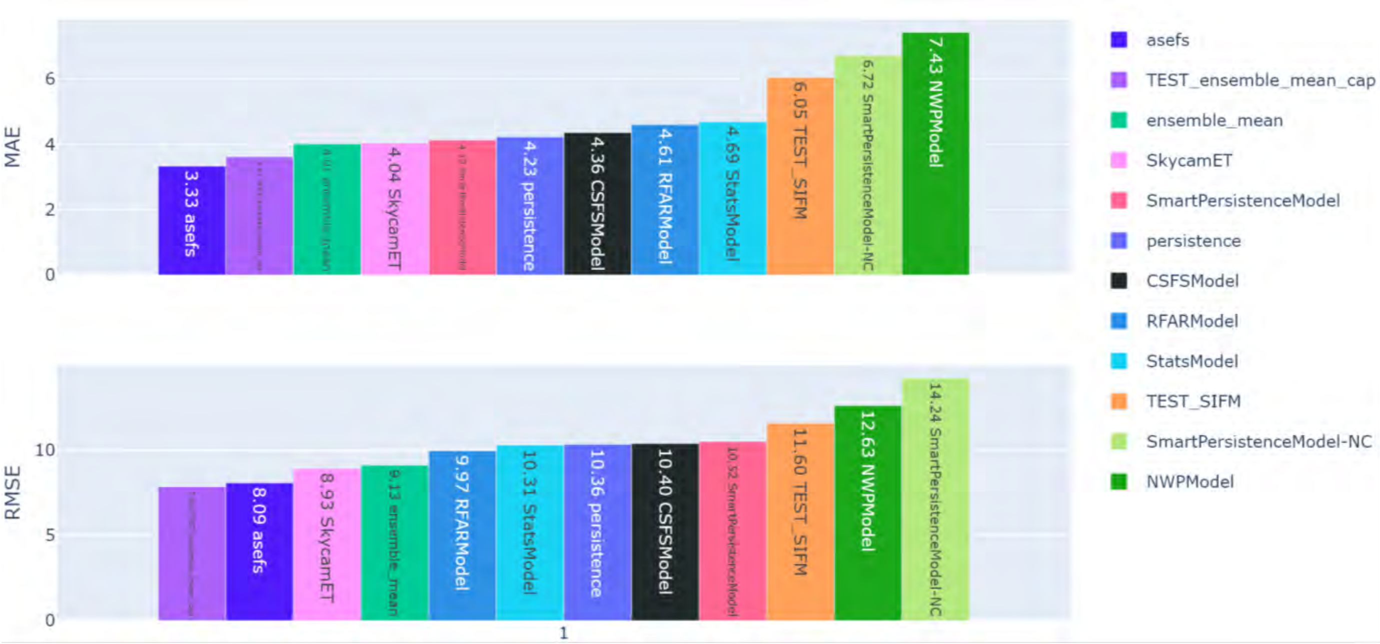
model	mean_gen	start	end	n_days	n	NAs	RMSE Skill % vs ASEFS	RMSE	MAE Skill % vs ASEFS	MAE	NRMSE	NMAE
SmartPersistenceModel	14.49	2020-08-01	2020-11-11	100	17738	0	11.16	5.16	**16.81**	**2.02**	0.36	0.14
ensemble_mean	14.49	2020-08-01	2020-11-11	100	17738	0	18.05	4.76	14.4	2.08	0.33	0.14
persistence	14.49	2020-08-01	2020-11-11	100	17738	0	10.8	5.18	12.22	2.14	0.36	0.15
SkycamET	14.49	2020-08-01	2020-11-11	100	17738	0	**19.31**	**4.69**	9.76	2.2	0.32	0.15
StatsModel	14.49	2020-08-01	2020-11-11	100	17738	0	11.29	5.15	8.91	2.22	0.36	0.15
ensemble_median	14.49	2020-08-01	2020-11-11	100	17738	0	9.33	5.27	6.17	2.28	0.36	0.16
TEST_ensemble_mean_cap	14.49	2020-08-01	2020-11-11	100	17738	0	11.9	5.12	3.75	2.34	0.35	0.16
CSFSModel	14.49	2020-08-01	2020-11-11	100	17738	0	8.45	5.32	0.38	2.42	0.37	0.17
asefs	14.49	2020-08-01	2020-11-11	100	17738	0	0	5.81	0	2.43	0.4	0.17
RFARModel	14.49	2020-08-01	2020-11-11	100	17738	0	-1.53	5.9	-12.37	2.73	0.41	0.19
EnsembleML	14.49	2020-08-01	2020-11-11	100	17738	0	5.81	5.47	-16.48	2.83	0.38	0.2
SmartPersistenceModel-NC	14.49	2020-08-01	2020-11-11	100	17738	0	-20.16	6.98	-24.61	3.03	0.48	0.21
TEST_SIFM	14.49	2020-08-01	2020-11-11	100	17738	0	-100.83	11.67	-172.52	6.63	0.8	0.46
NWPModel	14.49	2020-08-01	2020-11-11	100	17738	0	-131.03	13.42	-209.4	7.53	0.93	0.52



SF4

ASEFS had the lowest MAE at SF4, mostly due to the issues with SCADA data and frequent curtailments. Mean Ensemble had the lowest RMSE over this period.

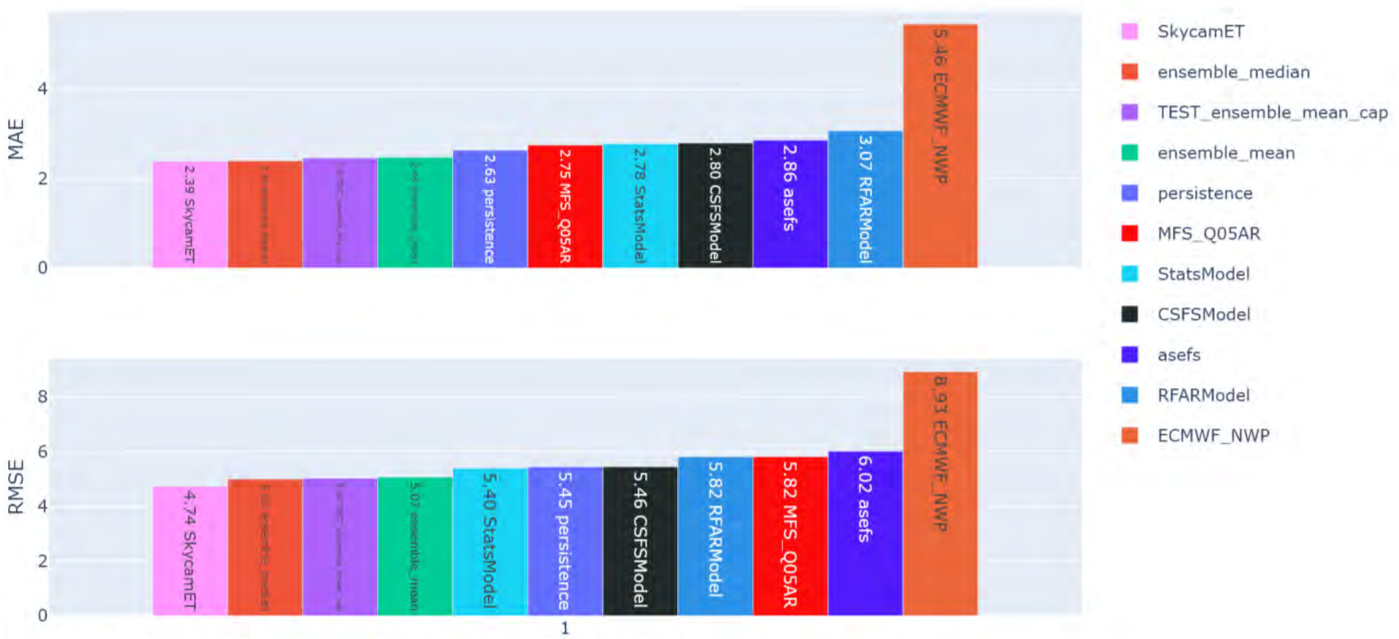
model	mean_gen	start	end	n_days	n	NAs	RMSE Skill % vs ASEFS	RMSE	MAE Skill % vs ASEFS	MAE	NRMSE	NMAE
asefs	28.3	2020-08-01	2020-11-11	91	14003	0	0	8.09	**0**	**3.33**	0.29	0.12
TEST_ensemble_mean_cap	28.3	2020-08-01	2020-11-11	91	14003	0	**2.84**	**7.86**	-8.61	3.62	0.28	0.13
ensemble_mean	28.3	2020-08-01	2020-11-11	91	14003	0	-12.81	9.13	-20.44	4.01	0.32	0.14
SkycamET	28.3	2020-08-01	2020-11-11	91	14003	0	-10.33	8.93	-21.27	4.04	0.32	0.14
SmartPersistenceModel	28.3	2020-08-01	2020-11-11	91	14003	0	-29.99	10.52	-24.04	4.13	0.37	0.15
persistence	28.3	2020-08-01	2020-11-11	91	14003	0	-27.99	10.36	-27	4.23	0.37	0.15
CSFSModel	28.3	2020-08-01	2020-11-11	91	14003	0	-28.53	10.4	-30.88	4.36	0.37	0.15
RFARModel	28.3	2020-08-01	2020-11-11	91	14003	0	-23.25	9.97	-38.22	4.61	0.35	0.16
StatsModel	28.3	2020-08-01	2020-11-11	91	14003	0	-27.39	10.31	-40.61	4.69	0.36	0.17
TEST_SIFM	28.3	2020-08-01	2020-11-11	91	14003	0	-43.3	11.6	-81.57	6.05	0.41	0.21
SmartPersistenceModel-NC	28.3	2020-08-01	2020-11-11	91	14003	0	-76.03	14.24	-101.61	6.72	0.5	0.24
NWPModel	28.3	2020-08-01	2020-11-11	91	14003	0	-56.08	12.63	-122.84	7.43	0.45	0.26



SF5

Unlike the other sites, just over a month of data with all models was available to SF5 at the time of writing. Skycam had the lowest RMSE and MAE in this period. All ensembles and individual models except RFAR and NWP outperformed ASEFS.

model	mean_gen	filter	start	end	n_days	n	NAs	RMSE Skill % vs ASEFS	RMSE	MAE Skill % vs ASEFS	MAE	NRMSE	NMAE
SkycamET	17.02		2020-09-26	2020-11-11	42	7408	0	**21.22**	**4.74**	**16.56**	**2.39**	0.28	0.14
ensemble_median	17.02		2020-09-26	2020-11-11	42	7408	0	16.92	5	16.21	2.4	0.29	0.14
TEST_ensemble_mean_cap	17.02		2020-09-26	2020-11-11	42	7408	0	16.35	5.04	13.98	2.46	0.3	0.14
ensemble_mean	17.02		2020-09-26	2020-11-11	42	7408	0	15.72	5.07	13.38	2.48	0.3	0.15
persistence	17.02		2020-09-26	2020-11-11	42	7408	0	9.4	5.45	7.97	2.63	0.32	0.15
MFS_Q05AR	17.02		2020-09-26	2020-11-11	42	7408	0	3.29	5.82	3.88	2.75	0.34	0.16
StatsModel	17.02		2020-09-26	2020-11-11	42	7408	0	10.27	5.4	2.81	2.78	0.32	0.16
CSFSModel	17.02		2020-09-26	2020-11-11	42	7408	0	9.36	5.46	2.27	2.8	0.32	0.16
asefs	17.02		2020-09-26	2020-11-11	42	7408	0	0	6.02	0	2.86	0.35	0.17
RFARModel	17.02		2020-09-26	2020-11-11	42	7408	0	3.34	5.82	-7.3	3.07	0.34	0.18
ECMWF_NWP	17.02		2020-09-26	2020-11-11	42	7408	0	-48.34	8.93	-90.74	5.46	0.52	0.32

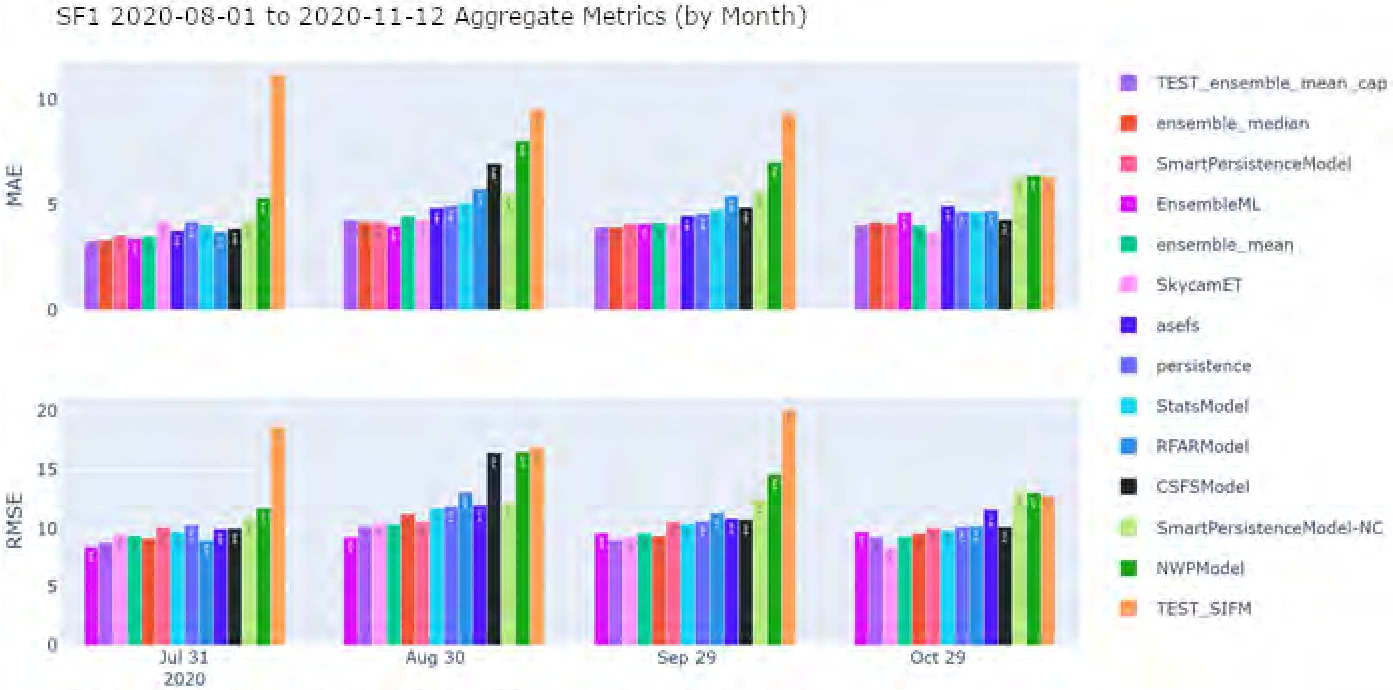


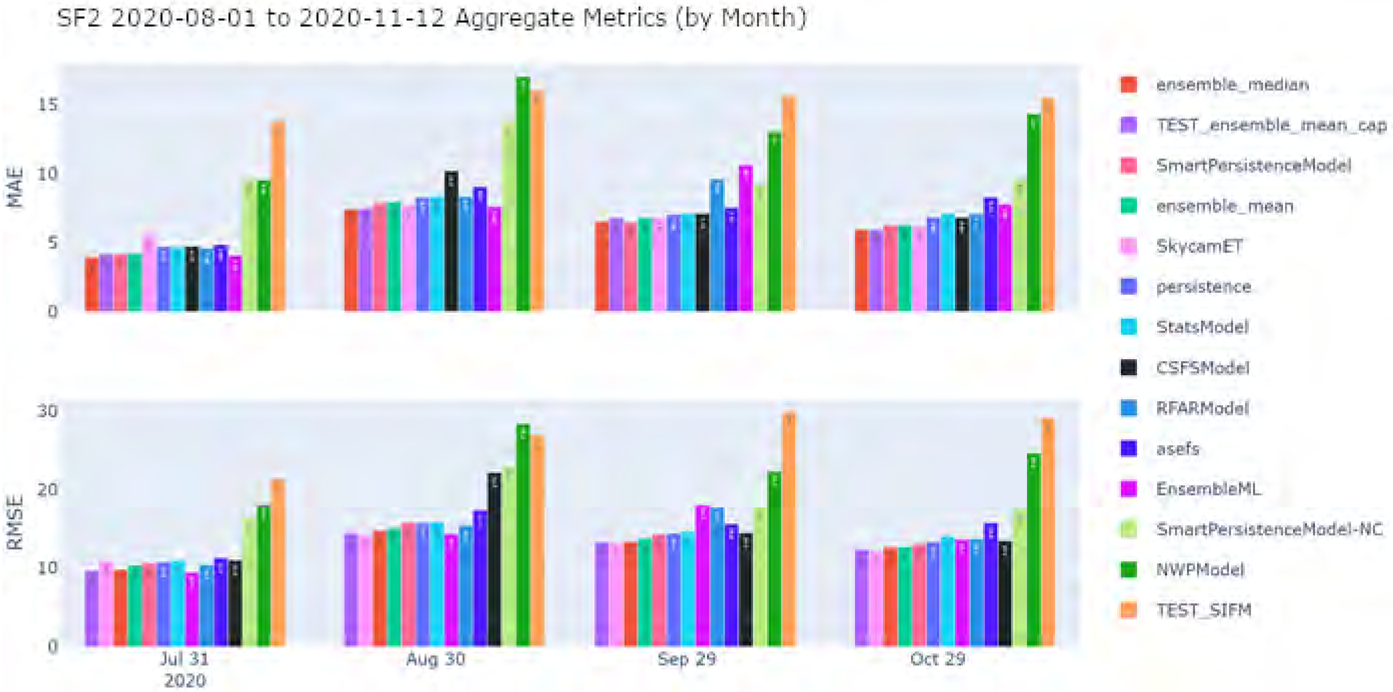
5.3.1.2 By month

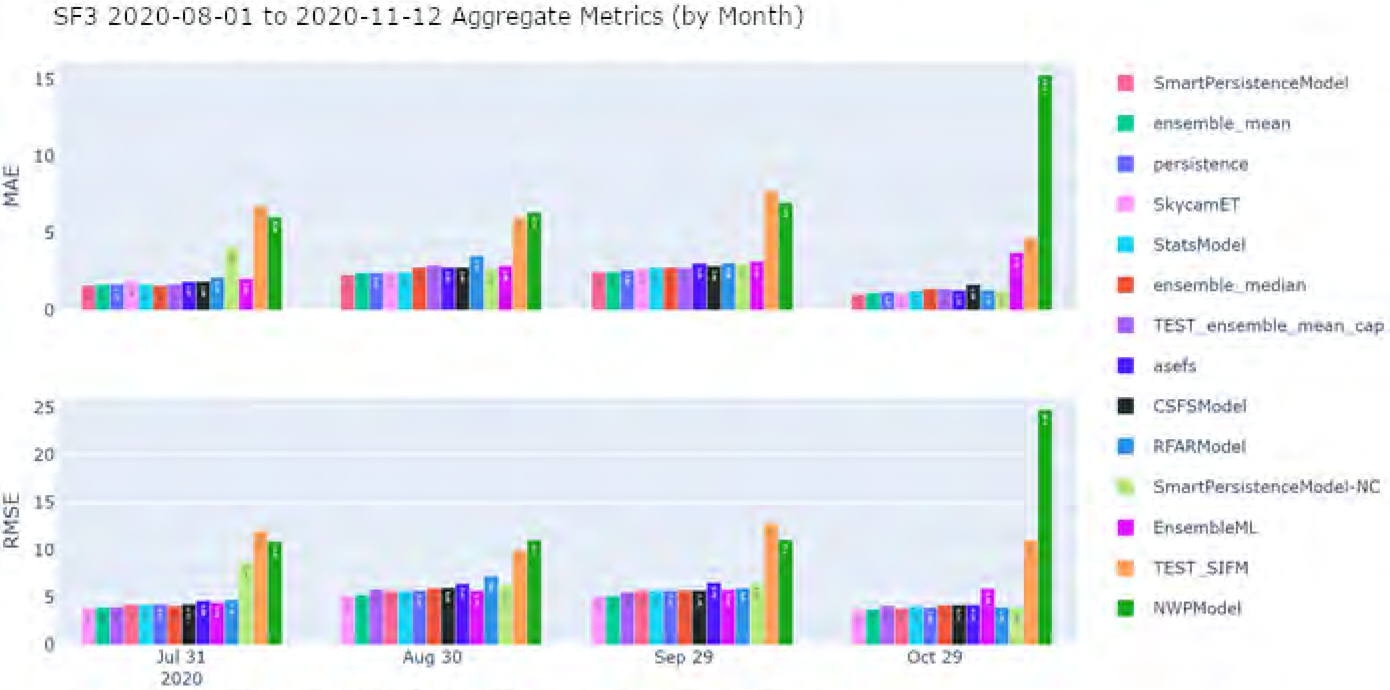
These plots show forecast performance for each model per month. Overall results are the same as the previous section but show the change in performance over time.

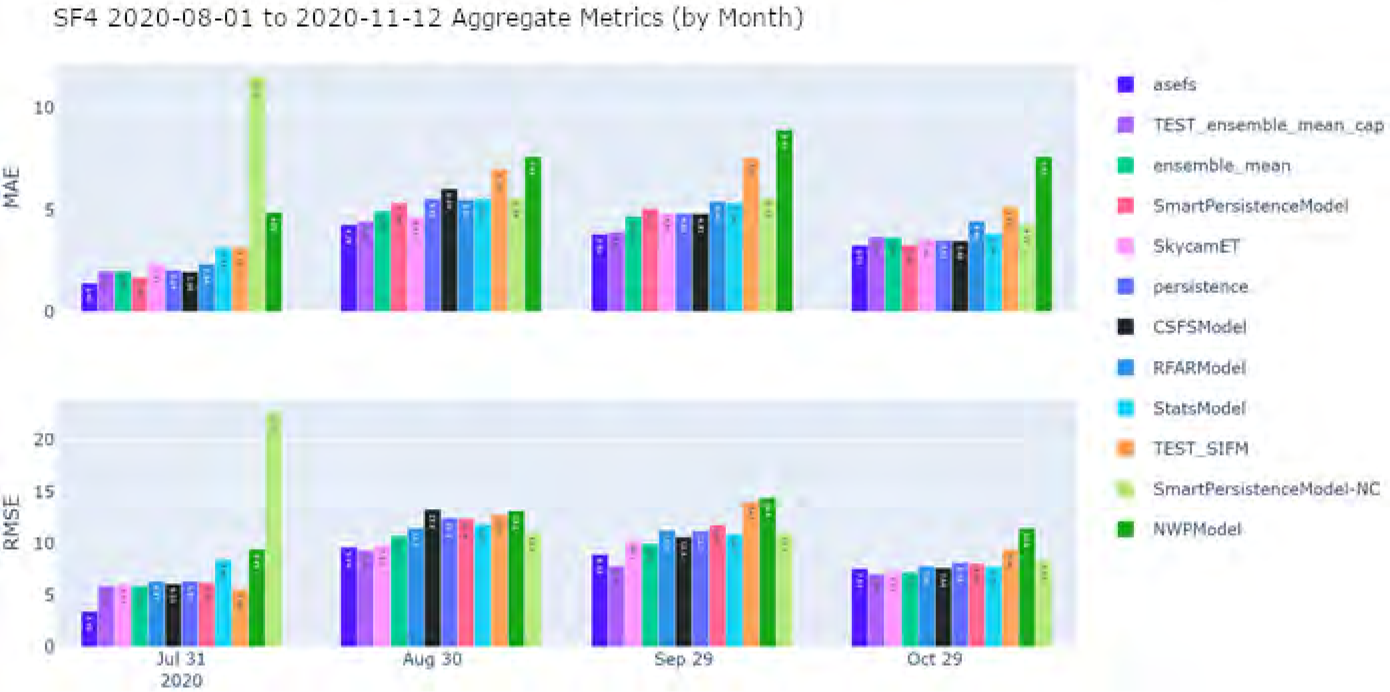
Analysis:

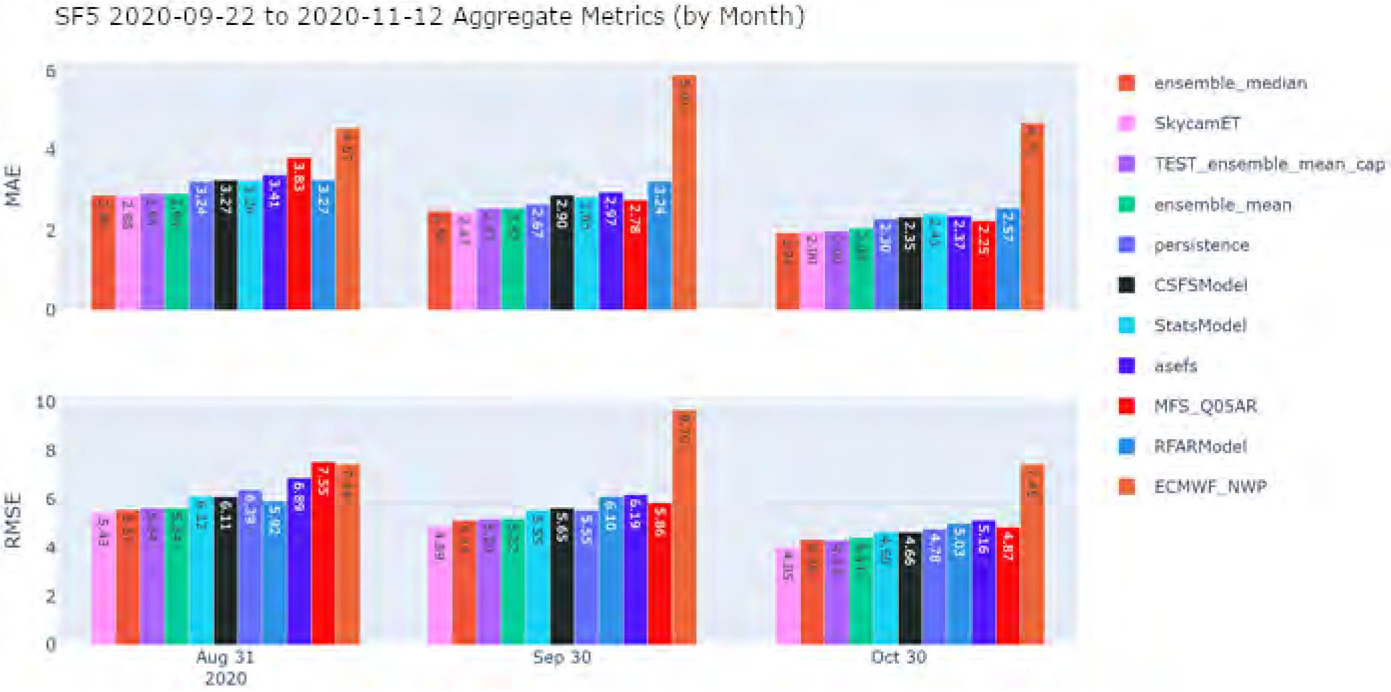
- Skycam performance improved from August to October relative to other models. This is probably due to the increasing solar intermittency associated with seasonal changes.
- SIFM showed a large improvement in October at SF1, SF3 and SF4 due to changes made to the code in late September. Improvements to the fitting functions used in converting cloud index to clear sky index and systematic biases in the clear sky model used in SIFM were implemented.


SF1 2020-08-01 to 2020-11-12 Aggregate Metrics (by Month)
MAE
RMSE
Jul 31
2020
Aug 30
Sep 29
Oct 29
TEST_ensemble_mean_cap
ensemble_median
SmartPersistenceModel
EnsembleML
ensemble_mean
SkycamET
asefs
persistence
StatsModel
RFARModel
CSFSModel
SmartPersistenceModel-NC
NWPModel
TEST_SIFM


SF2 2020-08-01 to 2020-11-12 Aggregate Metrics (by Month)
MAE
RMSE
Jul 31
2020
Aug 30
Sep 29
Oct 29
ensemble_median
TEST_ensemble_mean_cap
SmartPersistenceModel
ensemble_mean
SkycamET
persistence
StatsModel
CSFSModel
RFARModel
asefs
EnsembleML
SmartPersistenceModel-NC
NWPModel
TEST_SIFM


SF3 2020-08-01 to 2020-11-12 Aggregate Metrics (by Month)
MAE
RMSE
Jul 31
2020
Aug 30
Sep 29
Oct 29
SmartPersistenceModel
ensemble_mean
persistence
SkycamET
StatsModel
ensemble_median
TEST_ensemble_mean_cap
asefs
CSFSModel
RFARModel
SmartPersistenceModel-NC
EnsembleML
TEST_SIFM
NWPModel


SF4 2020-08-01 to 2020-11-12 Aggregate Metrics (by Month)
MAE
RMSE
Jul 31
2020
Aug 30
Sep 29
Oct 29
asefs
TEST_ensemble_mean_cap
ensemble_mean
SmartPersistenceModel
SkycamET
persistence
CSFSModel
RFARModel
StatsModel
TEST_SIFM
SmartPersistenceModel-NC
NWPModel



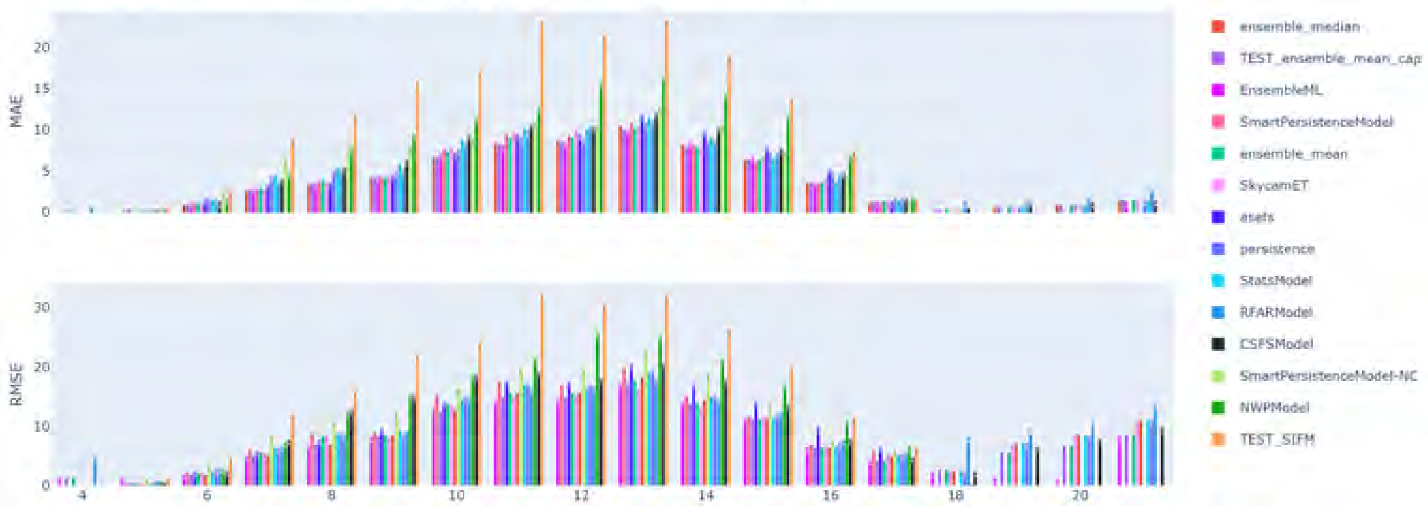
5.3.1.3 By hour of day

These plots show MAE and RMSE aggregated per hour of day over this 3-month period, to investigate relative model performance at different times of day.

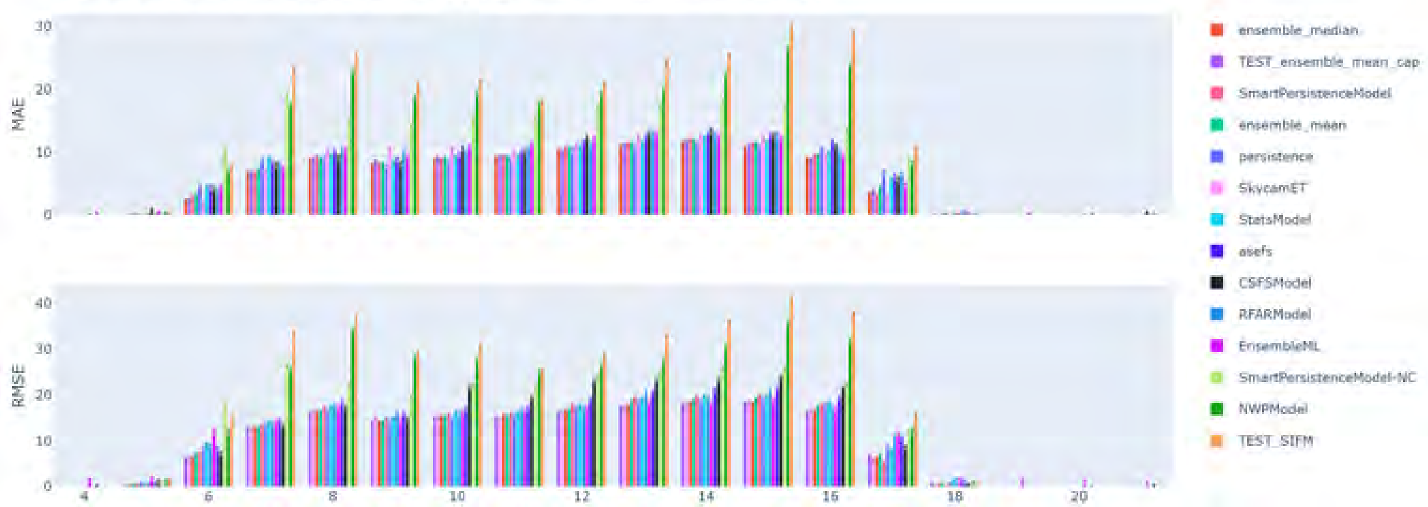
Analysis:

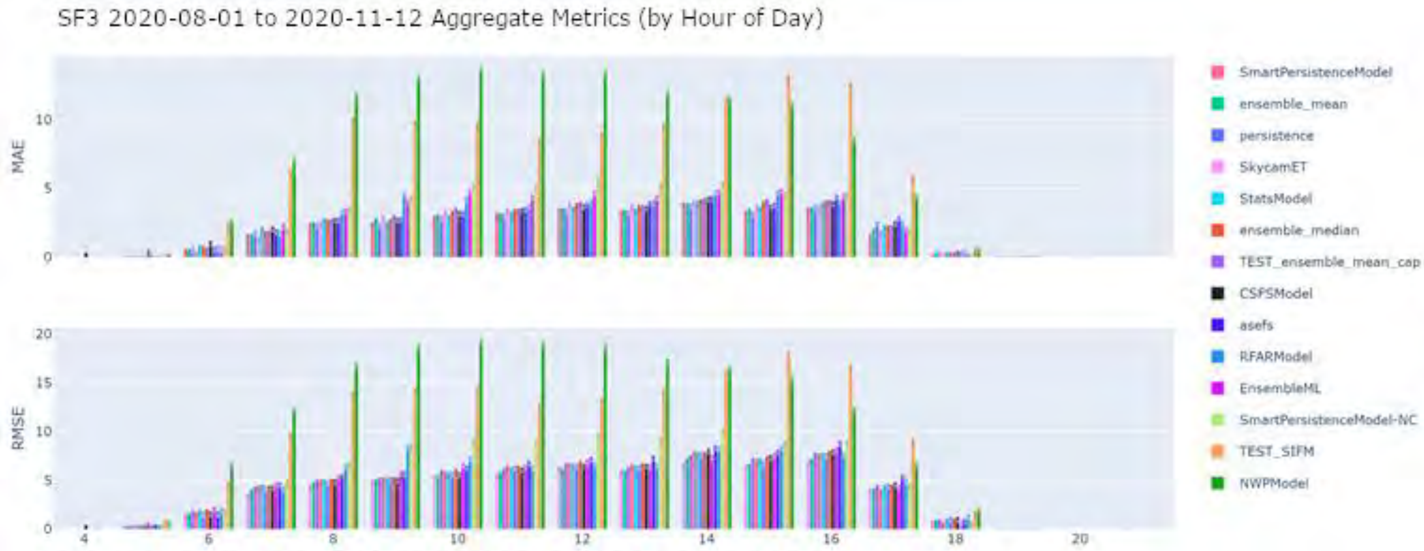
- ASEFS generally performs well in the mornings, but less so in the (often more intermittent) afternoons.
- Some models were predicting small non-zero values for sun-down hours at SF1 and SF2 due to training against SCADA power rather than INITIALMW data. These models have now been corrected.
- SIFM performed well at SF4 in many intervals, probably due to the relative advantage of not requiring unconstrained power input data during the frequent semi-dispatch cap periods.

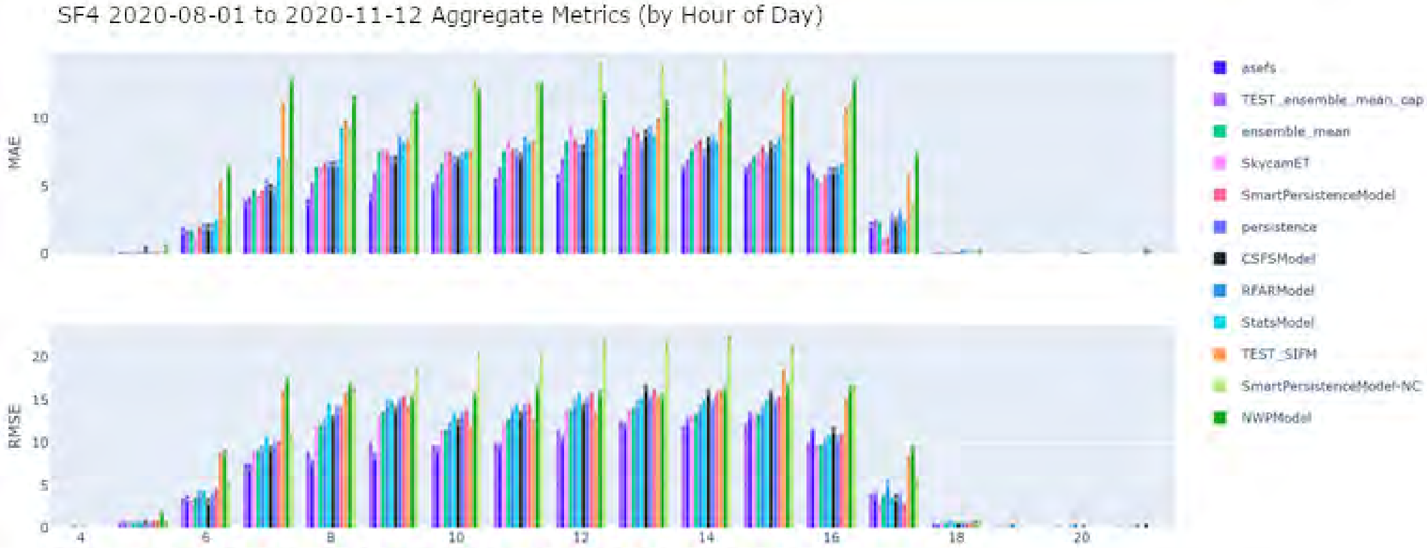
SF1 2020-08-01 to 2020-11-12 Aggregate Metrics (by Hour of Day)

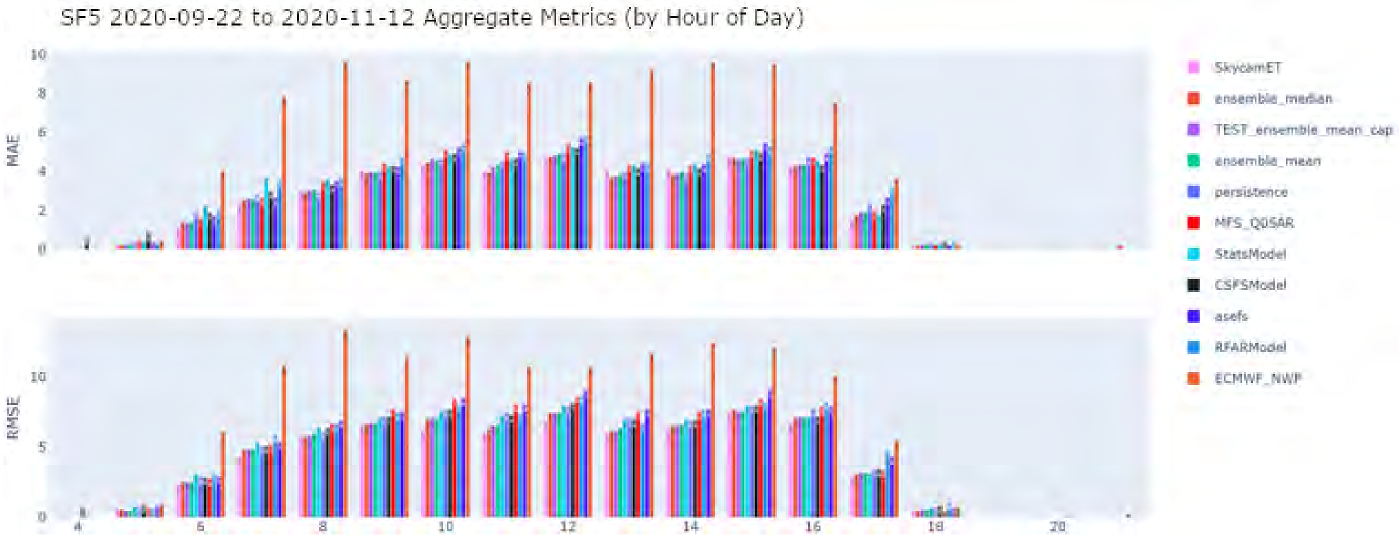


SF2 2020-08-01 to 2020-11-12 Aggregate Metrics (by Hour of Day)




SF3 2020-08-01 to 2020-11-12 Aggregate Metrics (by Hour of Day)
MAE
RMSE
SmartPersistenceModel
ensemble_mean
persistence
SkycamET
StatsModel
ensemble_median
TEST_ensemble_mean_cap
CSFSModel
asefs
RFARModel
EnsembleML
SmartPersistenceModel-NC
TEST_SIFM
NWPModel


SF4 2020-08-01 to 2020-11-12 Aggregate Metrics (by Hour of Day)
MAE
RMSE
asefs
TEST_ensemble_mean_cap
ensemble_mean
SkycamET
SmartPersistenceModel
persistence
CSFSModel
RFARModel
StatsModel
TEST_SIFM
SmartPersistenceModel-NC
NWPModel

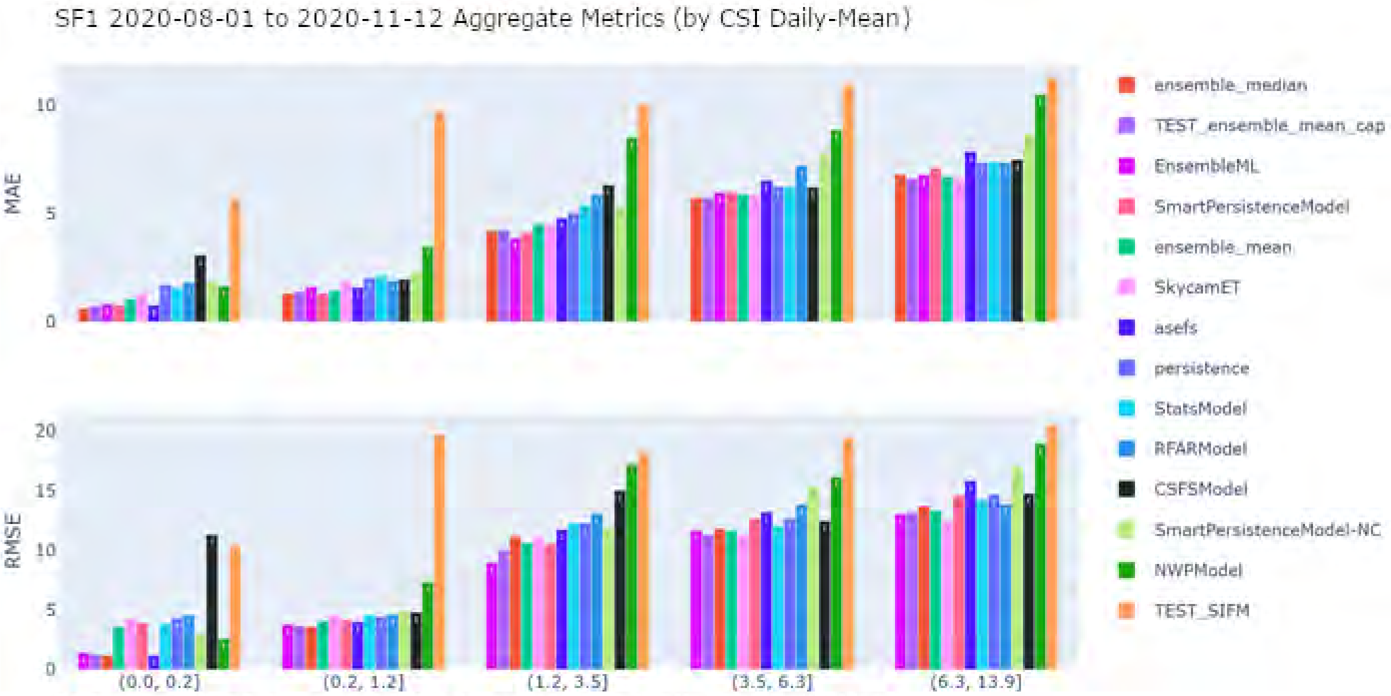

SF5 2020-09-22 to 2020-11-12 Aggregate Metrics (by Hour of Day)
MAE
RMSE
SkycamET
ensemble_median
TEST_ensemble_mean_cap
ensemble_mean
persistence
MFS_QOSAR
StatsModel
CSFSModel
asefs
RFARModel
ECMWF_NWP

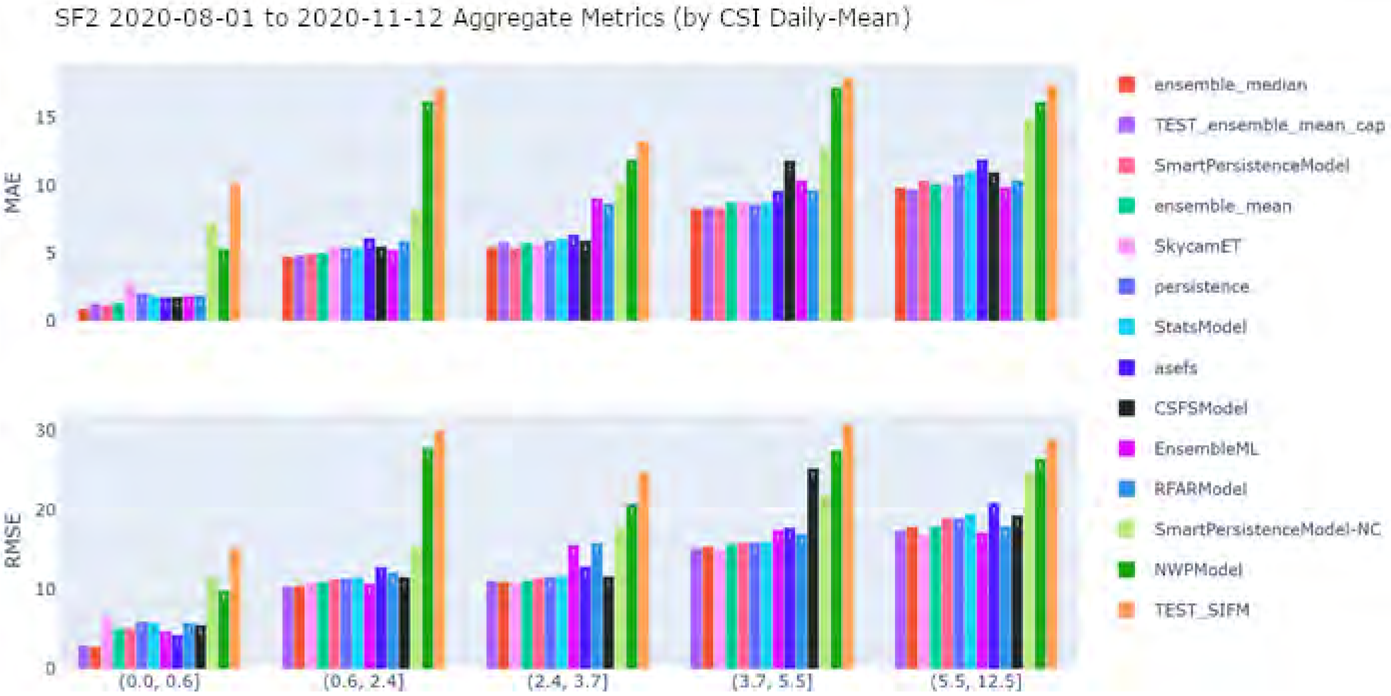
5.3.1.4 By binned Clear Sky Index (CSI)

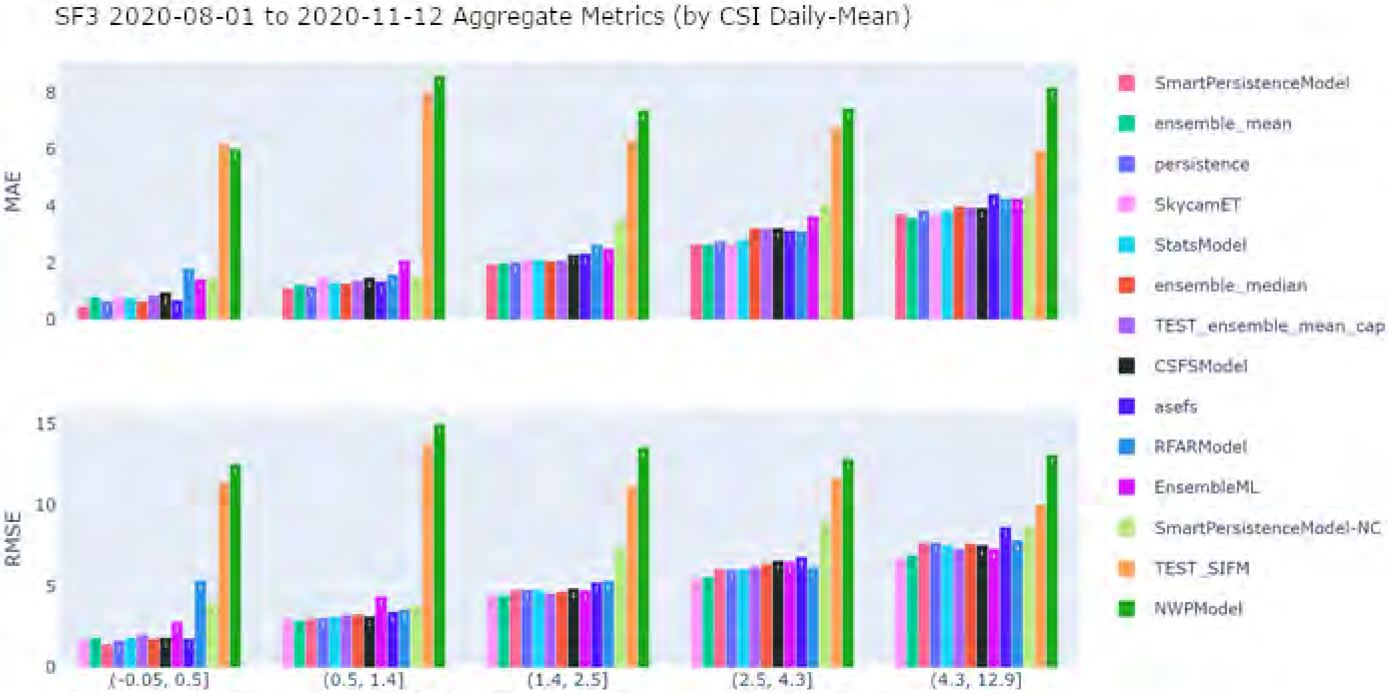
These plots show model performance split into Clear Sky Index (CSI) bins. Clear Sky Index is a measure of how much the observed irradiance fluctuated relative to the clear sky irradiance curve and is a good indicator of how intermittent the irradiance was. CSI values close to zero (left side of plots) indicate clear-sky conditions, and higher values (right side of plots) indicate higher intermittency, or statistical variance, in the observed irradiance. We have selected five bins, or ranges, of CSI to evaluate model performance in different weather conditions for each day.

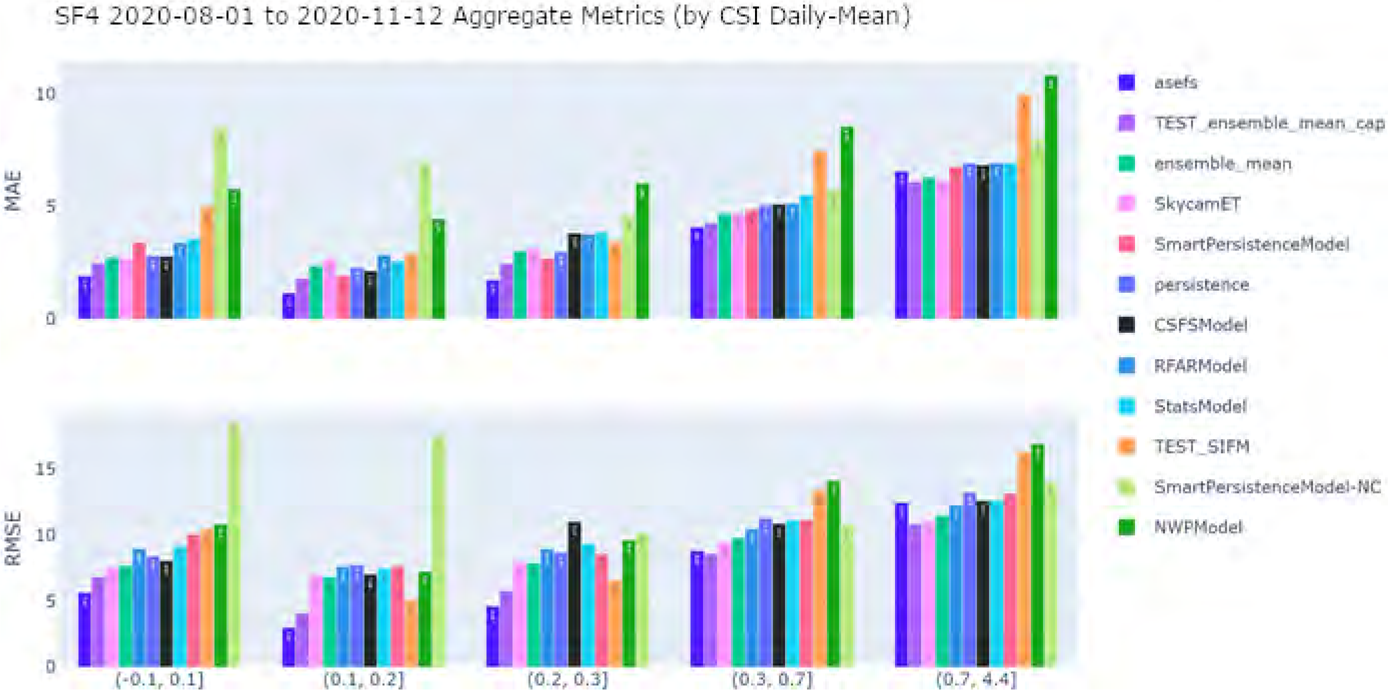
Analysis:

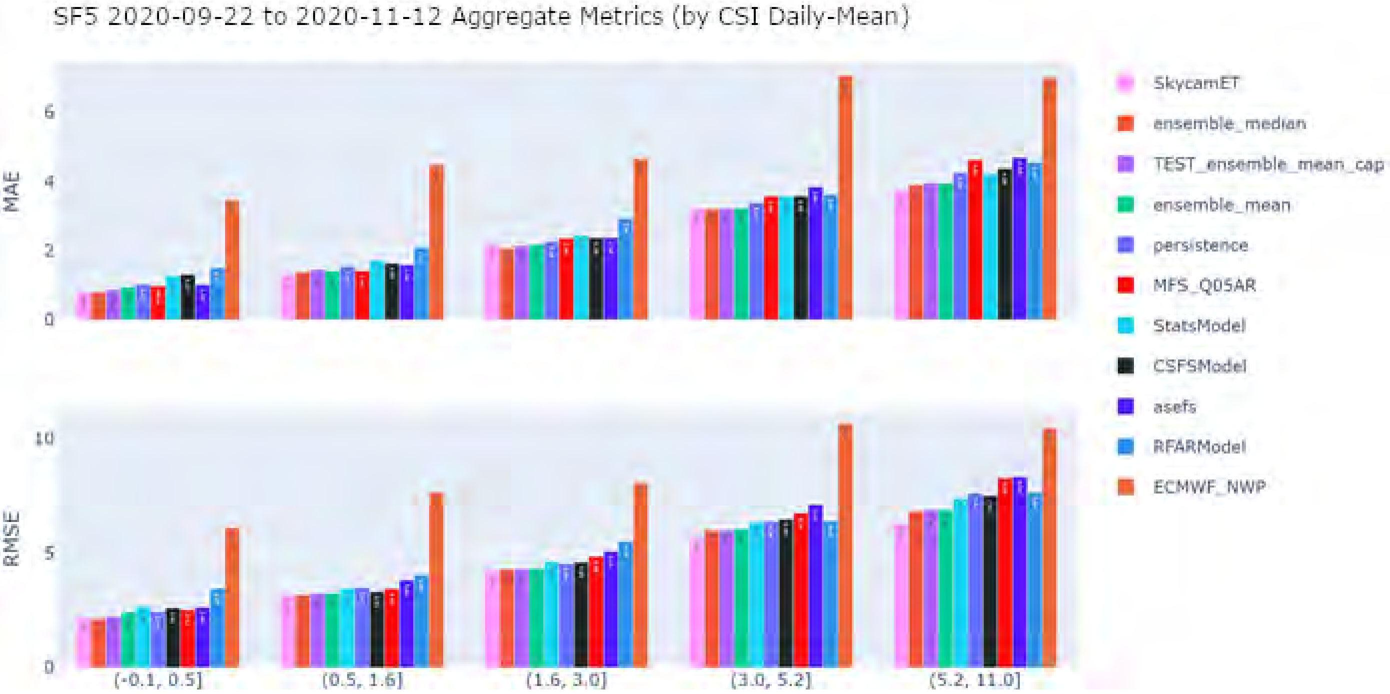
- ASEFS generally performs well at in clear sky (low CSI) conditions but is less accurate than other models on more intermittent days.
- Skycam is reasonably accurate on clear sky days but performs best in highly intermittent (high CSI) conditions. It had the lowest RMSE in the top CSI bin for all sites except SF4.
- There was no other single model that outperformed the others across all sites, though one of the ensemble models was in the top two models for every CSI bin for every site, indicating that the ensembles models adapt well to changing weather conditions.


SF1 2020-08-01 to 2020-11-12 Aggregate Metrics (by CSI Daily-Mean)
MAE
RMSE
(0.0, 0.2]
(0.2, 1.2]
(1.2, 3.5]
(3.5, 6.3]
(6.3, 13.9]
ensemble_median
TEST_ensemble_mean_cap
EnsembleML
SmartPersistenceModel
ensemble_mean
SkycamET
asefs
persistence
StatsModel
RFARModel
CSFSModel
SmartPersistenceModel-NC
NWPModel
TEST_SIFM


SF2 2020-08-01 to 2020-11-12 Aggregate Metrics (by CSI Daily-Mean)
MAE
RMSE
(0.0, 0.6]
(0.6, 2.4]
(2.4, 3.7]
(3.7, 5.5]
(5.5, 12.5]
ensemble_median
TEST_ensemble_mean_cap
SmartPersistenceModel
ensemble_mean
SkycamET
persistence
StatsModel
asefs
CSFSModel
EnsembleML
RFARModel
SmartPersistenceModel-NC
NWPModel
TEST_SIFM


SF3 2020-08-01 to 2020-11-12 Aggregate Metrics (by CSI Daily-Mean)
MAE
RMSE
(-0.05, 0.5]
(0.5, 1.4]
(1.4, 2.5]
(2.5, 4.3]
(4.3, 12.9]
SmartPersistenceModel
ensemble_mean
persistence
SkycamET
StatsModel
ensemble_median
TEST_ensemble_mean_cap
CSFSModel
asefs
RFARModel
EnsembleML
SmartPersistenceModel-NC
TEST_SIFM
NWPModel


SF4 2020-08-01 to 2020-11-12 Aggregate Metrics (by CSI Daily-Mean)
MAE
RMSE
(-0.1, 0.1]
(0.1, 0.2]
(0.2, 0.3]
(0.3, 0.7]
(0.7, 4.4]
asefs
TEST_ensemble_mean_cap
ensemble_mean
SkycamET
SmartPersistenceModel
persistence
CSFSModel
RFARModel
StatsModel
TEST_SIFM
SmartPersistenceModel-NC
NWPModel



5.3.2 Maximum period

The 'maximum period' was chosen to be 2020-05-01 to 2020-11-11, as this six-month period captured the first eight models to come online for most sites, allowing a direct comparison of each model at specific site and a reasonable comparison between sites. Note that SF5 is omitted from these results are its first models didn't come online until September 2020.

5.3.2.1 Overall

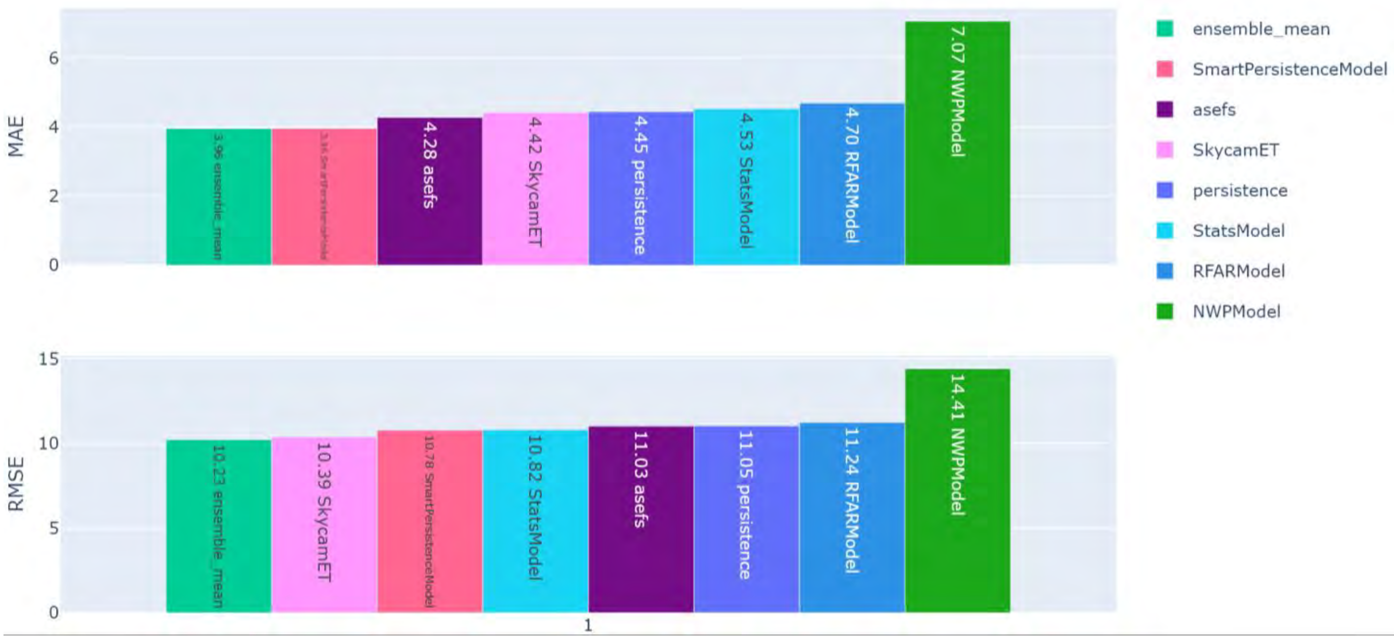
These results show the performance of and model in MAE and RMSE, aggregated over this period per site.

Analysis:

- Generally, the Ensemble Mean and Smart Persistence Model were the most accurate forecast models, outperforming ASEFS by a significant margin in both metrics at all sites except SF4 (due to frequent semi-dispatch cap issues outlined above).
- Skycam was in the top two models for RMSE at all sites except SF4 (where it was omitted from the results, as it didn't start running early enough).

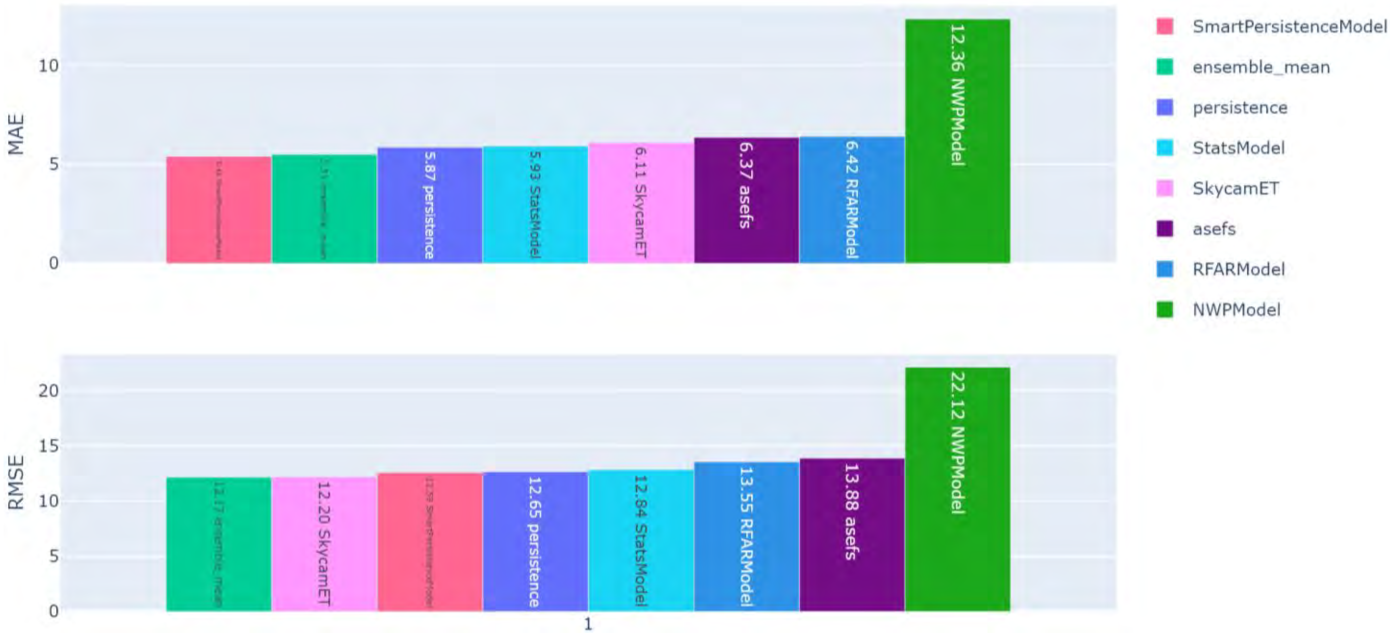
SF1

site	model	mean_gen	start	end	n_days	n	RMSE	RMSE Skill % vs ASEFS	MAE	MAE Skill % vs ASEFS	NRMSE	NMAE
SF11	ensemble_mean	36.99	2020-06-02	2020-11-11	157	28934	10.23	7.26	3.96	7.52	0.28	0.11
SF11	SmartPersistenceModel	36.99	2020-06-02	2020-11-11	157	28934	10.78	2.22	3.96	7.4	0.29	0.11
SF11	asefs	36.99	2020-06-02	2020-11-11	157	28934	11.03	0	4.28	0	0.3	0.12
SF11	SkycamET	36.99	2020-06-02	2020-11-11	157	28934	10.39	5.81	4.42	-3.37	0.28	0.12
SF11	persistence	36.99	2020-06-02	2020-11-11	157	28934	11.05	-0.19	4.45	-3.95	0.3	0.12
SF11	StatsModel	36.99	2020-06-02	2020-11-11	157	28934	10.82	1.92	4.53	-5.9	0.29	0.12
SF11	RFARModel	36.99	2020-06-02	2020-11-11	157	28934	11.24	-1.97	4.7	-9.77	0.3	0.13
SF11	NWPModel	36.99	2020-06-02	2020-11-11	157	28934	14.41	-30.7	7.07	-65.1	0.39	0.19



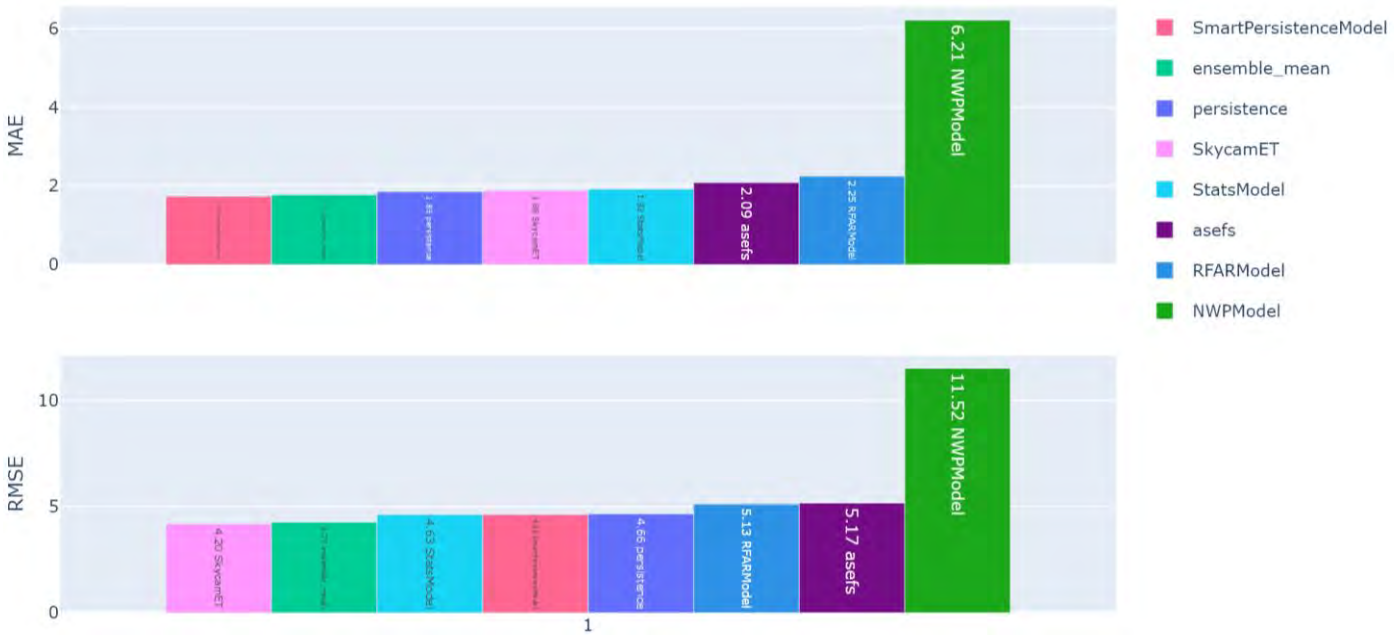
SF2

site	model	mean_gen	start	end	n_days	n	RMSE	RMSE Skill % vs ASEFS	MAE	MAE Skill % vs ASEFS	NRMSE	NMAE
SF21	SmartPersistenceModel	54.36	2020-06-02	2020-11-11	160	28541	12.59	9.34	5.41	15.09	0.23	0.1
SF21	ensemble_mean	54.36	2020-06-02	2020-11-11	160	28541	12.17	12.35	5.51	13.6	0.22	0.1
SF21	Persistence	54.36	2020-06-02	2020-11-11	160	28541	12.65	8.86	5.87	7.82	0.23	0.11
SF21	StatsModel	54.36	2020-06-02	2020-11-11	160	28541	12.84	7.55	5.93	6.97	0.24	0.11
SF21	SkycamET	54.36	2020-06-02	2020-11-11	160	28541	12.2	12.13	6.11	4.09	0.22	0.11
SF21	asefs	54.36	2020-06-02	2020-11-11	160	28541	13.88	0	6.37	0	0.26	0.12
SF21	RFARModel	54.36	2020-06-02	2020-11-11	160	28541	13.55	2.41	6.42	-0.69	0.25	0.12
SF21	NWPModel	54.36	2020-06-02	2020-11-11	160	28541	22.12	-59.31	12.36	-93.89	0.41	0.23



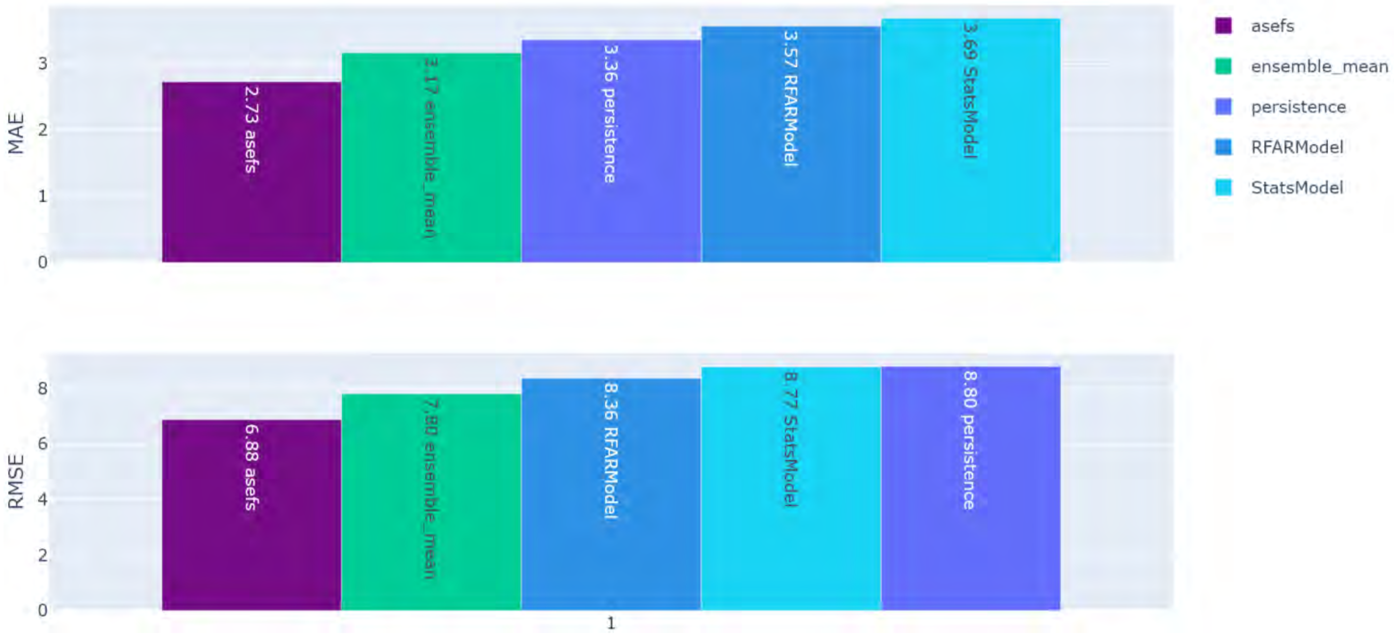
SF3

site	model	mean_gen	start	end	n_days	n	RMSE	RMSE Skill % vs ASEFS	MAE	MAE Skill % vs ASEFS	NRMSE	NMAE
SF31	SmartPersistenceModel	13.13	2020-06-02	2020-11-11	157	28947	4.63	10.38	1.74	16.74	0.35	0.13
SF31	ensemble_mean	13.13	2020-06-02	2020-11-11	157	28947	4.27	17.4	1.77	14.96	0.33	0.14
SF31	persistence	13.13	2020-06-02	2020-11-11	157	28947	4.66	9.83	1.85	11.15	0.36	0.14
SF31	SkycamET	13.13	2020-06-02	2020-11-11	157	28947	4.2	18.85	1.88	9.78	0.32	0.14
SF31	StatsModel	13.13	2020-06-02	2020-11-11	157	28947	4.63	10.44	1.92	8.02	0.35	0.15
SF31	asefs	13.13	2020-06-02	2020-11-11	157	28947	5.17	0	2.09	0	0.39	0.16
SF31	RFARModel	13.13	2020-06-02	2020-11-11	157	28947	5.13	0.86	2.25	-7.72	0.39	0.17
SF31	NWPModel	13.13	2020-06-02	2020-11-11	157	28947	11.52	-122.71	6.21	-197.68	0.88	0.47



SF4

site	model	mean_gen	start	end	n_days	n	RMSE	RMSE Skill % vs ASEFS	MAE	MAE Skill % vs ASEFS	NRMSE	NMAE
SF41	asefs	24.23	2020-05-07	2020-11-11	178	31368	6.88	0	2.73	0	0.28	0.11
SF41	ensemble_mean	24.23	2020-05-07	2020-11-11	178	31368	7.8	-13.48	3.17	-16.06	0.32	0.13
SF41	persistence	24.23	2020-05-07	2020-11-11	178	31368	8.8	-27.89	3.36	-23.21	0.36	0.14
SF41	RFARModel	24.23	2020-05-07	2020-11-11	178	31368	8.36	-21.61	3.57	-30.76	0.35	0.15
SF41	StatsModel	24.23	2020-05-07	2020-11-11	178	31368	8.77	-27.55	3.69	-35.13	0.36	0.15



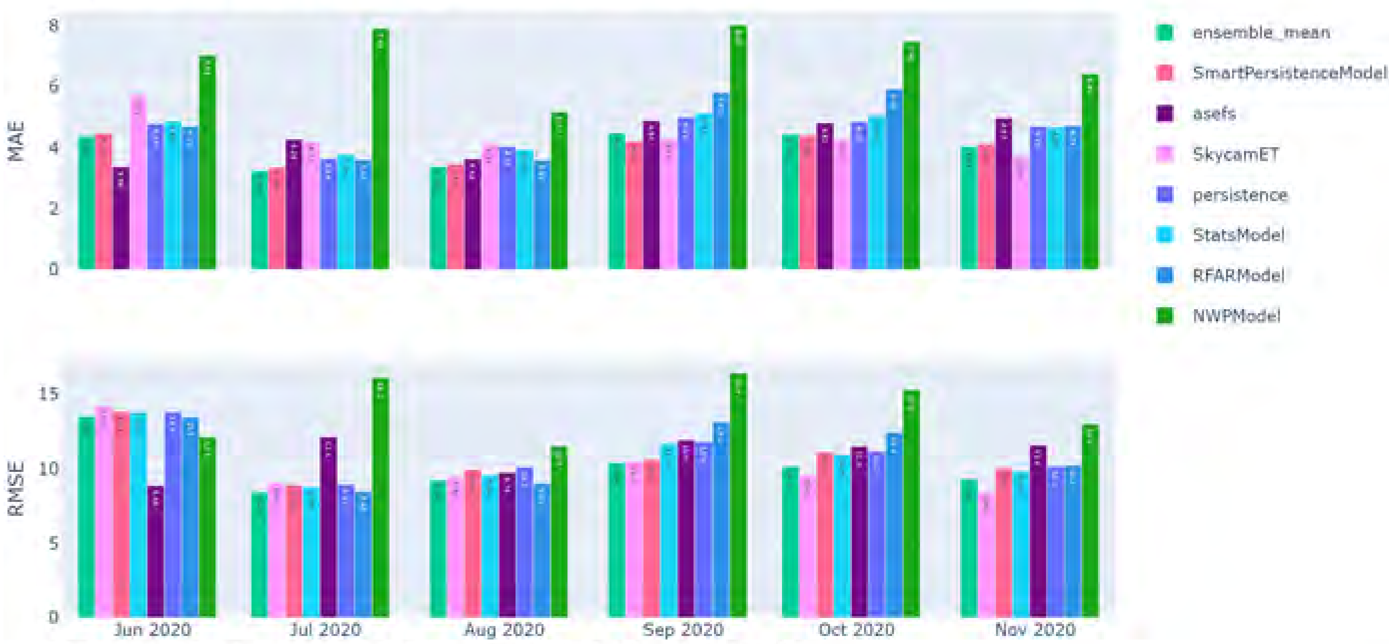
5.3.2.2 By month

These plots show forecast performance for each model per month. Overall results are the same as the previous section but show the change in performance over time.

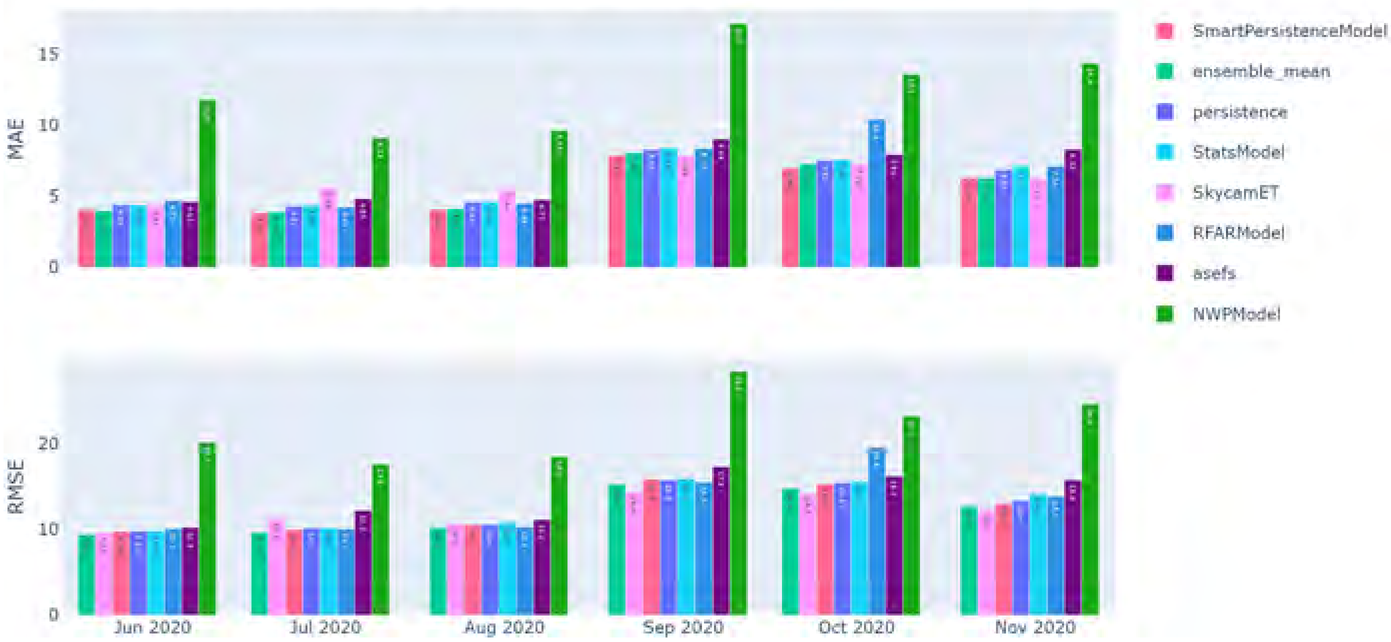
Analysis:

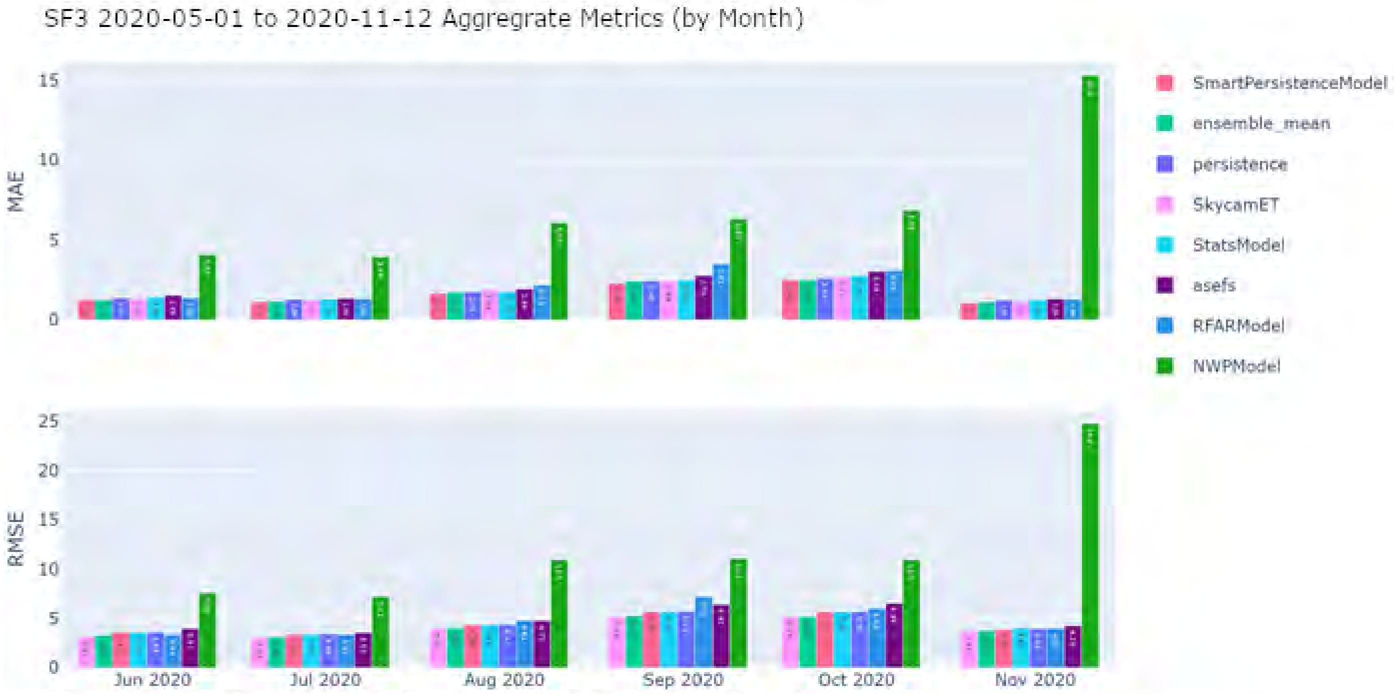
- The Skycam model generally improved relative to the other models at SF1 and SF2 over this period as data quality issues were identified and corrected, nightly training came online, and hyper-parameter tuning was introduced. September/October also exhibited higher solar intermittency as the weather changed, which favoured the Skycam forecasts. In contrast, Skycam performance at SF3 was more consistent, as the data quality issues were not as prevalent at this site.
- The persistence and autoregression models (Stats, RFAR, SmartPersistence) all performed consistently over this period, generally doing well in the winter period when clear sky days dominated. An update to RFAR in September that accidentally introduced a five-minute forecast delay caused performance to suffer but was corrected in mid-October.

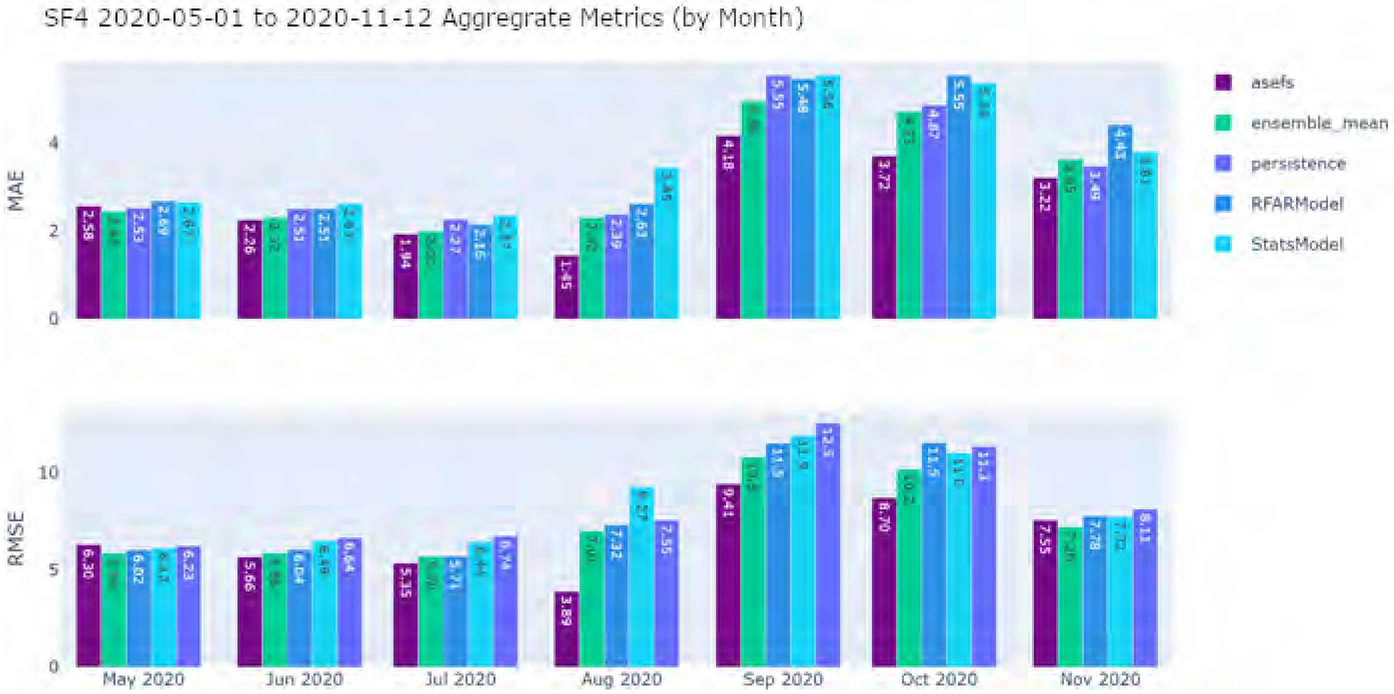
SF1 2020-05-01 to 2020-11-12 Aggregrate Metrics (by Month)



SF2 2020-05-01 to 2020-11-12 Aggregrate Metrics (by Month)







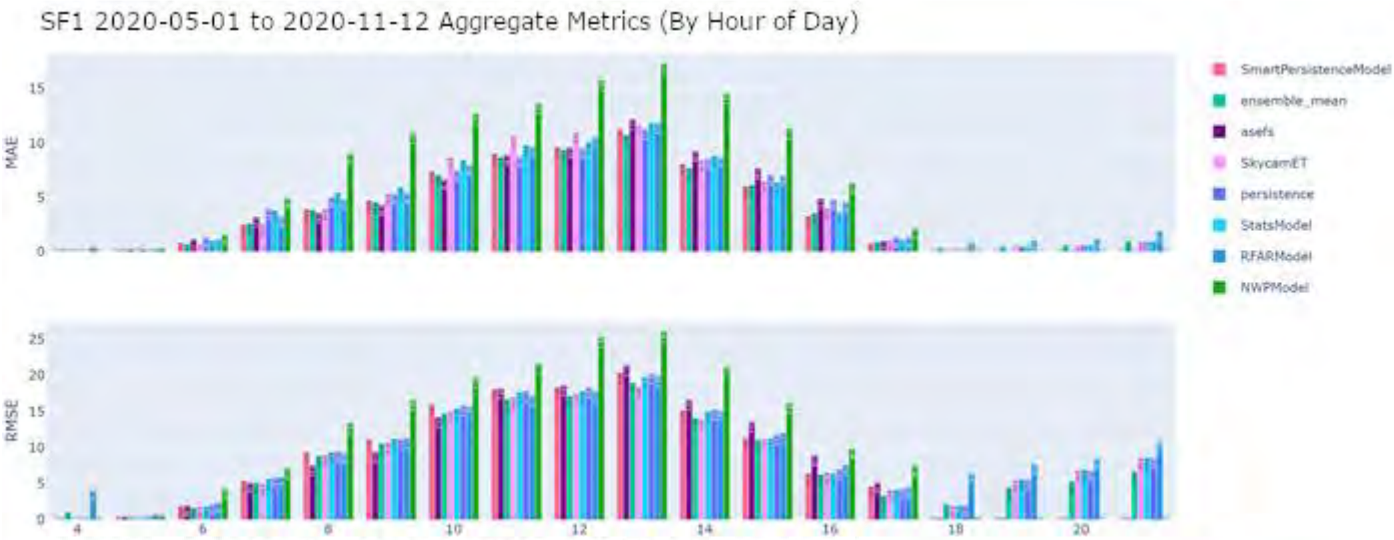
5.3.2.3 By hour of day

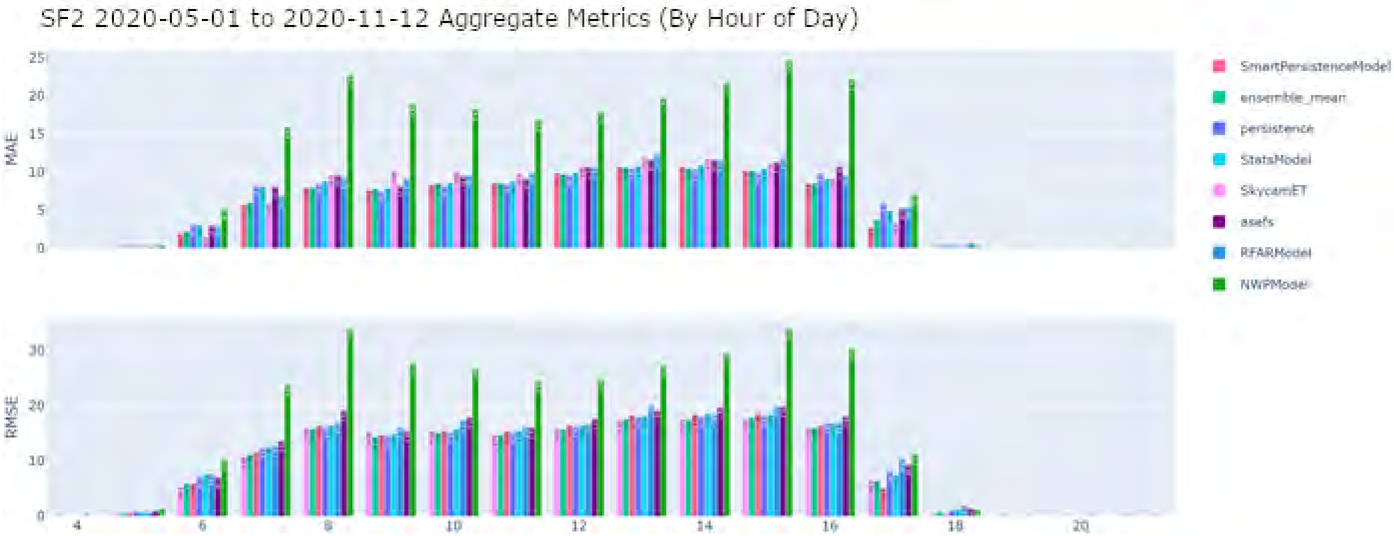
These plots show MAE and RMSE aggregated per hour of day over this whole six-month period, to investigate relative model performance at different times of day.

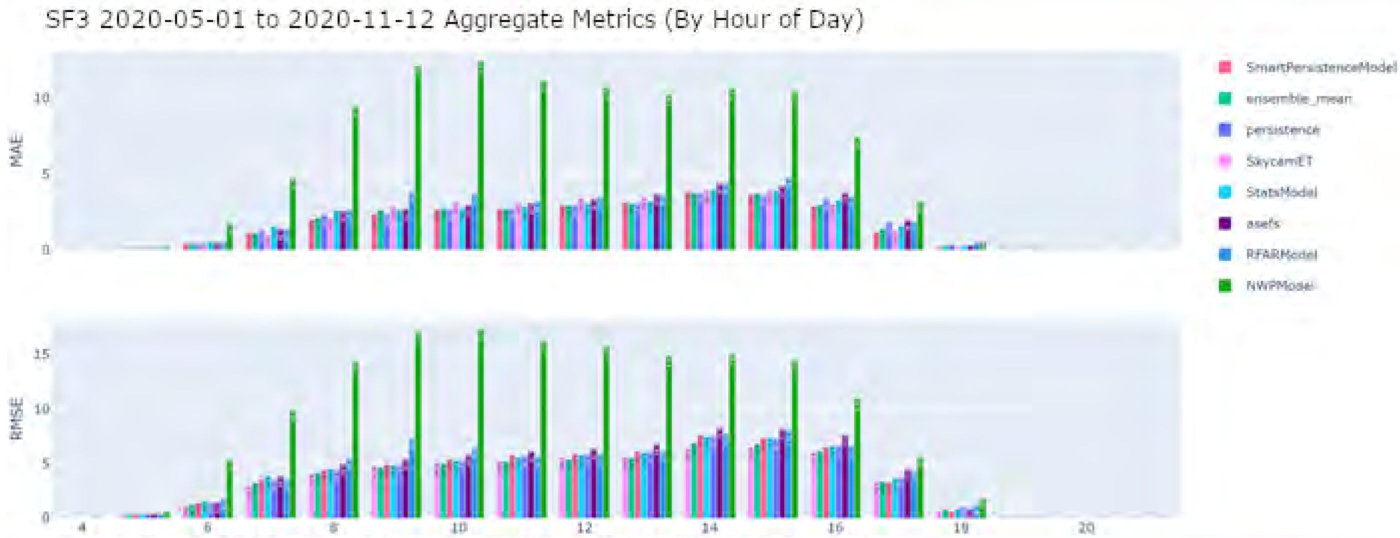
Analysis:

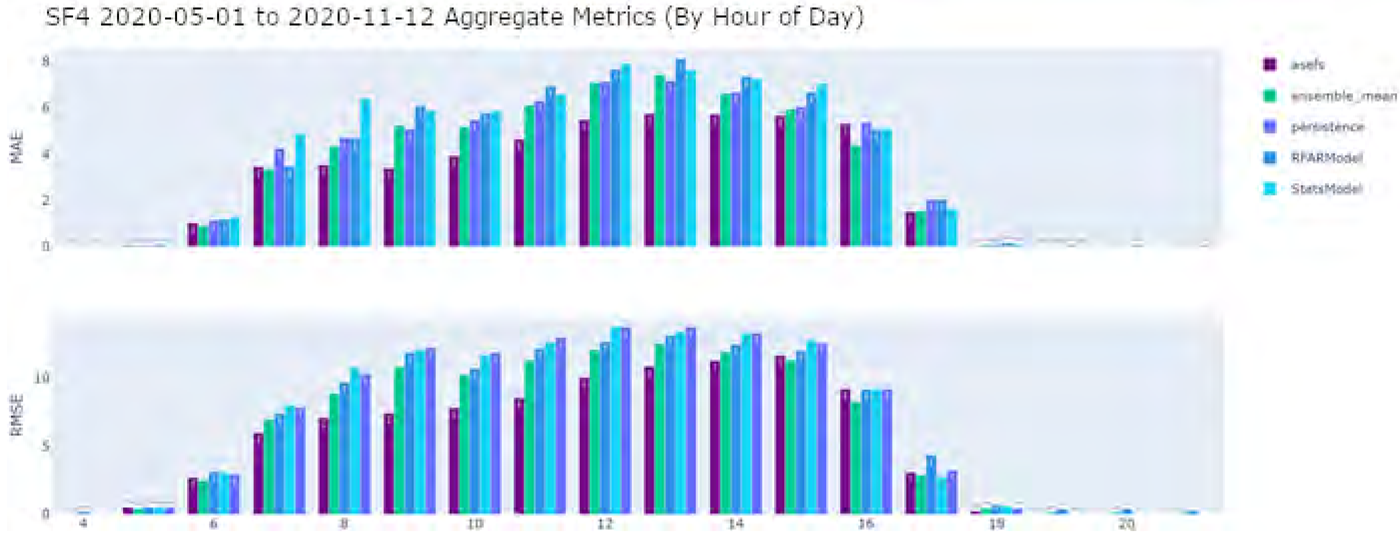
- ASEFS generally does well in the mornings, which tend to have clearer skies, possibly because this is where bias correction is most important but was not running in our models for a large part of this period. Other models do better in the afternoons, which are generally more intermittent.
- Initially, models using SCADA power as targets predicted small negative power values at SF1 during sun-down hours, resulting in errors before 6am and after 6pm. This was corrected in July but is still included in the aggregates shown on the plot below.

- Smart Persistence predicts well in the afternoons, particularly in the last two daylight hours, where it outperformed other models at SF2 and SF3 in both metrics.









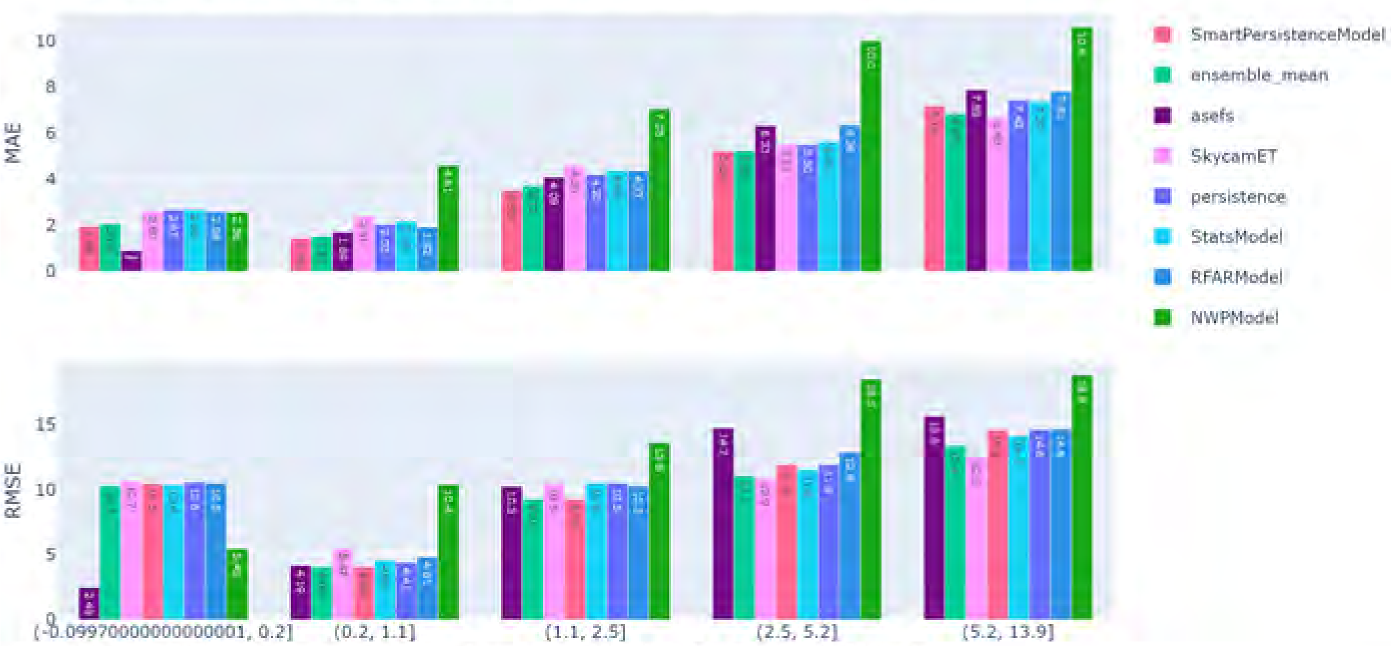
5.3.2.4 By binned Clear Sky Index (CSI)

These plots show model performance split into Clear Sky Index (CSI) bins. CSI values close to zero (left side of plots) indicate clear-sky conditions, and higher values (right side of plots) indicate higher intermittency, or statistical variance, in the observed irradiance.

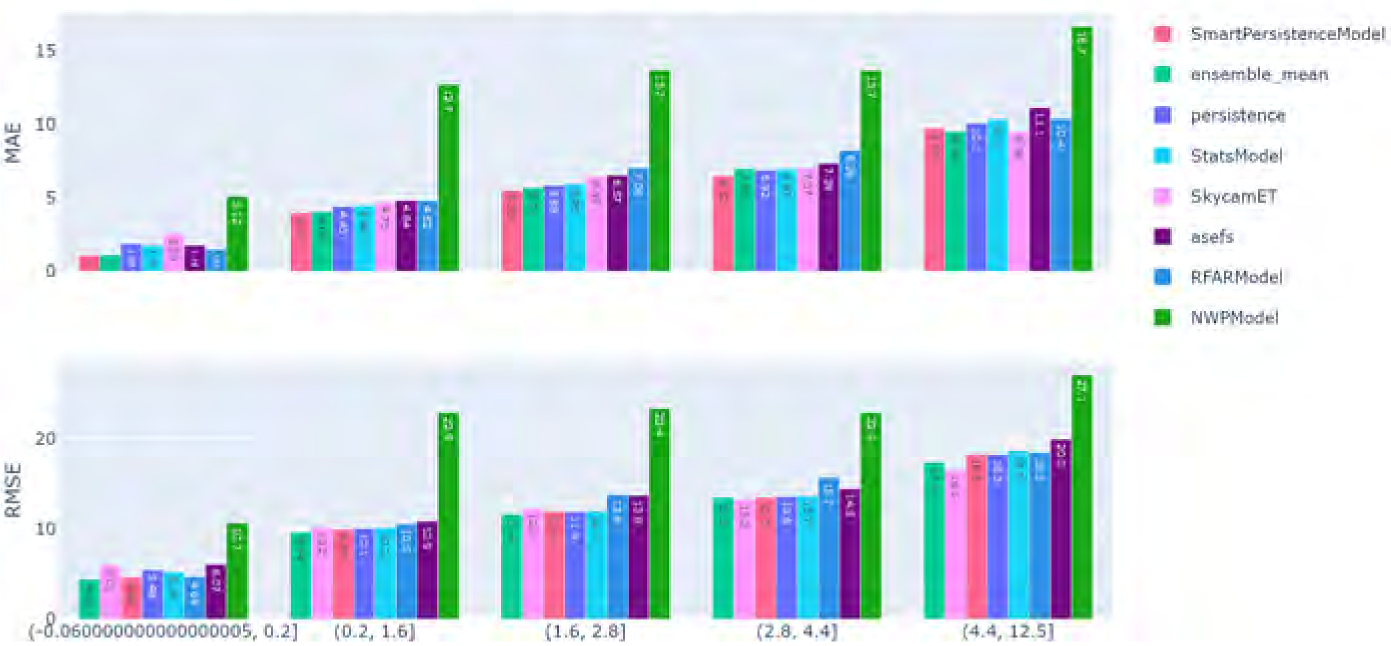
Analysis:

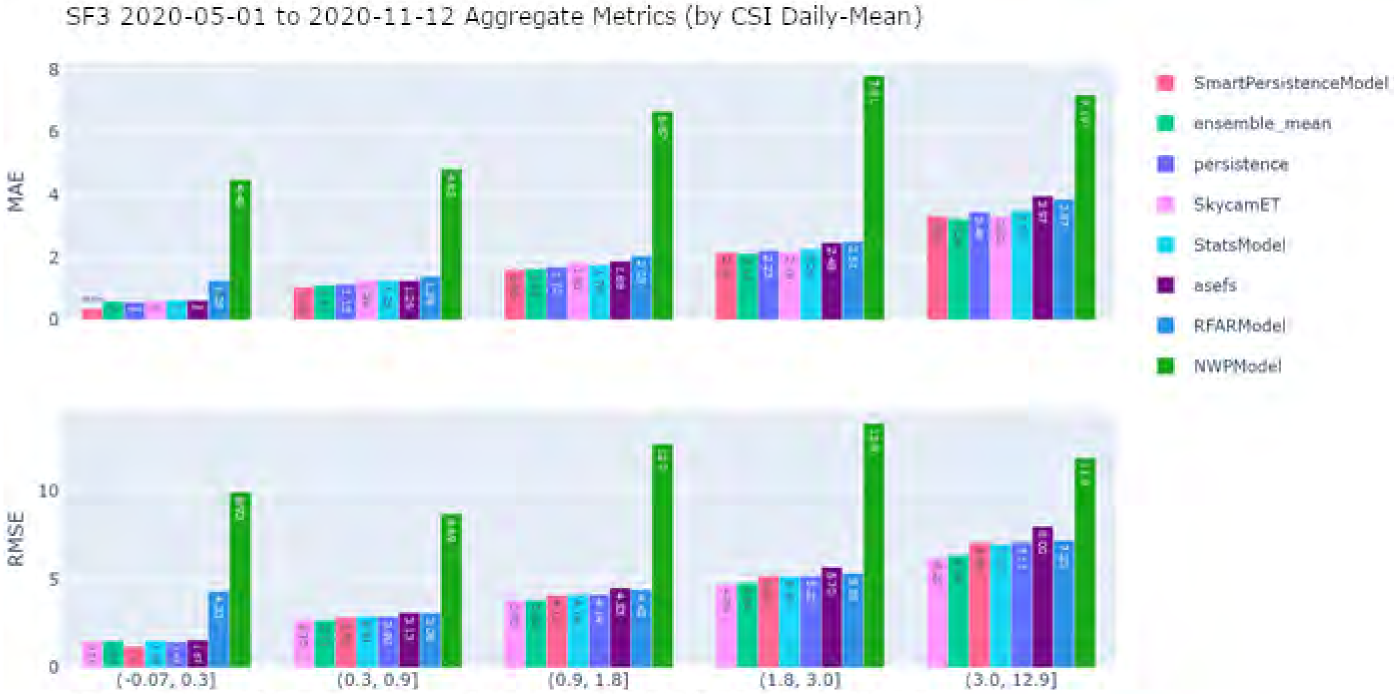
- Skycam is more accurate, relative to other models, on more intermittent (higher CSI) days.
- The other models tend to perform well in the lower CSI brackets, particularly Ensemble Mean and Smart Persistence.

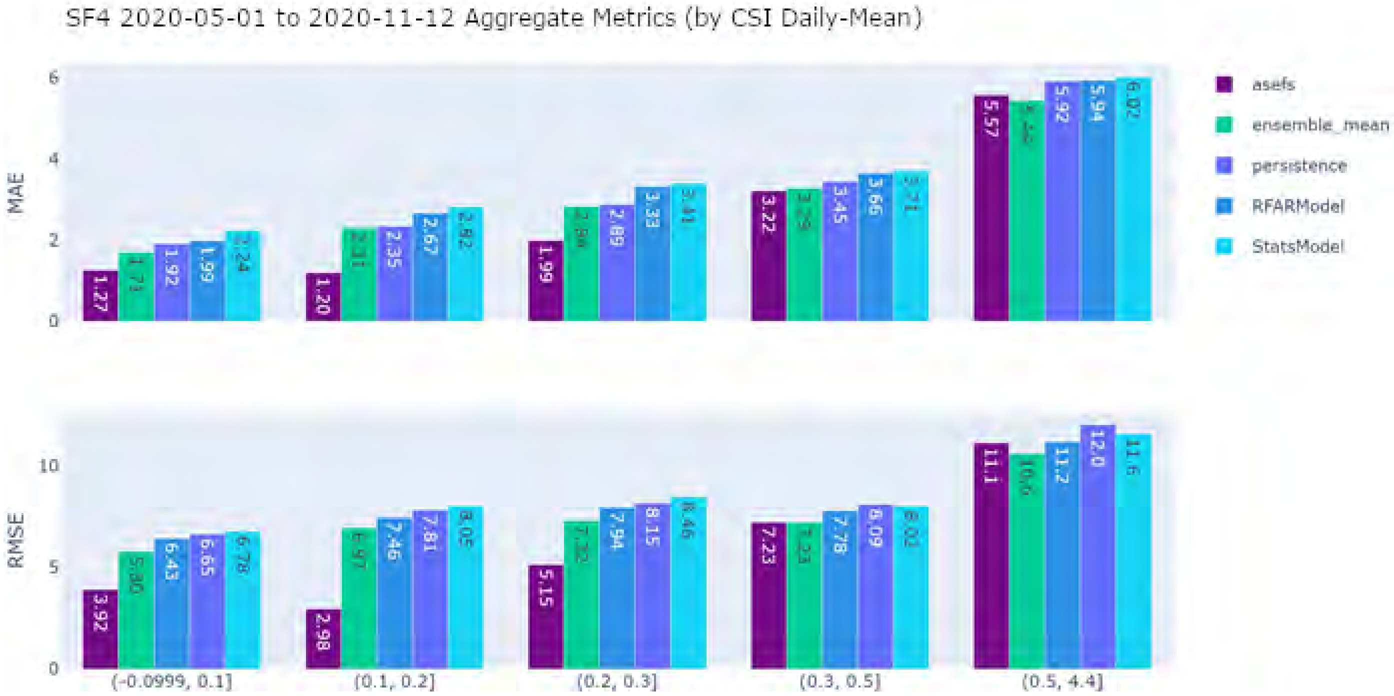
SF1 2020-05-01 to 2020-11-12 Aggregate Metrics (by CSI Daily-Mean)



SF2 2020-05-01 to 2020-11-12 Aggregate Metrics (by CSI Daily-Mean)







5.4. Forecast evolution

There weren't any large step-changes in forecast performance over the trial periods, but gradual improvements in model performance and stability as data quality issues were corrected, additional information about farm operation was discovered and operational issues were fixed. These incremental improvements included:

- zeroing forecasts when the sun zenith angle was below the horizon or negative power values were predicted
- setting a hard-maximum farm output level for the submitting model
- filtering out obvious periods of poor-quality power data from site SCADA systems
- automatic filtering of bad data in SCADA irradiance data
- splitting data into training/validation sets based on daily Clear Sky Index
- nightly model retraining

- regular hyperparameter tuning of machine learning models.

There was some variation in normalised forecast error (raw error divided by mean power generated) from the best model (in MAE) at each site over the last three months, as shown in Figure 4. Plotting these normalised errors against site latitude (Figure 5) results in a possible trend where more southerly latitude sites exhibited higher forecast error in both error metrics. The number of sites (n = 4) is too small to draw solid conclusions, but this may indicate that solar intermittency at more southerly locations is higher (at least over winter/spring), increasing the normalised forecast error at those sites. A larger dataset, ideally capturing all seasons and additional sites, would be needed to confirm this effect.

Figure 4 Normalised overall forecast error per site over the last three months (left) and six months (right) for the best model at each site

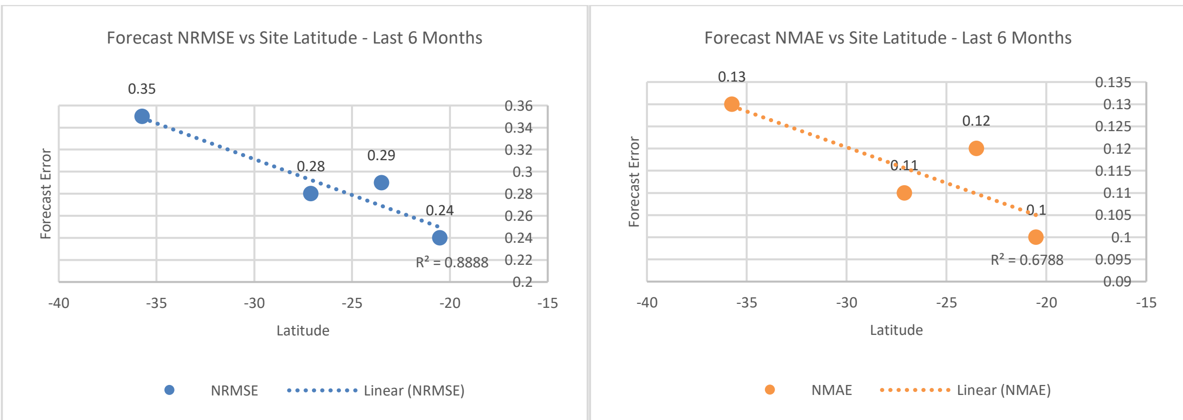


Figure 5 Six-month normalised forecast RMSE and MAE error vs site latitude, with linear fits, for best model at each site

To determine the relative contribution of each of the models and other features to the ensemble forecast, we extracted the relative feature importance values from a recently trained EnsembleML model for each site. Feature importance is based on the frequency with which each input (or *feature*) appears in the tree-based model and is an indicator of relative predictive skill of a feature in each model.

Figure 6 shows the relative importance of each model (prefixed: *forecast_*) and other inputs including various measures of irradiance (shown as GHI and DNI) and Clear Sky Index (prefixed: *CSI_*). Generally, the median and mean ensembles, Skycam, and Smart Persistence, Stats and RFAR models ranked highly, indicating that their contribution to the final forecast was more skilful than other models and inputs as far as the trained EnsembleML model could determine during training and validation.

The irradiance and CSI inputs were most highly ranked at SF5, but this is because the other models have not been running long enough to provide a robust training set for EnsembleML at this site.

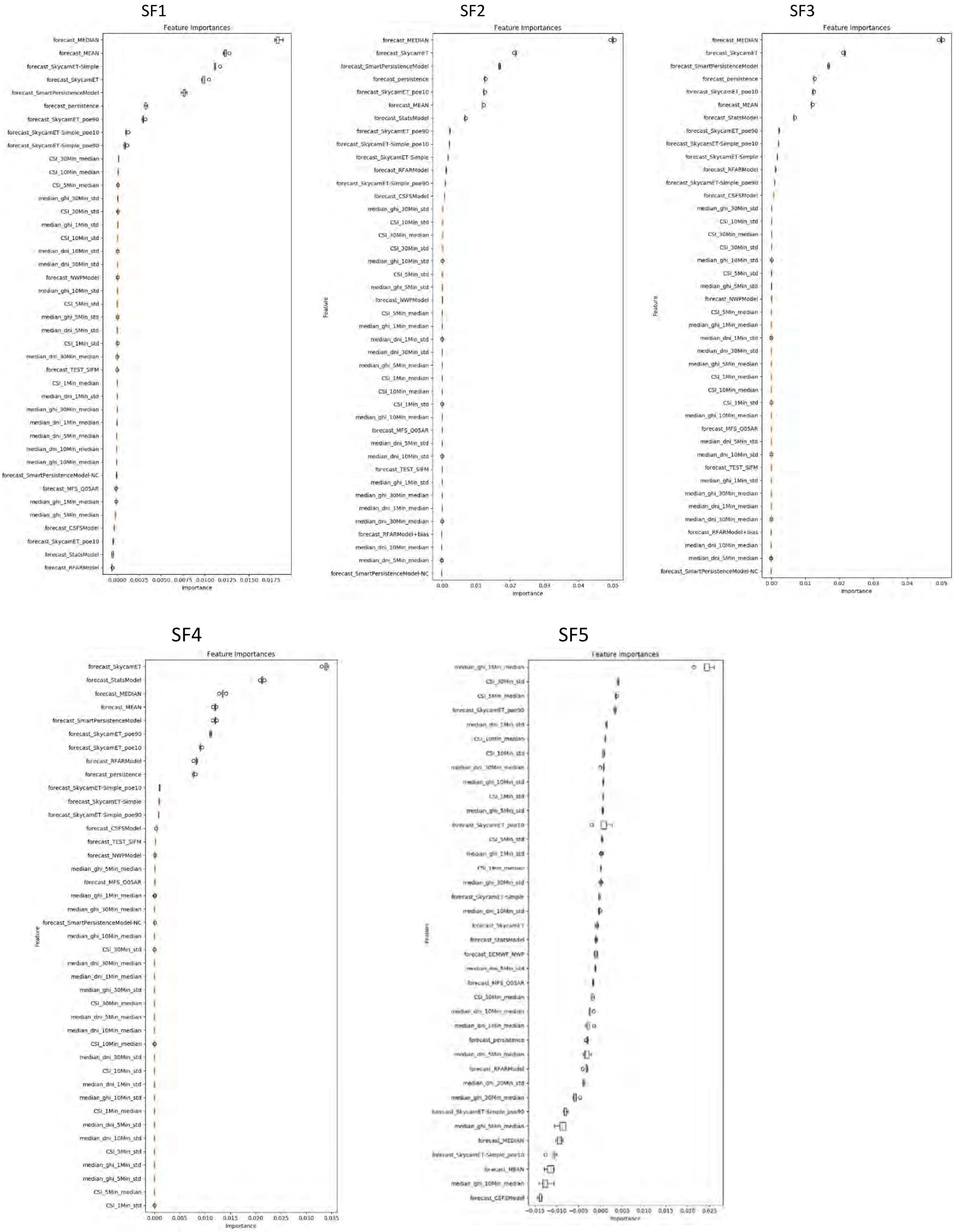


Figure 6 Example of the relative model & feature importance to the trained EnsembleML models at each site

5.5. Highlights and breakthroughs

This section summarises the notable model-specific highlights and breakthroughs that occurred during development over the course of this trial. Section 3.2 contains more a more general discussion of difficulties and methods of overcoming them.

5.5.1 Statistics Model

The key highlights and breakthroughs in developing Statistical Model include:

- This is a very simple model to implement and is versatile in that the procedure can be applied to any solar, or wind, farm and only uses the real-time output to retrain the model, on any required time scale. No extra equipment needs to be installed.
- It has allowed us to begin setting up a robust and versatile probabilistic forecasting tool, wherein error bounds on the forecast can be constructed in order to help minimise risk.

5.5.2 SIFM Model

The key highlights and breakthroughs in developing SIFM include:

- Capacity to nowcast GHI with conversion to Power for all sites in near real-time.
- Benchmark period shows GHI errors at sites comparable to other studies around the globe using a similar approach (Bright, 2019; Kamath & Srinivasan, 2020).
- Development of ensemble forecasts based on a diverse range of clear sky models.
- SIFM works well on clear sky days, including capturing the intermittent periods of power generation with modest accuracy.

6. Financial performance

The primary benefit of improved forecasts to the grid is to reduce the mismatch between forecast (or bid) energy and the dispatched (or actual) energy generated and delivered to the grid. This improves grid stability by allowing the market to more closely match supply to demand and keep grid frequency stable and avoid using more expensive methods to account for the difference.

6.1. FCAS Causer Pays

The direct financial benefit to renewable generators is in the reduction of so-called Causer Pays fees. Put simply, these fees are apportioned by AEMO based on the difference between a generator's bid and actual power, and the effect on grid frequency fluctuations, assessed at four second resolution.

This section contains results from reproducing AEMO's partially documented procedure for calculating and apportioning these Causer Pays fees, models the relationship between fees and model forecast accuracy, and applies the process to each of our models to estimate the fee savings compared to business as usual (i.e. using the ASEFS forecasts to determine the solar farm's market bids).

For this trial, we needed to choose a single model's output to submit to AEMO for use in dispatch and calculation of Causer Pays fees charged to the generators. Reimplementation and remodelling were necessary for two reasons:

1. We had many models running and aimed to determine real changes in these fees caused by each forecast individually, and
2. A causal relationship between a change in fee and a change in forecast error cannot be determined purely from a before-and-after study with no control, so this error-to-fee model was required.

We planned to use this financial model to optimise the forecast models to minimise FCAS fees rather than just minimising forecast error, as these aims are not identical. Due to long delays from AEMO in clarifying key aspects of the Causer Pays procedure though, insufficient time remained in the project to allow this. In theory, the models starting with the most accurate raw forecast (in terms of MAE and RMSE) should have been able to be optimised to provide the lower Causer Pays fees, by adding a step which selected the best confidence interval to minimise the forecast's effect on grid frequency. Because this was not completed in time, the estimated Causer Pays savings achieved by the forecast models are somewhat correlated to RMSE and MAE metrics but savings are lower than they could be if the models were optimised for FCAS directly, as planned.

6.2. FCAS modelling procedure

This section outlines the process used to reproduce and approximate AEMO's Causer Pays calculations.

If in a given four second interval a generator supplies more or less power than a straight line between its five-minute instantaneous power forecasts, *and* that difference makes a grid frequency deviation worse (higher or lower respectively), its relative contribution to the deviation will be recorded. The frequency correction costs for the market in a 28-day application period are allocated to each generator according to its relative contributions in a previous 28-day sampling period.

The lag between the sampling period and application period is presumably on the assumption that their actual power generation performance relative to their dispatch instructions (or forecasts in the case of renewable generators) doesn't change between these periods. This is probably a valid assumption for dispatched generation, however, the monthly results in Section 1 show that performance for the generators in this trial is not at all constant from month to month (mostly due to seasonal variation), invalidating this assumption – probably allocating fees amongst individual generators unfairly.

The full process to calculate these 28-day performance factors is shown as a flowchart in Figure 7 and is quite detailed. This process produces the Raise Not-Enabled Factors (RNEF) and Lower Not-Enabled Factors (LNEF), indicating the generator's impact on the grid frequency during frequency raise and lower events respectively. It should be noted that this calculation requires Frequency Indicator (FI) values not publicly published by AEMO and were instead supplied by one of the market participants in this trial.

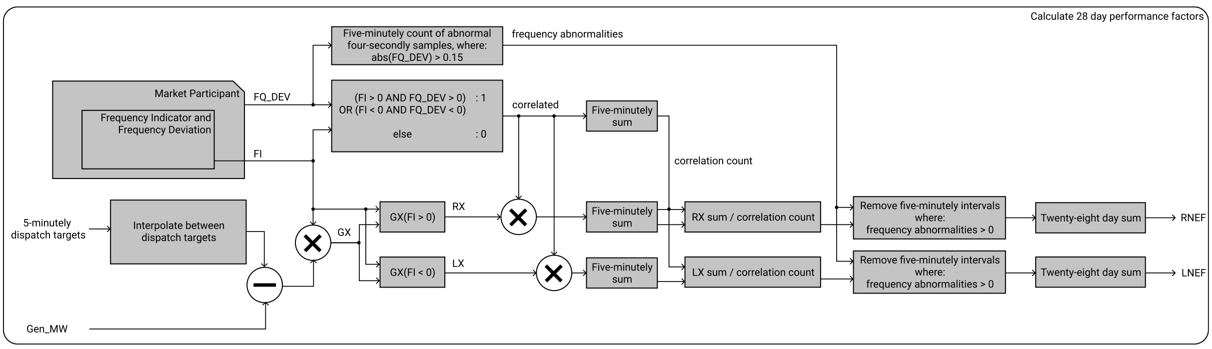


Figure 7 Overview of the process to calculate 28-day performance factors

To assess the financial impact of each forecast model, the respective Contribution Factors for each model per-site were estimated using RNEF and LNEF. Figure 8 details the modelling process used to fit Contribution Factor models and to calculate Regulation FCAS costs for each forecast model.

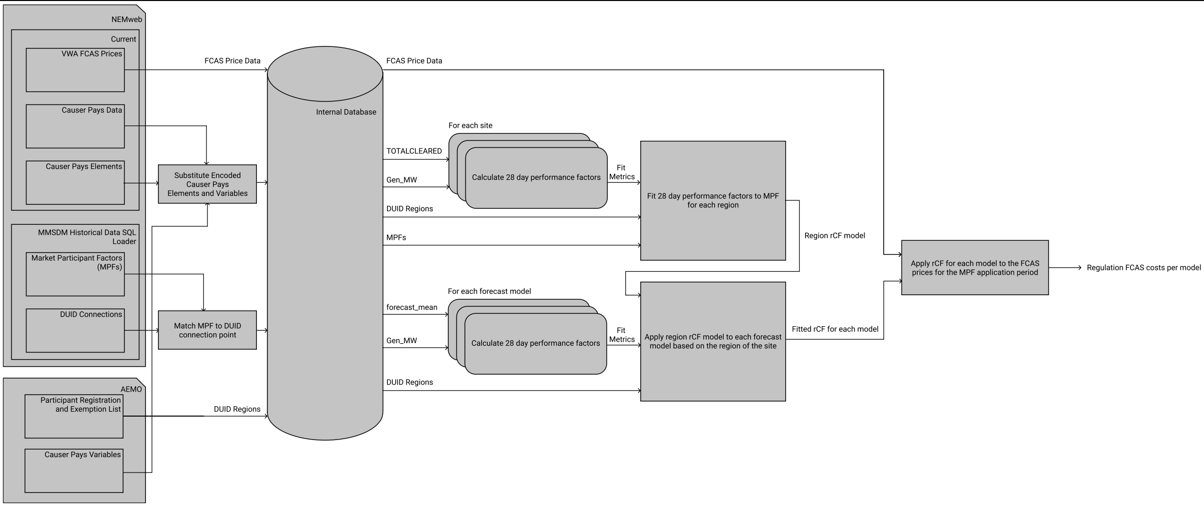


Figure 8 Visualisation of the process used to generate contribution factors and performance costings per forecast model, including sourcing the data from disparate sources, data pre-processing and storage as well as post-processing of the data to obtain forecast model costs

To allow Causer Pays fees estimation for each model's forecast errors individually, we then constructed a multi-input linear regression model for all solar farms in each state which predicted relative Contribution Factor (rCF) from calculated RNEF/LNEF and that farm's published dispatch forecast error metrics, aggregated over 28-day periods to match AEMO's procedure. These regression models can then be used to predict rCF for all other forecast models and ensembles as if their forecasts had been submitted to AEMO and used for dispatch; allowing a comparison of the Causer Pays fee that would have been charged, had each one *been* submitted. Unfortunately, because RNEF and LNEF require FI data to calculate, and AEMO do not publish FI values publicly or frequently, this process can only be used for after-the-fact fee comparison, and not for realtime model tuning to efficiently minimise fees. Such fee optimisation is still possible but would be much more effective were FI values published frequently.

After observing that the models using just MAE metrics did not produce good fits as expected, we explored additional error metrics. As shown in Figure 9, a combination of LNEF and RNEF was found to give the best linear fit to the published contribution factors[1] for each region. On average, models using RNEF and LNEF, were seen to outperform models based on MAE or RMSE with respect to R^2.

[1] AEMO Contribution Factors (https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/system-operations/ancillary-services/ancillary-services-causer-pays-contribution-factors)

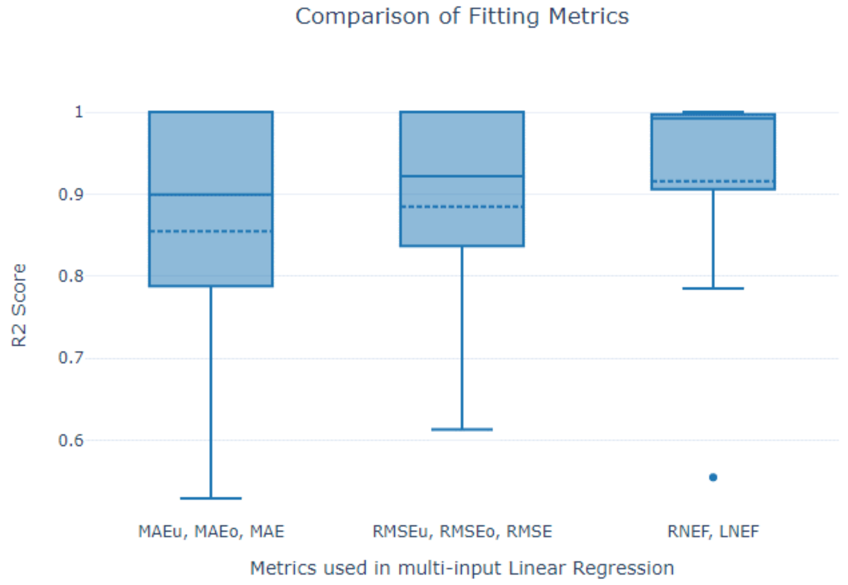


Figure 9 R^2 performance comparison between fitting using various combinations of the fitting metrics over four one-month assessment periods for NSW, QLD and VIC.

Due to this fitting performance, the best combination (the right-most combination in Figure 9) was used to construct the final rCF models. For example, rCF model fits for four assessment periods are shown in Figure 10. The obtained R^2 metrics for the remaining fits can be seen in Table 3.

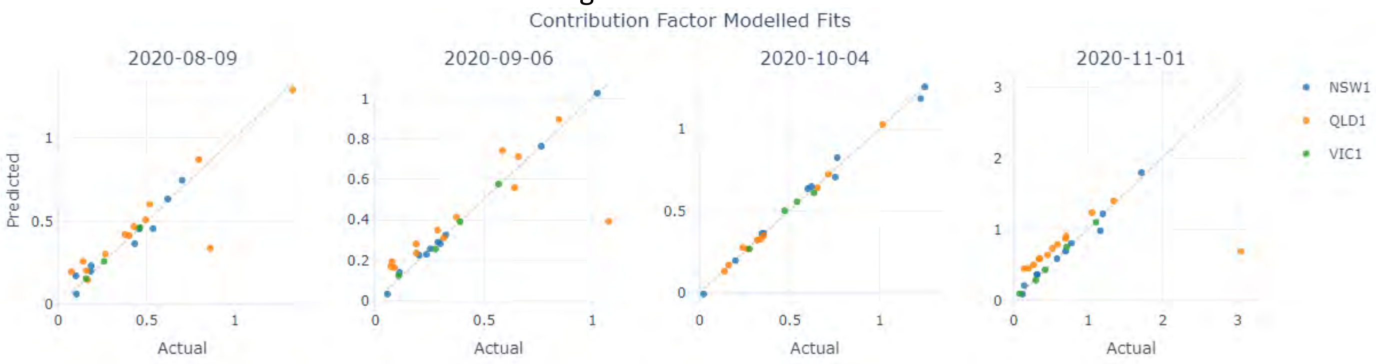


Figure 10 Predicted Contribution Factor (rCF) vs Published Contribution Factor (CF) for the NSW1, QLD1 and VIC1 regions over four Contribution Factor assessment periods

Table 3 R^2 results for the Contribution Factor model fits performed across four assessment periods

Region	Assessment Period Starting Date			
	2020-08-09	2020-09-06	2020-10-04	2020-11-01
NSW1	0.95	1.00	0.99	0.98
QLD1	0.78	0.55	1.00	0.14
VIC1	1.00	0.99	0.98	1.00

The fit metrics were then calculated for each forecast model, and the relevant rCF model for the region applied to the forecast's fit metrics, to produce rCFs for each forecast model. Using a following 28-day application period, the rCFs for each model were applied to the Regulation FCAS costs to produce per-model Causer Pays fees, giving estimates of the fees if these forecasts had been submitted to AEMO and used in dispatch.

The accuracy of cost estimates for all models (including ASEFS) and sites are limited by the accuracy of these linear fits, and so will not match exactly with the final charges to each site. Fortunately, the fits match the published data well across NSW and VIC, and the same fit applies to all forecast models at each site, so even if the absolute dollar estimates are not exactly correct this should still be an accurate method of estimating *relative* cost savings for each model at that site. The reduced fit R^2 scores in Queensland are likely a result of some DUIDs which are frequently capped (notably, the SF4 site that was part of this trial), and so produce reduced LNEF and RNEF metrics resulting in an underpredicted Contribution Factor. A comparison between published and modelled average contribution factors per-site can be seen in Table 4.

Table 4 Comparison between published and modelled contribution factors for the participant farms for the August-November 2020 period.

Site	Average CF	Average rCF	Error	Error %
SF2	1.139	1.153	0.036	3.243
SF3	0.231	0.226	0.016	8.846
SF5	0.265	0.288	0.024	9.251
SF1	0.620	0.724	0.113	18.106
SF4	1.669	0.473	1.195	67.311

While SF2, SF3 and SF5 all displayed an average error below 10%, SF1 shows a larger error, while the modelling error at SF4 highlights the distinct discrepancy between the fitted model for the Queensland region. Capped intervals are not included in RNEF/LNEF calculations, so frequently capped sites are likely to have lower RNEF/LNEF values, because there are fewer dispatch intervals in which Causer Pays fees apply to that farm, resulting in lowered modelled contribution factors for these sites. Additionally, suppressed forecasts are not included in the modelling process and so similarly would lead to lower modelled factors. As the dispatch forecast at SF4 has been AEMO-suppressed from the 8th of September onwards, this large modelling error is likely caused by this suppression.

6.3. Analysis of individual model contribution to causer-pays charges

Table 5 summarises the estimated FCAS savings for the dispatch forecasts at each site produced by an ensemble model. On average, the dispatch ensemble forecasts saved 13% in regulation FCAS costs when compared to the ASEFS forecast. Dispatch Forecasts refer to the final set of forecasts used by AEMO in dispatch and published publicly. These may differ slightly to the internal results due to various filtering procedures and server outages, but should be similar to *ensemble_median*, which was the forecast model used to generate the submitted forecasts to AEMO.

Table 5 Estimated FCAS saving for the August-November period for participant solar farms. Refer to Section A.4 for the full set of results.

Site	ASEFS Fee	Dispatch Model	Dispatch Ensemble Fee	FCAS Saving vs ASEFS	% FCAS Saving vs ASEFS
SF1	$109,720.54	SF_10	$106,642.18	$3,078.36	3%
SF2	$206,003.92	SF_12	$168,702.56	$37,300.44	18%
SF3	$32,963.14	SF_5	$27,130.32	$5,832.82	18%
			Average Saving	**$15,403.87**	**13%**

We could only perform this analysis over three months because of the lack of public availability of the necessary FCAS data (AEMO only publish data from the previous two months). Before this period, not all models were running and most were still being improved, so this recent period should give a better indication of ongoing performance than including additional older data. Additionally, SF4 and SF5 were omitted from financial assessment due to dispatch forecasts being suppressed either due to poor SCADA data quality or being within the self-forecasting initial assessment period, resulting in the dispatch forecast not producing forecasts for at least 80% of the expected number of intervals.

For each site, the top performing models which were applicable for dispatch and the respective regulation

FCAS fee savings are summarised in Table 6. On average, the top performing models saved 47% on regulation FCAS costs compared to the ASEFS forecast. The increase in individual model savings (+34%) compared to the dispatch ensemble model savings indicates the potential for additional savings by using forecast optimisation based on per-model FCAS costs, or for direct optimisation of the ensemble model for fee reduction and forecast accuracy.

Table 6 Summary of the top-performing models at each site, their Causer-Pays fee reduction and percentage savings versus modelled ASEFS. Only models which produced forecasts for at least 80% of the expected number of intervals and outperformed ASEFS in MAE and RMSE for the period were eligible.

Site	Top Performing	FCAS Saving vs ASEFS	% FCAS Saving vs ASEFS
SF1	SmartPersistenceModel	$43,819	40%
SF2	SmartPersistenceModel	$88,983	43%
SF3	SkycamET	$23,702	72%
	Average Saving	**$48,740**	**53%**

The average modelled contribution factors for each model are shown in Figure 11**Error! Reference source not found.** for each of the four sites modelled. Most forecast models displayed a lower modelled contribution factor than ASEFS, except for some irradiance-based forecast models (*TEST_SIFM*, *NWPModel*) and power models with consistent, significant under-prediction (*MFS_Q05AR*).

Models which outperformed ASEFS in terms of MAE and RMSE and which submitted at or above 80% of expected forecasts are distinguished from those models which did not. Missed model forecasts are not accounted for in the RNEF/LNEF or other fitting metrics. As a result, models with a larger number of missing forecasts will likely have a disproportionately low modelled contribution factor. For this reason, those models which do not provide unsuppressed forecasts for at least 80% of the expected number of intervals are deemed ineligible for dispatch. Additionally, those models which did not outperform ASEFS with respect to MAE and RMSE performance were also deemed ineligible.

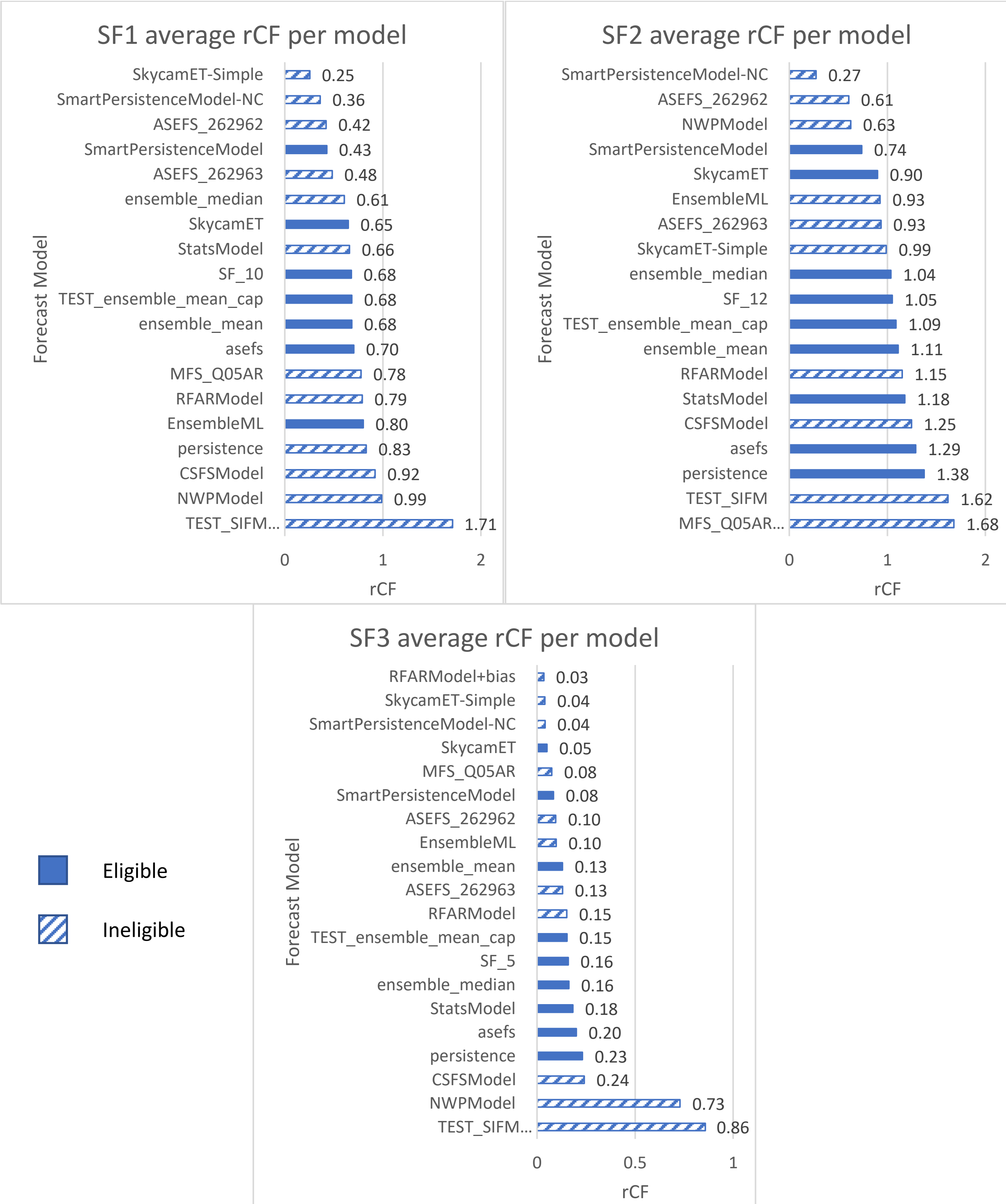


Figure 11 Average modelled contribution factors for the August-November period for each forecast model across sites (a) SF1 (b) SF2 (c) SF3. The *SF_* prefix indicates dispatch self-forecasts, while the ASEFS model is taken as the highest priority ASEFS forecast available per interval. Eligible models are defined as those meeting 80% or above of forecast intervals, as well as outperforming ASEFS in MAE and RMSE.

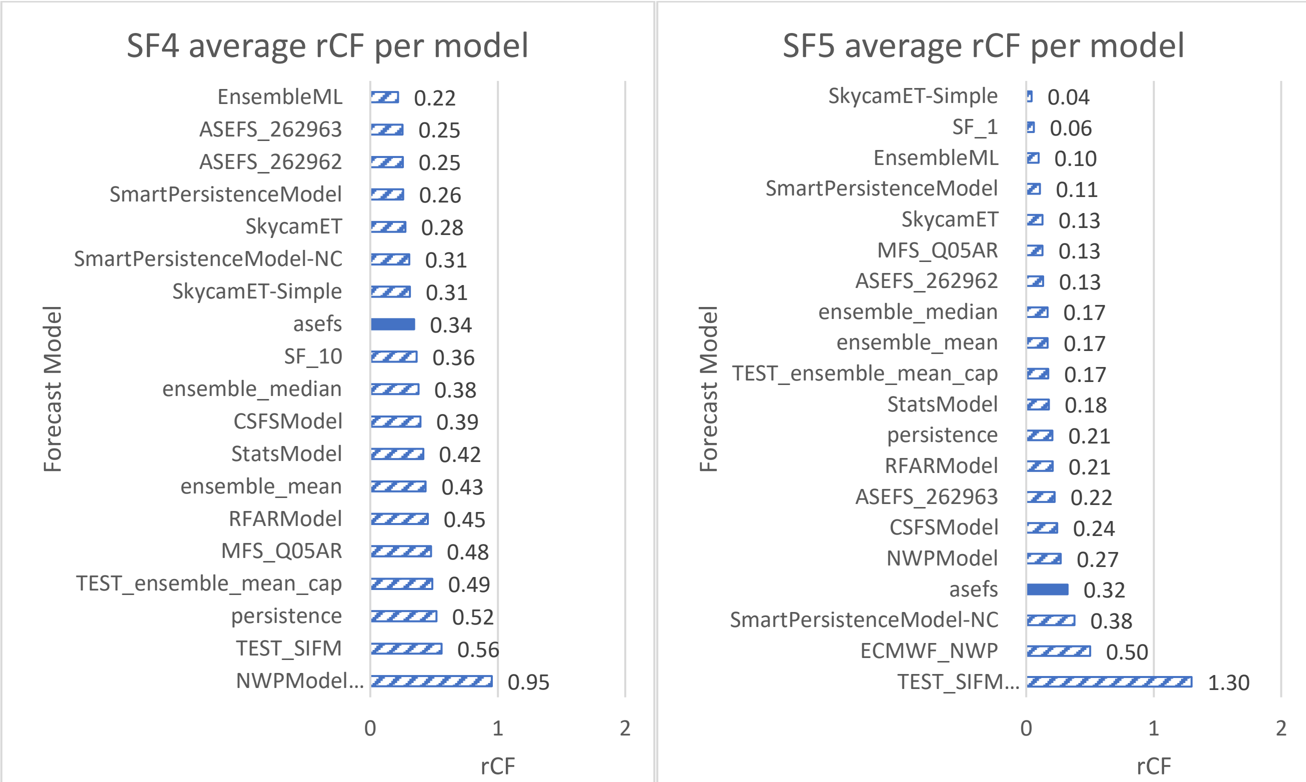
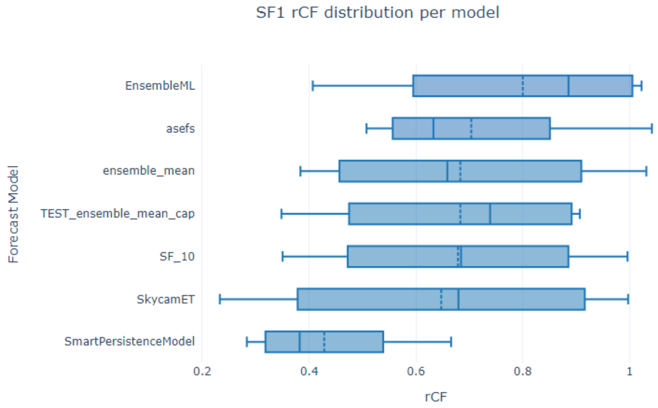
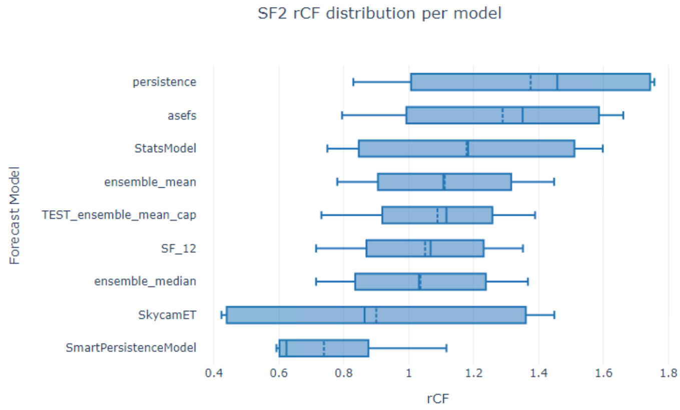


Figure 12 Average modelled contribution factors for the August-November period for each forecast model across sites (a) SF4 (b) SF5. For both sites, there was an insufficient number of model forecasts to analyse and/or not enough models outperforming ASEFS with respect to MAE and RMSE performance to select a best-performing model.

The distribution of fitted rCF values for each eligible model across sites SF1, SF2 and SF3 for the August-November period can be seen in Figure 13. Across all sites the SmartPersistenceModel consistently outperformed ASEFS with respect to the modelled rCF values.





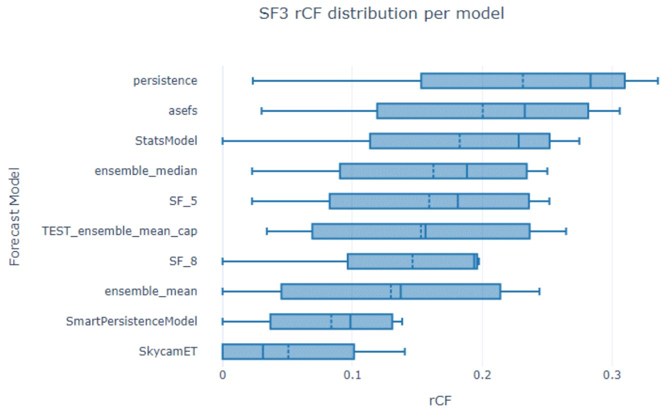


Figure 13 Distribution of eligible forecast models rCFs for the August-November period across sites (a) SF1 (b) SF2 and (c) SF3.

To assess the average financial performance of the various forecast models over a longer period, the average Regulation FCAS market cost was calculated for the July to December 2020 period. The distribution of Regulation Raise and Regulation Lower costs can be seen in Figure 14.

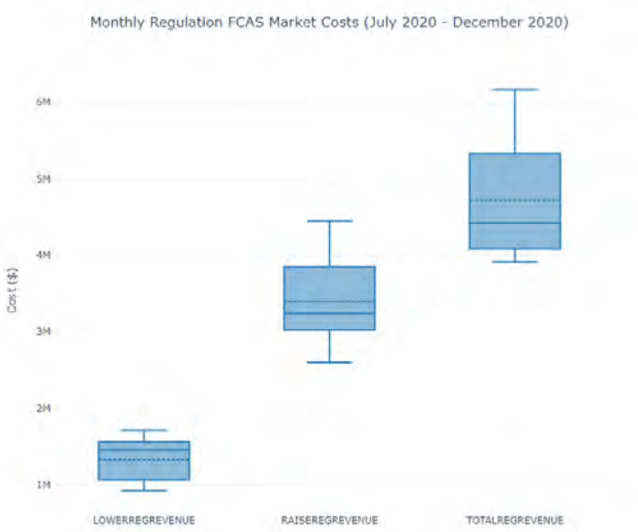


Figure 14 Distribution of market costs for Regulation FCAS Raise and Lower events for the July - December 2020 period. Average Regulation Raise events costs substantially outweighed Regulation Lower events costs at $3.24 million vs $1.32 million.

The calculated average monthly market cost for the July – December period was $4.72 million, 1.24 times greater than the August – November monthly Regulation FCAS total cost. To gain a greater understanding of expected model financial benefit based on average market costs, the July – December monthly market cost was applied to the fitted rCFs for each eligible model. This was intended to separate the fairly consistent rCF model performance from the more volatile FCAS market pricing, by averaging the pricing over a longer period.

Table 7 Summary of the top-performing eligible models per site, as well as the average Regulation FCAS savings for four months compared to ASEFS, based on the average monthly total Regulation FCAS cost over the July - December 2020 period. Table 7 shows the Regulation FCAS savings for top-performing eligible models at SF1, SF2 and SF3, calculated based on the longer-term July - December averaged monthly market cost. To be directly comparable to Table 6, the fitted rCFs over four application periods along with the average market cost was used to produce Table 7. When examining the Regulation FCAS cost over a longer period, the average estimated Regulation FCAS saving increased from $48,740 to $61,419 (+$12,679 or +26%). These levels of saving are likely more indicative of long-run savings from these models at each site, as they are less coupled with market price volatility (being calculated from prices averaged over a somewhat longer period) and more reflective of relative model Contribution Factor performance than the values in Table 6.

Table 7 Summary of the top-performing eligible models per site, as well as the average Regulation FCAS savings for four months compared to ASEFS, based on the average monthly total Regulation FCAS cost over the July - December 2020 period.

Site	Top Performing	Avg. FCAS Saving vs ASEFS	Avg. % FCAS Saving vs ASEFS
SF1	SmartPersistenceModel	$51,991	40%
SF2	SmartPersistenceModel	$103,974	41%
SF3	SkycamET	$28,292	77%
	Average Saving	**$61,419**	**53%**

6.4. Other financial impacts (Non-FCAS Causer Pays)

No other financial impacts were identified that may be attributable to increased accuracy of self-forecasts or participation in MP5F trial.

6.5. Financial benefits: incremental dispatch revenue

All farms in this trial export their maximum power output to the grid when not capped, rather than attempting to follow their linear ramp between five-minute dispatch targets, because they earn more income from power purchases than they are charged in Causer Pays fees caused by deviations from this ramp.

During capped periods, the dispatched power is either:

- the semi-dispatch cap level, if the available power is greater than the cap level, or
- the available power if it is lower than the cap level.

As such, we understand that any change in the dispatch power would only be determined by a change in the cap level. Determining a change in the cap level for a forecast would require reimplementing AEMO's procedure for setting a cap level. Given the difficulties in obtaining information for re-implementing the Causer Pays procedure, the cap's apparent dependency on intra-regional constraints and the lack of detailed published information about how this cap level is set, reimplementation was deemed to be beyond the scope of this trial and not considered in assessing financial benefits. However, most sites in this trial experienced relatively few cap-constrained periods, so we have assumed that the change in dispatch revenue due to changes in dispatch power for each model is negligible.

6.5.1 Incremental FCAS revenue

None of the generators included in this trial participated in the FCAS regulation market (beyond their compulsory exposure to Causer Pays costs), so there was no FCAS revenue received by the generators before or after self-forecasts were implemented.

6.5.2 Other revenue benefits

We have no information about any changes in Power Purchase Agreements as a result of more reliable forecasting, so have assumed this to be zero.

6.6. Non-financial benefits

No non-financial benefits were identified during this project.

6.7. Financial benefits to market/consumers

End-use energy consumers benefit from improved renewable energy generation forecasts because increased forecast accuracy will increase grid stability by reducing the number and magnitude of frequency deviations caused by supply-demand imbalances. This, in turn, should lower the grid's reliance on the FCAS regulation market to compensate for these frequency excursions, resulting in lower overall costs and a more efficient energy market. Increased stability should also allow a higher penetration of renewable generation, which will lower average energy wholesale costs and decrease carbon emissions by displacing more expensive fossil-fuel generation.

These benefits have not been quantified as they are well out of scope of this project but warrant further investigation in wider ranging research.

7. Web dashboard

A publicly accessible web dashboard has been developed to help demonstrate the technology and its performance on all trial sites. This dashboard includes:

- 30-min delayed power, forecast & weather station data
- a map with site locations
- satellite imagery
- skycam images
- skycam stereo ray-traced images.

The dashboard can be accessed here http://dash.wattcloud.ai/.

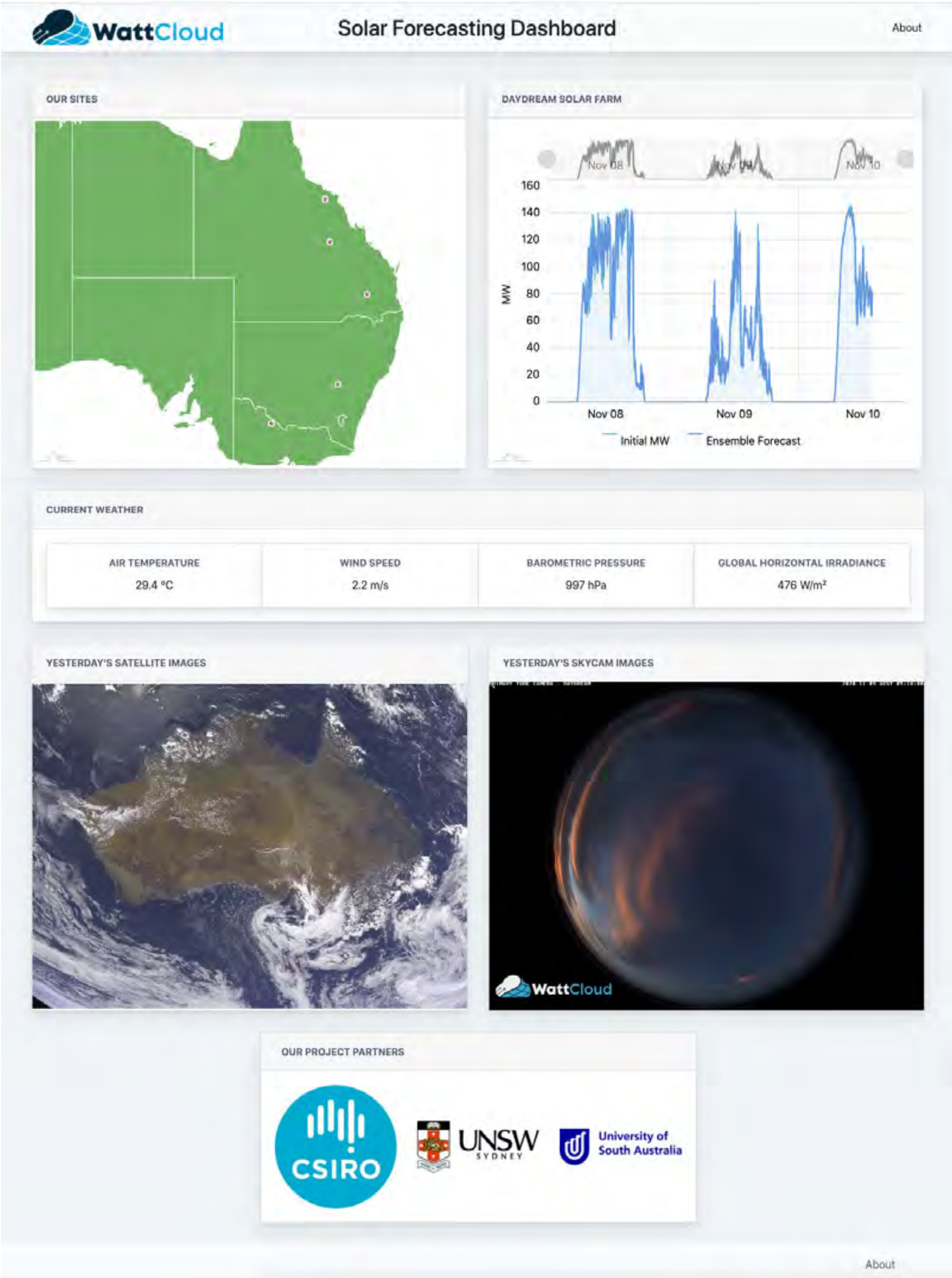


Figure 15 Sample view of WattCloud web dashboard

8. Dataset

A public dataset was released alongside this report to allow students, researchers and industry to use and analyse the high-quality data collected in this project from the five solar farms in this trial. This data comprises a six-month dataset sampled at one-minute frequency, collected from a supplied weather station and power, weather and irradiance data supplied by the solar farm's SCADA systems. These tables form the core of the datasets used to train the models presented in this report.

This dataset can be accessed by searching for '*Solar Power Ensemble Forecaster*' at https://data.csiro.au/collections/ or using this persistent DOI link: http://hdl.handle.net/102.100.100/384938

If this data is used in any publications, please reference this report and the dataset together.

The data can be used with attribution for non-commercial purposes. Full license details are available via the links above.

Data is raw and unprocessed, except for resampling (with no averaging) to one-minute intervals. Not all fields contain data and missing data is generally denoted by NA values. The dataset is released as simple gzip-compressed comma separated value (CSV) files and as more modern parquet files, an efficient typed binary format readable in most languages. One file per table per site is provided. Timestamps are in NEM-standard time, which is equivalent to Australia/Brisbane (i.e. Australian East Standard Time with no daylight savings adjustments). Tables can generally be joined using their timestamps, rounded to a common interval, as keys.

Column names are prefixed with their table names for clarity. Table name notation is:

- pvexport_data: power values measured at the solar farm's high voltage grid connection transformer
- peripheral_data: weather and irradiance values measured by the supplied weather station hardware and pyranometer, collocated with the sky camera
- site_peripherals: data provided by the solar farm via their SCADA systems. Most farms had multiple stations spread around the site. Status columns are somewhat site-dependant but are generally a binary mask flagging which values in this table are valid for each sample. Status quality varies across the sites.

Table 8 Column names and descriptions of the public dataset fields

Column Name	Description	Unit
pvexport_data_frequency	Grid frequency	Hz
pvexport_data_power_real	Real power exported to grid	kW
pvexport_data_power_reactive	Reactive power exported to grid	kVa
pvexport_data_voltage	Export voltage	V
pvexport_data_current	Export current	A
peripheral_data_baropress	Barometric Pressure	kPa
peripheral_data_airtemp	Ambient air temperature	°C
peripheral_data_relhumid	Relative Humidity	%
peripheral_data_dewpoint	Dewpoint Temperature	°C
peripheral_data_windangle	Wind angle relative to North	°
peripheral_data_windspeed	Wind speed	m/s
peripheral_data_pyroup	Global Horizontal Irradiance (GHI)	W/m^2
peripheral_data_pyroangle	Plane of Array Irradiance	W/m^2
site_peripheral_airtemps_status	Binary status mask	n/a
site_peripheral_airtemps_station_1 ... 6	Ambient air temperatures	°C
site_peripheral_dcpowers_status	Binary status mask	n/a
site_peripheral_dcpowers_dcpu_1 … 64	Inverter DC power	kW
site_peripheral_humidities_status	Binary status mask	n/a

Column Name	Description	Unit
site_peripheral_humidities_station_1 … 6	Station relative humidity	%
site_peripheral_irradiances_station_1 … 6_normal	Station plane of array irradiance	W/m^2
site_peripheral_irradiances_station_1 … 6_hrzntl	Station Global Horizontal Irradiance	W/m^2
site_peripheral_paneltemps_status	Binary status mask	n/a
site_peripheral_paneltemps_station_1 … 6	Station solar panel temperature	°C
site_peripheral_rainfalls_status	Binary status mask	n/a
site_peripheral_rainfalls_station_1 … 6	Station rainfall	Mm
site_peripheral_winds_status	Binary status mask	n/a
site_peripheral_winds_station_1 … 6_speed	Station wind speed	m/s
site_peripheral_winds_station_1 … 6_angle	Station wind angle	°

9. Technology development

This project has resulted in the launch of WattCloud, which aims to commercialise the solar forecasting technologies that have been developed.

9.1. IP and collaboration

All consortium members intend to continue development of the existing models and additional models for a minimum of six months following the completion of this trial. Each member has agreed to negotiate the licencing of relevant IP in good faith to allow the effective commercialisation of the technology.

10. Lessons learnt

- **Forecasts should perform well in *both* RMSE and MAE metrics to minimise Causer Pays costs**. A range of error metrics were tested when modelling causer-pays Contribution Factors (CF) to match AEMO's procedure, and the best models to estimate these fees from forecast errors generally relied equally on MAE and RMSE metrics, mostly with a slightly heavier weighting toward RMSE. This is an important finding that is not obvious from the Causer Pays procedure documentation, from which we had previously assumed that MAE would be more important. Most forecast models can be optimised to minimise one (not both) of these competing error metrics, but the ensemble models generally did well in both, suggesting that ensembles are a good approach to this problem, though more work remains to optimise for fee minimisation directly.

- **Power autoregression gives better performance, but semi-dispatch caps mean irradiance-only models are also critical**. Most models gave the best predictions when some indication of the actual current and past farm power output was used as model input. This is due to errors and uncertainties in the irradiance measurements and the complexity of modelling the irradiance-to-power conversion for the entire farm in a way that captures all the possible variables, which is very complex. Modelling power directly often bypasses these issues but requires additional logic to produce an unconstrained power forecast. While forecasts *during* a semi-dispatch cap being imposed are not assessed, these autoregressive power models perform poorly for a small number of intervals immediately after the cap is released, while power levels return to unconstrained levels and their historical power window still contains the constrained power values. Our Ensemble models deal with this by switching to using irradiance-only models during and immediately after constrained periods, but generally provide less accurate predictions than the power models. On farms which are frequently constrained (in particular, SF4), this causes overall forecast performance to suffer, making improvements in irradiance-only models important.

- **Poor SCADA data quality benefits the incumbent forecast**. When the SCADA power data sent from the farm to AEMO is flagged as poor quality, the INITIALMW value is sometimes replaced with the UIGF. This quality data is difficult to obtain, and if not considered when assessing forecasts makes the incumbent forecast (e.g. ASEFS at SF4) appear to be much higher performing than it may be. This is likely a major factor explaining why ASEFS performs so well at SF4, but needs further investigation to confirm.

- **Balancing training data by clear sky index is important**. For a given dataset, how it is split into training and validation sets is very important, both for getting an accurate estimate of live forecast performance of a model and for any kind of model parameter tuning which relies on validation performance to select optimal parameters or feature subsets efficiently. Of the various methods tested, ranking each day by clear-sky index (CSI) and putting an equal share in both sets gave the best results.

- **Export power ramps occurring before and after the semi-dispatch cap flag is set need to be removed from training data.** For models training on data sampled more frequently than five minutes, data needs to be removed from the training set for the entire interval before and after a cap causes the farm's output to be curtailed, so the pre- and post-cap ramps aren't used for modelling. Significant forecast improvements in validation (4-5% improvements in MAE skill vs ASEFS) were observed after implementing this filtering in several models.

- **Data identifying curtailment from network constraints should be considered**. There were several occasions at most farms where the farm output appeared to be constrained outside a semi-dispatch cap period. This resulted from a network constraint (e.g. SF3) and some cases of unscheduled maintenance. These periods need to be flagged and removed from model training data sets to ensure an accurate unconstrained forecast model.

- **Other types of curtailment need to be accounted for**. Data from other types of curtailment, such as maintenance and unplanned outages, should be removed from data when training models to avoid unintended bias they may create in the models that rely on historical power to train against. Data flagging these periods was often difficult to obtain from the farms, and probably explains some difference in forecast performance between farms, but this could not be quantified.

- **Bias correction improves performance**. We observed (in late September) that most models with a persistence component exhibited statistically significant biases to underpredict in the morning and overpredict in the afternoon, Figure 16 illustrates this effect visually. This was because 'dumb' persistence doesn't explicitly deal with the clear sky slope and tends to persist the current power level horizontally, despite usually having implicit information about the slope. It was surprising that even most advanced machine learning models still didn't capture this trend. The exception was Smart Persistence (which had a slight bias in the reverse direction) because it explicitly models the power slope; this may explain why Smart Persistence generally performs well. Automatic bias correction has now been added to most models to address this issue but was not put into production early enough to have a significant improvement in the results in this report.

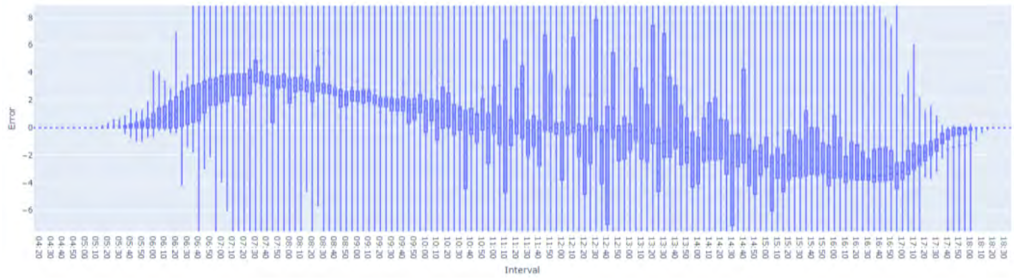


Figure 16 Boxplot showing the distribution of forecast errors from the persistence model per five-minute interval, aggregated over a three-month period. The trend of morning underprediction (positive errors) and afternoon overprediction (negative errors) can be observed.

- There were several lessons specific to satellite forecasting models:
  - Errors in short-term irradiance forecasts using the advection of derived Global Horizontal Irradiance (GHI) with Cloud Motion Vectors (CMVs) depend more on the quality of derived GHI than the CMVs.
  - The CMV accuracy is dependent on image inhomogeneity and contrast. Images with little contrast (e.g. very overcast conditions) produce larger errors in CMV estimation.
  - HELIOSAT technique for deriving GHI is fast but lacks precision because of several assumptions relating to clouds and aerosols in the atmosphere. On the other hand, physical models using rapid radiative transfer may be more accurate, but these lack speed, making usage in real-time applications with minimum computational resource more difficult.

10.1. Implications for future projects

General Recommendations:

- AEMO should publish Frequency Indicator data and a short, worked example of a Causer Pays calculation procedure publicly.
- Direct optimisation of Causer Pays fees should be possible, but further work is required to develop this.
- Some ongoing maintenance of production models will probably be necessary, particularly when commissioning new sites with new data cleaning requirements.
- Improvements in irradiance-only forecast models will likely increase forecast skill for frequently capped generators.
- Extra data flagging other types of generation constraints (network, maintenance, outages) would likely improve forecast skill. These periods should also be classed as non-assessable when evaluating forecast performance, similarly to how cap-constrained periods are dealt within the current AEMO guidelines.

Satellite models:

- Improvements in the accuracy of GHI derived using multiple satellite channels may require fast radiative transfer models, but this may increase computational costs and processing times for forecasts at short time scales.
- Errors resulting from power conversion with GHI and DNI inputs would require proper evaluation with observations, especially DNI observations with longer records.
- Use of machine learning techniques in capturing the relationships between clear sky index and the cloud index by taking multiple images captured at different bands (wavelengths) may also minimise errors in prediction of solar power from GHI estimates.

11. Conclusions

Forecast accuracy and Causer Pays reduction results are discussed in Sections 5 and 6. The main findings included:

- One of the three ensemble models (Mean, Median or ML) was ranked in the top two models overall on every site for both MAE and RMSE.
- Ensemble Median and Ensemble Mean both outperformed ASEFS on all sites (except SF4) over the last two months in both RMSE and MAE. Ensemble Mean also outperformed ASEFS over the last six months (it was the only ensemble running for the entire period).
- Two models, Smart Persistence and Skycam, both individually outperformed ASEFS at all sites except SF4 in both RMSE and MAE over the two months.
- Overall, the best performing models and skill (percentage improvement) vs ASEFS at each site over the last two months were:

Site	Best Model - RMSE	% RMSE Skill vs ASEFS	Best Model - MAE	% MAE Skill vs ASEFS
SF1	EnsembleML	9.19	Ensemble Median	13.07
SF2	Ensemble Mean	16.23	Ensemble Median	18.19
SF3	Skycam	19.31	Smart Persistence	16.81
SF4	Ensemble Mean	2.84	ASEFS	-
SF5	Skycam ET	21.22	Skycam ET	16.56

- This is a relatively small sample size (approximately three months) and these results may change over a longer period.
- Raw forecast performance in terms of RMSE and MAE (or other error metrics) was found not to be a good indicator of modelled or actual Causer Pays fees charged to the generators in this trial. Model prediction quality can be effectively judged by these metrics, but fee reduction cannot and requires additional modelling. We believe that this is due to several factors:
    - Causer Pays fees in contribution factor periods are based on a generator's Contribution Factor from a previous 28-day period.
    - Only forecast errors which occur concurrently with grid frequency fluctuations under a threshold value contribute to the fees, and these are not uniformly distributed or easily predictable.
    - Fees also depend on the behaviour of other farms in the region during frequency fluctuations, so are not directly tied to just the incumbent farms' forecast.
- A trend was observed where more southerly sites exhibited higher normalised forecast error over the last three and six months, possibly indicating that solar intermittency is higher there over winter/spring. The sample size was small, however, so more data would need to be collected to confirm this effect.
- Because of the delays discussed below, additional Causer Pays modelling to estimate fees from raw error metrics was completed late in the project, so models were not able to be optimised for fee reduction, and deeper financial savings were almost certainly possible.
- Estimated Causer Pays fee reduction over the three-month August to November 2020 period from the best financially performing model at each site were found to be:

Site	Top Performing	FCAS Saving vs ASEFS	% FCAS Saving vs ASEFS
SF1	SmartPersistenceModel	$43,819	40%
SF2	SmartPersistenceModel	$88,983	43%
SF3	SkycamET	$23,702	72%
	Average Saving	**$48,740**	**53%**

11.1. Recommendations

- Further work is needed to fully utilise the re-implementation of AEMO's Causer Pays procedure for optimising forecasts, to realise the full financial benefit of accurate models for minimising Causer Pays fees for generators. Most modelling approaches trialled allow customised optimisation metrics and can produce probabilistic outputs. It is recommended that these facilities are used to modify models to minimise Causer Pays fees, while maintaining a reasonable MAE/RMSE forecast error. Realtime (or close to) access to Frequency Indicator (FI) data feeds will be required to implement this process. We recommend this data is made publicly available.
- The published AEMO Causer Pays procedure is opaque, incomplete and extremely difficult to reimplement and validate using public information and data. AEMO staff even stated that some final steps have never been successfully reimplemented to their knowledge. To increase transparency and assist in grid stability, we recommend a full worked example spreadsheet or full operational code implementation be made available so that market participants and forecast providers can optimise forecasts, improving grid operation and minimising exposure to Causer Pays fees.
- Large variation in forecast accuracy was observed in all models from month to month, particularly around the change of season, where more intermittent conditions became more prevalent, increasing forecast error. The Causer Pays procedure seems to assume that forecast error remains relatively constant, as it applies contribution factors from the *previous* month to FCAS regulation costs for the current month to calculate fees for generators. This system allocates fees to solar (and presumably other weather-dependant renewable) generators which aren't reflective of their actual forecast performance. We recommend the Causer Pays procedure be reviewed considering these findings.
- Given the significant demonstrated improvement over the incumbent ASEFS forecasts at the 5-minute ahead horizon, and the flow-on benefits to market participants, grid stability and energy consumers, we recommend that similar funding be considered for the trial and deployment of forecasting systems for longer time frames.
- Consideration should also be given to the utility of including probabilistic forecast outputs, similar to the Probability of Exceedance (PoE) values that ASEFS and AEMO's Load Forecasting system already output. This would give a better indication of any disagreements within individual models or between multiple ensembled models and provide additional information about the probabilistic distribution of possible forecast values, allowing better dispatch decisions to be made.

12. Supporting information

Full details of per-model error metrics and FCAS Causer Pays results can be found in the supporting material. This material includes:

- *All Sites Aggregate Metrics 2 Months (Overall) (Table).xlsx* — the full set of error metrics discussed in Section 5.3.1
- *All Sites Aggregate Metrics 6 months (Overall) (Table).xlsx* — the full set of error metrics discussed in Section 5.3.2
- *FCAS_Summary.xlsx* — The full set of fee estimates and intermediate calculations discussed in Section 6.1.

12.1. Knowledge sharing

A high-level project article was posted to the CSIRO ECOS portal here https://ecos.csiro.au/nothing-but-blue-skies-how-solar-forecasting-works/. Marketing and promotional materials are being uploaded to the http://wattcloud.solar/ website. Several papers were published as a result of the trial, these are detailed in the sections below under each forecast model and respective organisation.

12.1.1 SIFM

- Initial results from SIFM were presented at a national conference in Perth, Australia with reference:
  Prasad A; Kay M, 2020, 'Short-term satellite irradiance forecast model', presented at 27th AMOS Annual Meeting and International Conference on Indian Ocean Meteorology and Oceanography (AMOS 2020), Fremantle, Western Australia, 10 February 2020 - 14 February 2020.
- Final results from SIFM will be submitted for a publication in the Journal of Remote Sensing (ISSN 2072-4292).
- Future Conferences
  AMOS Annual Conference, 8-12 (to be held online) February 2021

12.1.2 Statistics

12.1.2.1 *Publications*

John Boland, Characterising Seasonality of Solar Radiation and Solar Farm Output (2020), Energies, 13, 471; doi:10.3390/en13020471

12.1.2.2 *Presentations*

John Boland, Sleiman Farah and Lei Bai, Short term forecasting of solar farm output, ISES Solar World Congress, Santiago, Chile, Nov 2019 (virtual). John Boland, Sleiman Farah and Lei Bai, Short term forecasting of solar farm output, MODSIM 2019, Canberra, Dec 2019.

John Boland and Sleiman Farah, Forecasting solar radiation and farm power using combined seasonality and autoregressive model, 2020 Asia-Pacific Solar Research Conference
30th November to 2nd December, 2020.

12.1.2.3 *Projected Publications*

Sleiman Farah and John Boland, Forecasting photovoltaic solar farm electrical output, International Journal of Forecasting or Solar Energy (probably the latter).

Sleiman Farah and John Boland, Probabilistic forecasting of solar farm output, International Journal of Forecasting or Journal of Renewable and Sustainable Energy (probably the latter).

12.1.2.4 *Future Conferences*

AMOS 2021, SCIENCE FOR IMPACT, 8-12 February 2021