Internship report

Reuniwatt SAS, 14 rue de la Guadeloupe, 97490 Sainte-Clotilde
Tel : 06.92.64.43.13 / Fax : 02.62.92.10.20 / courriel : info@reuniwatt.com
Société par Actions Simplifiée au capital de 191.700€ RCS Saint-Denis 518 919 345

Contents

Abstract

For day-ahead forecasts, the combination of Numerical Weather Prediction (NWP) models and post-processing algorithms is the most effective method.

However, it is hard to extract from all the literature on the subject the best algorithm to use because of the lack of consistency in the different approaches.

During my Internship, my mission was to investigate the best algorithms according to the literature so as to improve the day-ahead irradiance forecasts.

My final results demonstrated improved metrics in comparison to the current algorithm used by Reuniwatt.

1 Introduction

1.1 Background and motivation of the intership project

This report is the result of my 6-month internship in Reuniwatt, a leader in cloud observation and forecasting. My internship extended from March 1st to August 31st, taking place during the second semester of my academic gap.

The main subject of the intership was the post-processing of the day ahead NWP irradiance forecasts. Despite their proven utility for day-ahead irradiance forecasting, NWP models predictions can still be improved thanks to post-processing.

As I will show in 2.3, many models have been investigated in the literature, and it is thus important to draw a clean benchmark of all the available state-of-the-art models.

Statistical models will be investigated, whose applications do not restrict to day-ahead irradiance forecasts. Indeed any timestamped weather-related variable could benefit of the post-processing I am going to discuss in this paper.

1.2 Objectives of the internship

Hereafter the main objectives of the internship:

- Benchmark promising models on the post-processing of a single NWP model.
- Sensitivity study of the models.
- Comparison of the results with the current model used by Reuniwatt for day-ahead forecasting (LT CONT).

2 Methodology

2.1 Data source

Verbois et al. demonstrated that using a large set of predictors can significantly improve the performances of post-processing models, while Suksamosorn et al. selected WRF forecasts of irradiance, temperature, relative humidity and the solar zenith angleas relevant inputs of the models.

Our initial data source for the forecasts was GFS, and we opted for the following set of predictors (1), both simple and easily available for any location.

The forecasted data is for each day the one relative to the origin 00:00 UTC of the day before. The irradiance explored during my internship is the global horizontal irradiance (GHI), which is the total solar radiation incident on a horizontal surface.

ghi_{GFS}	T^{2m}_{GFS}	θ	ϕ	ghi_{cs}
Irradiance forecasted	Temperature forecasted 2 meters above the ground	Zenith angle	Azimuth angle	Clear-sky irradiance

Table 1: Set of predictors.

Verbois et al. advises researchers to analyze their models performances over several years but I was at this point limited by the Reuniwatt API, thus I initially opted for learning during 2020 and testing during 2021.

The four initial study sites are the following (1):

Figure 1: Four initial study sites

2.2 Metrics

Even if papers like Mayer and Yang state that the correlation coefficient is the recommended metrics to use when no clear directive is given, the metrics that I am going to investigate are the one already prefered by Reuniwatt, the mean absolute error (MAE) and the root mean square error (RMSE).

I also wanted to investigate the MBE optimization, but the results were not convincing and MBE is more seen in our study as a metrics to be verified after post-processing. We indeed aim at the lowest absolute MBE.

- The mean absolute error

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|I_{forecast,i} - I_{measure,i}|$$

- The mean bias error

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(I_{forecast,i} - I_{measure,i})$$

- The root mean square error

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_{forecast,i} - I_{measure,i})^2}$$

- The skill score s of a certain accuracy measure A, with R denoting the reference irradiance

$$s = 1 - \frac{A(X,Y)}{A(R,Y)}$$

2.3 Models investigated

Our bibliography study leads us toward the most relevant models to be tested.

Concerning the reference model, Lorenz et al. and others suggested the use of the persistence model, consisting in taking as prediction the latest measure available, but this model turned out to have too poor results to be a good reference model. I opted for using the raw forecasted value as the reference model.

Suksamosorn et al. proposed a really interesting linear model based on a Kalman filter scheme. Hence the Kalman filter was first used as a promising linear model to be assessed against heavier non-linear machine learning models.

On the hand of machine learning models, Verbois et al. distinguished the models effective to reduce the RMSE, including multi-layer perceptron (MLP) and gradient boosting machine (GBM), and the models promising for reducing the MAE, notably the standard vector regression (SVR). Suksamosorn et al. also pointed out the effectiveness of the random forest (RF) model for a RMSE-optimization.

It's why I am going to compare the following models, against the reference raw forecasted irradiance.

- Kalman filter model (KF).

  The Kalman filter is a recursive estimator. This means that only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state.

  The correction procedure involves two groups of equations: time update equations and measurements update equations, time update equations are responsible for making a first guess of the next solar irradiance prediction error, based on the last state of the measured error and error covariance estimates, obtaining an a priori prediction for the next time step; the measurement update equations will then incorporate new measurements into the first guess, obtaining improved a posteriori predictions.

  My understanding of the general Kalman filter was greatly thanks to Becker, and I practised the filter thanks to Labbe.

  In the context of irradiance forecasting, I followed the path from Suksamosorn et al..

- Gradient boosting machine model (GBM). GBM creates an ensemble of weak learners, meaning that it combines several smaller, simpler models in order to obtain a more accurate prediction than what an individual model would produce. Gradient boosting works by iteratively training the weak learners on gradient-based functions and incorporating them into the model as "boosted" participants.

  For more information, see notably Kumar for the theory and Bento (a) for the python practise.

- Support vector regression model (SVR).

  The support vector regression method is often used in cases where there are multiple input variables, each of which may have an effect on the output variable. The goal is to find the best linear combination of these input variables to predict the output variable.

To estimate the coefficients of the linear function, standard vector regression uses a method called least squares regression. This involves finding the values of the coefficients that minimize the sum of the squared differences between the predicted and actual values. Here is an interesting article on the subject: Sharp.

- Random forest model (RF).

  The Random Forest algorithm is an ensemble method used for machine learning. It creates multiple decision trees, each trained on a different subset of data and considering random features for splitting. The final prediction is made by combining the predictions of these trees through voting (for classification) or averaging (for regression), resulting in improved accuracy and reduced overfitting.

  Again, here is a link for a hands-on practise of the RF algorithm: Bento (b).

- Multiple-layer perceptron model (MLP).

  The Multilayer Perceptron (MLP) is a type of artificial neural network used in machine learning. It consists of multiple layers of interconnected nodes (neurons) where each node computes a weighted sum of its inputs, passes it through an activation function, and then forwards the result to the next layer. MLPs are commonly used for various tasks such as classification, regression, and pattern recognition, and they can learn complex relationships in data. They can be trained using backpropagation, adjusting the weights between nodes to minimize the difference between predicted and actual outputs.

All this models will be compared during the internhip, and all the data wrangling architecture around it can be found either in the README or more specifically in the source code of my repo ACCOU.

After having post-processed a single NWP forecast model individually, the next step will be to assess the performances of our hybrid model in comparison to the one currently used by Reuniwatt (LT CONT).

2.4 Model performances evaluation strategy

The benchmarking consists in evaluating the performances on each metrics of each one of the model optimized with the corresponding metrics.

A grid search which details can be explored in ACCOU is used for the learning year in order to find the best hyperparameters for any of the model.

We assess the performances of the trained models on their performances on the test year.

The big picture will be given by a global significance matrix that will compare all the models performances regarding the particular metrics across the 4 sites. This matrix will allow us to discern the most pertinent models for each of our study metrics. Then, to verify that the models indeed perform well in the detail and across the different times of the day, we are going to plot scatter plots and data distributions of the MBE for each site.

This dual approach will ensure us that global results indeed translate into improved performances for each hour of the day, and especially in term of bias (MBE).

3 Results and discussion

As explained in section 2, the first objective is to effectively post-process a single NWP forecast model, by benchmarking the models altogether.

The following step will be to look at the results in the details to prove the true effectiveness of the post-processing.

We will each time first study the MAE optimisation and then the RMSE optimisation. Be aware that the models hyperparameters may vary between the two optimisations because the target metrics to minimize is not the same during the computations.

3.1 Post-processing of a single forecasting NWP model

3.1.1 Study of the models altogether

MAE

Figure 2: Significance matrix for MAE. The value $V_{(i,j)}$ of the (i, j) cell indicates how often the model of line i performs better than the one of column j, across the 4 sites. For example, the MAE of the MLP model post-processed data is 3 times lower than the MAE of the reference model, and for 1 site (4 - 3 = 1), it is higher.

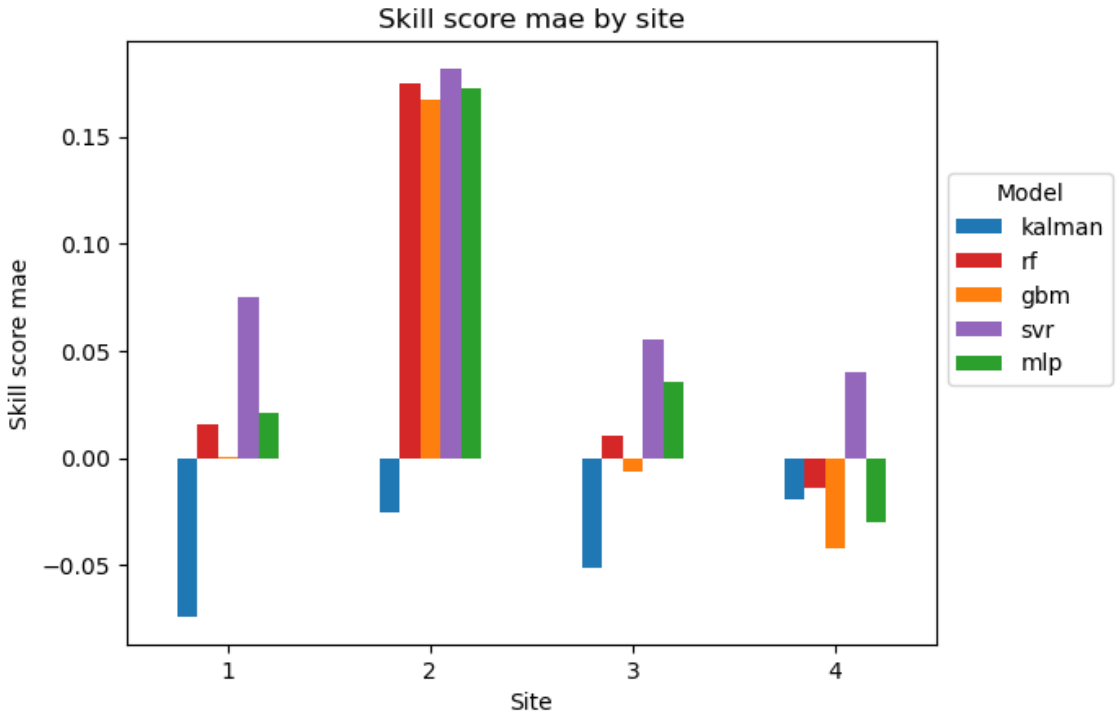


Figure 3: MAE skill score plot across the 4 sites.

Figure 2 clearly shows that the SVR model is the best one for MAE. It performs better than any of the other model on any of the 4 study cases.

Figure 3 demonstrates that the sites 1, 3 and 4 heavily benefit from this model with respect to the other ones. On site 4, the only positive post-processing is given by the SVR model.

The Kalman filter showcases really poor performances across the 4 sites.

RMSE

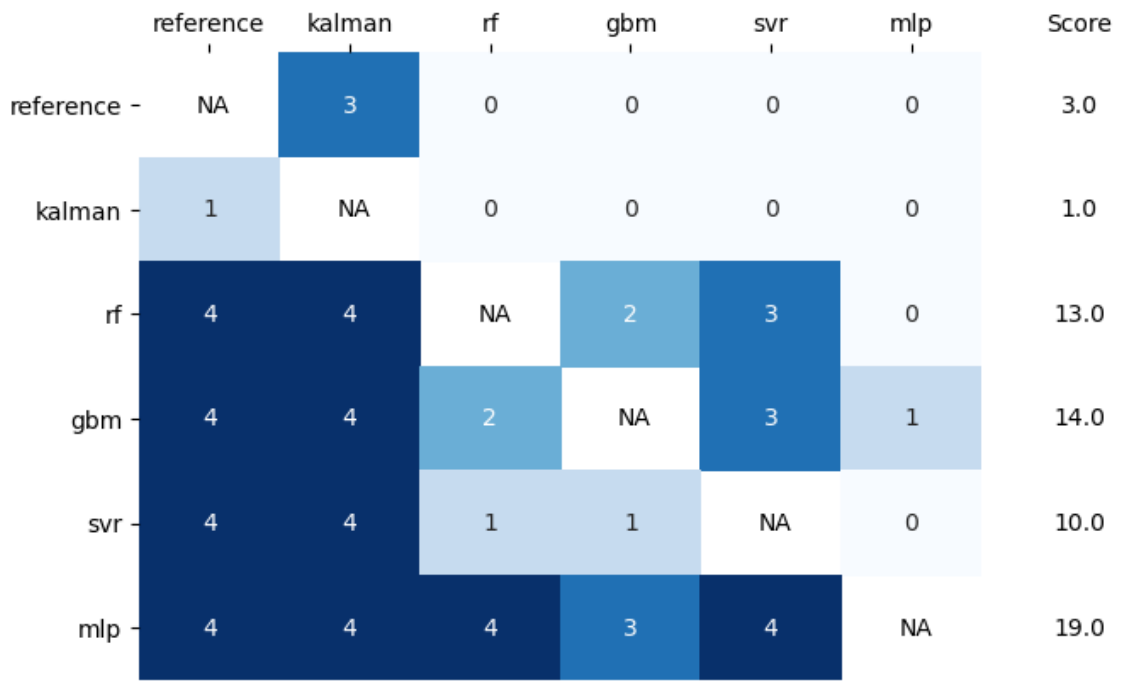


Figure 4: Significance matrix for RMSE. The value $V_{(i,j)}$ of the (i, j) cell indicates how often the model of line i performs better than the one of column j, across the 4 sites. For example, the RMSE of the MLP model post-processed data is 3 times lower than the RMSE of the reference model, and for 1 site (4 - 3 = 1), it is higher.

The results of the RMSE are not the same, and it is the MLP model that performs the best, achieving the highest score in the matrix of Figure 4.

This is confirmed by Figure 5 where the MLP model bar is the highest for 3 sites out of 4.

3.1.2 Detailled study of the most performing model

With the aim of clarity, only the plots of the single site 2 will be shown here. The overall similarity of the results across the 4 sites also motivate this choice.

The ones of the other sites can be found in the appendix to fortify the belief in the analysis drawn for a single site.

MAE

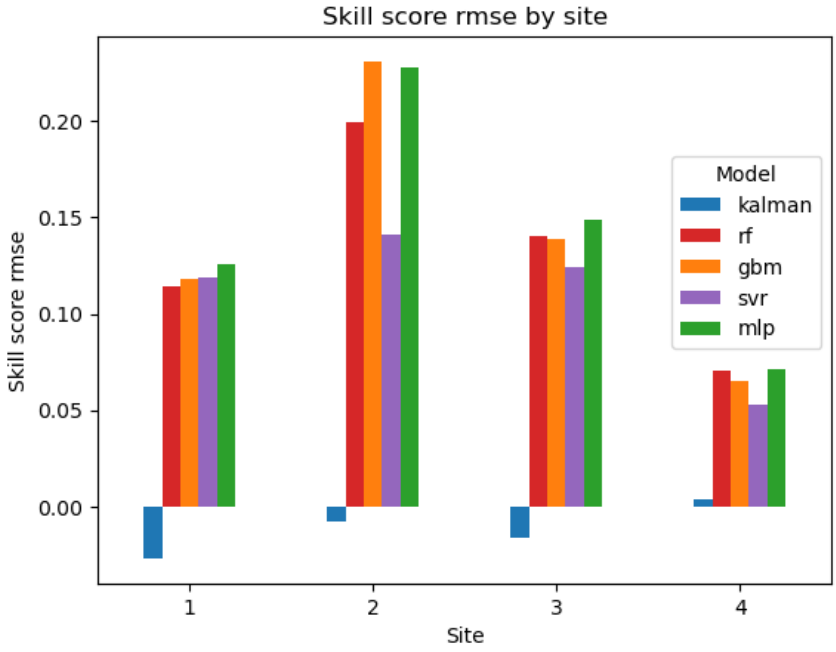


Figure 5: RMSE skill score plot across the 4 sites.

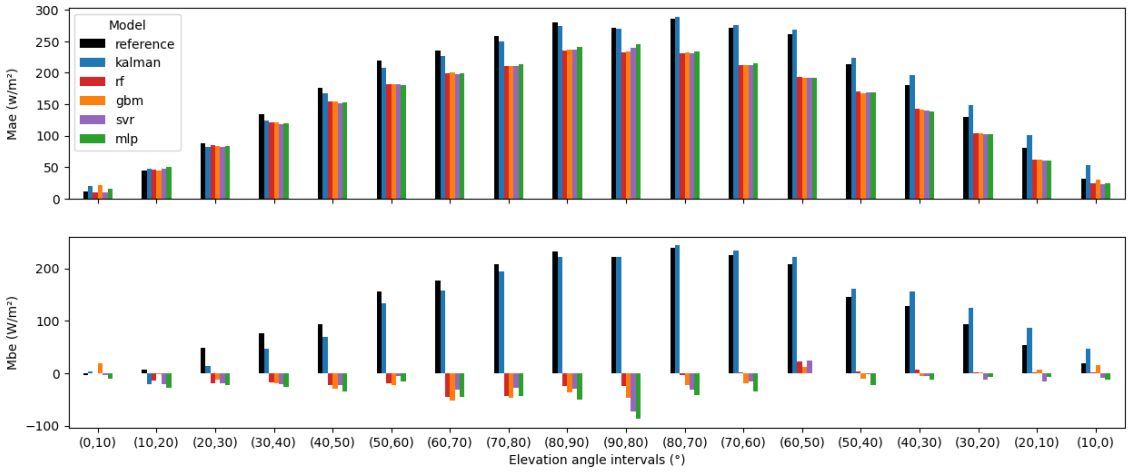


Figure 6: MAE and MBE levels across all elevation angle intervals of a day, for site 2.

?? supports our previous conclusions, but most importantly shows the decrease of the MBE across all the elevation angle intervals of a day. Indeed, it is a know fact that NWP models present a positive bias that is higher around noon, and the

MBE of all the machine learning models tends to be much closer to 0.

On site 2, the bias is negative but this bias can be either positive or negative depending on the study site, as confirmed with the other graphs in Appendix A. Only site 4 shows a different behavior for the bias across the day, but this does not change the MAE results.

The probability density functions across the day support these conclusions, where we can clearly see that the curve is shifted towards 0, as are the scatter plots in Appendix A. The whole dataset cloud is shifted towards the line y=x, which is the line where the corrected forecast exactly match the measured irradiance.

These plots about the MAE minimisation thanks to the SVR model show that this model is not only able to reduce the MAE across the day, but also to reduce the MBE across the day.

The global lowering of the global metrics is indeed reflected in the lowering of the metrics at each elevation angle interval of particular day. The bias is also improved in 3 sites out of 4.

RMSE

Similar conclusions can be drawn about the RMSE minimisation, with the MLP model.

The only difference is that the improvement of the MBE now stands for all the study cases.

All in all, the improvement of the metrics of interest (MAE and RMSE) is reflected in an improvement of the bias.

This behavior stands for each period of a day, which was not perfectly obvious after I performed the global post-processing.

Once the relevant models have been found, one has to perform a sensitivity study of these ones in order to better grasp how they work.

3.1.3 Sensitivity study

It is also necessary to perform a sensitivity study of the parameters that are not tuned in our process. It is why the influence of the choices of predictors, of learning periods, of learning window type (fixed or sliding) and of forecasting model are successively performed.

With the aim of clarity, the results will be here presented with the MAE optimisation, but the results of the RMSE optimisation lead the same conclusions, and can also be found in Appendix A.

It's important to notice that for each sensitivity study, only the one parameter investigated is modified from the framework defined in section 2, all the rest is kept constant.

Influence of the choice of the predictors The first sensitivity study is performed with the set of predictors that was presented in Table 1.

	0	1	2	3
ghi_{GFS}	X	X	X	X
$temperature^{2m}_{GFS}$	X	X		
θ	X	X	X	
ϕ	X	X	X	
ghi_{cs}	X		X	

Table 2: Description of the configurations of the sets of predictors.

Interestingly, the configuration with all the predictors is not the configuration that seem to give the best results if we consider Figure 10.

If we consider now 11, the performances of the configurations 0, 1 and 2 are nearly identical. They both provide great improvements to the post-processing using only the forecasted irradiance.

In the following, I chose to always rely on the full configuration of the set of predictors. This is because I was working with only 5 predicting values and a feature selection is generally not advised when working with so few predictors. This is a strong choice that could be questioned if the situation made to work with much more predictors (more than 10), where a feature selection would be necessary, both for metrics and computational performances reasons.

Influence of the learning period Both Figure 12 and Figure **??** go to show that a larger learning period is beneficial for the model, up to one year when the benefits from increasing the learning period do not prove to be significant.

It is interesting to notice that a one-month learning period provides a negative post-processing.

Influence of the window of learning Another question from Reuniwatt was the pertinence of a sliding learning window.

I thus performed two types of learning:

- A fixed-window learning where the testing period was the full year of 2021 and and the learning period was the full year of 2020.
- A sliding-window learning where each month of the test 2021 year used a model trained during the moving year that preceded this month. For instance, March 2021 used a model trained from February 2020 to February 2021.

Interestingly, the fixed window performs best that the sliding window for all the study cases.

Influence of the NWP forecasting model All the previous optimisations were done on the GFS forecast, as it was explained in section 2.

Reuniwatt uses the data from several different numerical weather prediction (NWP) models, including ECMWF, AROME and ARPEGE. It is interesting to see how our models perform on this different data source.

Figure 16 clearly shows that the NWP models differently benefit from the post-processing. There seems to be higher skill scores for less elaborate models such as AROME in comparison to more advanced models such as ECMWF. The higher is the raw forecast metrics, the better is the post-processing.

This is confirmed by **??** where it can be clearly seen that post-processed metrics fall in a much thiner range than the initial one. It is worthwhile to point out that the ECMWF forecast still remains the best forecast after post-processing.

Now that each forecast model has been post-processed, it would be interesting to hybride them so as to compare the performances against the current LT CONT model used by Reuniwatt.

Before doing it, we need to find another linear model to compare against the machine-learning models, since the Kalman filter did not prove its effectiveness.

We are now going to discuss which linear regression model suits best for each one of our metrics of interest.

3.2 Benchmarking the linear regression models

We investigate the linear models that are implemented inside the skikit-learn library, and that are listed on their website scikit learn.

Alike to what we did for our machine learning models, a grid search is performed to find for each model evaluated which set of hyperparameters is the best.

MAE Figure 18 shows that the best model for MAE is the stochastic gradient descent. The grid search allowed us to select the best set of hyperparameters of

this algorithm. The loss function is the "epsilon insensitive" function and the other hyperparameters are the default ones given by scikit-learn.

Figure 19 confirms the differences between this model and classical models such as the "LinearRegression" that implements the ordinary least squares method to find the optimal regression parameters. This more classical method will perform best for a RMSE minimisation as we are going to see.

RMSE While Figure 20 seems to indicate that some models such as "TweedieRegressor" may perform really better than 'LinearRegression', a closer look thanks to Figure 21 shows that models from 'LinearRegression' up to 'TweedieRegressor' has really comparable performances, indeed almost identical. It's why it is the 'LinearRegression' model that is kept for a RMSE minimisation because of its unbeatable computation time.

Now that the appropriate linear model to use has been fixed, we can compare the results of our models with respect to the one currently used by Reuniwatt.

3.3 Showcase of the hybrid model

The efficiency of both the SVR model and the MLP model has been proven for post-processing a single NWP forecast model, respectively for MAE and RMSE.

It is also important to notice that, while the MLP model and the linear model have really similar performances, the SVR model performs much better than the linear model in terms of MAE. The preference was given to the MLP model over the linear model concerning the RMSE minimisation because the MLP model can be constantly improved whereas the linear model is fixed.

The following models are going to be assessed in this part:

- The linear model (different according to the considered metrics). It is learnt with the initial set of predictors described in Table 1.
- The best machine learning model (ie SVR for MAE and MLP for RMSE) learnt with the initial set of predictors. The NWP forecast model used for the irradiance used in the predictors set in the best one, in practise this means the ECMWF model.
- The best machine learning model learnt with the concatenation of the initial set of preditors and all the other NWP models irradiance forecast available. This constitutes the big model where all the data is fed into the model.
- The LT CONT model, learnt with the raw NWP models data.

- The LT CONT POST model, learnt with the improved NWP models data.

As depicted in Figure 22, it was necessary to switch from the 2020-2021 period to the 2020-2021-2022 period, because of the need of LT CONT POST of post-processed NWP data.

Following the flow of the data that was available on my machine, I first performed the analysis on the 4 inital sites, and then performed a validation study in 25 additional sites, located in Germany.

3.3.1 Study on the four inital sites

MAE Figure 23 shows that the MAE skill score is improved by a few percent both by the LT CONT POST and the ALL SVR model, which is the big model that takes as predictors all the NWP models. THE LT CONT POST and the ALL SVR model share nearly the same performances.

Figure 24 confirms that the LT CONT POST is able to improve the already improved ECMWF forecast, which was not obvious as a starting point.

RMSE Figure 25 shows a less obvious contribution to our models comparatively to the LT CONT model for the 3 first study sites. However, site 4 is really promising since the skill score our models nearly doubles the skill score of the LT CONT model.

All in all, our 4 initial study cases demonstrated the potential benefits of our models, being the LT CONT POST and the big machine learning model, with respect to the LT CONT model. More validation sites are needed to be totally convinced of the prevalence of any of our model over the LT CONT model. What is more, it's still difficult to say which one of our 2 models is the best.

3.3.2 Study on the German sites

The following study cases are now investigated, all located in Germany.

MAE One has to be well aware that the reference model is now the LT CONT model. This way, it is easier to see the benefits of our models with respect to the LT CONT model.

Figure 28 clearly shows that any of the 25 validation sites benefits from both models, since all the skill scores are positive.

It is really interesting to note that LT CONT POST performs much worse than the big SVR model for these sites, clarifying our questions from subsubsection 3.3.1.

RMSE Concerning the RMSE study, the MLP model performed worse than LT CONT for nearly all of the 25 models, but the big linear model actually provides minor improvements over all the 25 sites.

It is highly interesting to see that, despite showcasing similar performances in subsubsection 3.3.1, the MLP and linear models perform differently when it comes to the big study.

This may be due to a too simple hyperparameters search space for the MLP that can't perform well enough when scaling the model.

It is now mandatory to check the good performances of the big linear model for our 4 inital study case, because the linear model studied in subsubsection 3.3.1 did not make advantage of all the NWP models for learning. Figure 30 indeed shows the benefits of the big linear model also for the 4 initial study sites.

4 Conclusions and perspectives

4.1 Results summary

- The Kalman filter has not been proven interesting enough for post-processing day-ahead irradiance forecasts, in opposition to what was said in Suksamosorn et al. paper. The difference may be due to the fact that we used a more precise GFS data in comparison to the WRF data used in the paper. What is more, I dealed with the extreme case of day-ahead data with origin 00:00 UTC, while the paper fixed the origin at 13:00 (local hour).

  A more appropriate use of the Kalman filter would be for filtering faulty data, as explored in **?**.

- For a MAE minimisation, the SVR model clearly yielded the best results, both for post-processing a single NWP model and for hybridation. This confirms Verbois et al..

- For a RMSE minimisation, there is no clear consensus for post-processing a single NWP model, as it was the case in the literature where the random forest (Suksamosorn et al.), GBM and MLP models (Verbois et al.) were deemed promising. Still the MLP model as well as the linear least square regression model stand out.

  The hybridation showed some limits of the MLP model, in favor of the linear model.

- Using the LT CONT model on post-processed data only yielded marginal enhancements in metrics, whereas using the adequate model (SVR for MAE, Linear regression for RMSE) on the whole dataset provided significant improvements as compared to the LT CONT model.

4.2 Suggestions for future improvements

4.3 My learnings from the internship

References

M. ACCOU. My intership gitlab repo. URL `https://gitlab.soleka.org/maccou/maccou`.

A. Becker. Kalmanfilter.net. URL `https://www.kalmanfilter.net/default.aspx`.

C. Bento. Gradient boosted decision trees explained with a real-life example and some python code, a. URL https://towardsdatascience.com/gradient-boosted-decision-trees-explained-with-a-real-life-example-and-some-python-code-77cee4ccf5e.

C. Bento. Random forests algorithm explained with a real-life example and some python code, b. URL https://towardsdatascience.com/random-forests-algorithm-explained-with-a-real-life-example-and-some-python-code-affbfa5a942c.

A. Kumar. Gradient boosting algorithm: Concepts, example. URL https://vitalflux.com/gradient-boosting-algorithm-concepts-example/?utm_content=cmp-true.

R. R. Labbe. Kalman and bayesian filters in python. URL https://github.com/rlabbe/Kalman-and-Bayesian-Filters-in-Python.

E. Lorenz, J. Remund, S. C. Müller, W. Traunmüller, G. Steinmaurer, D. Pozo, V. Lara, L. Ramirez, M. G. Romeo, C. Kurz, L. M. Pomares, and C. G. Guerrero. Benchmarking of different approaches to forecast solar irradiance.

M. J. Mayer and D. Yang. Calibration of deterministic NWP forecasts and its impact on verification. 39(2):981–991. ISSN 01692070. doi: 10.1016/j.ijforecast.2022.03.008. URL https://linkinghub.elsevier.com/retrieve/pii/S0169207022000486.

scikit learn. Scikit-learn regression models. URL https://scikit-learn.org/stable/modules/linear_model.html.

T. Sharp. An introduction to svr. URL https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2.

S. Suksamosorn, N. Hoonchareon, and J. Songsiri. Post-processing of NWP forecasts using kalman filtering with operational constraints for day-ahead solar power forecasting in thailand. 9:105409–105423. ISSN 2169-3536. doi: 10.1109/ACCESS.2021.3099481. URL https://ieeexplore.ieee.org/document/9494359/.

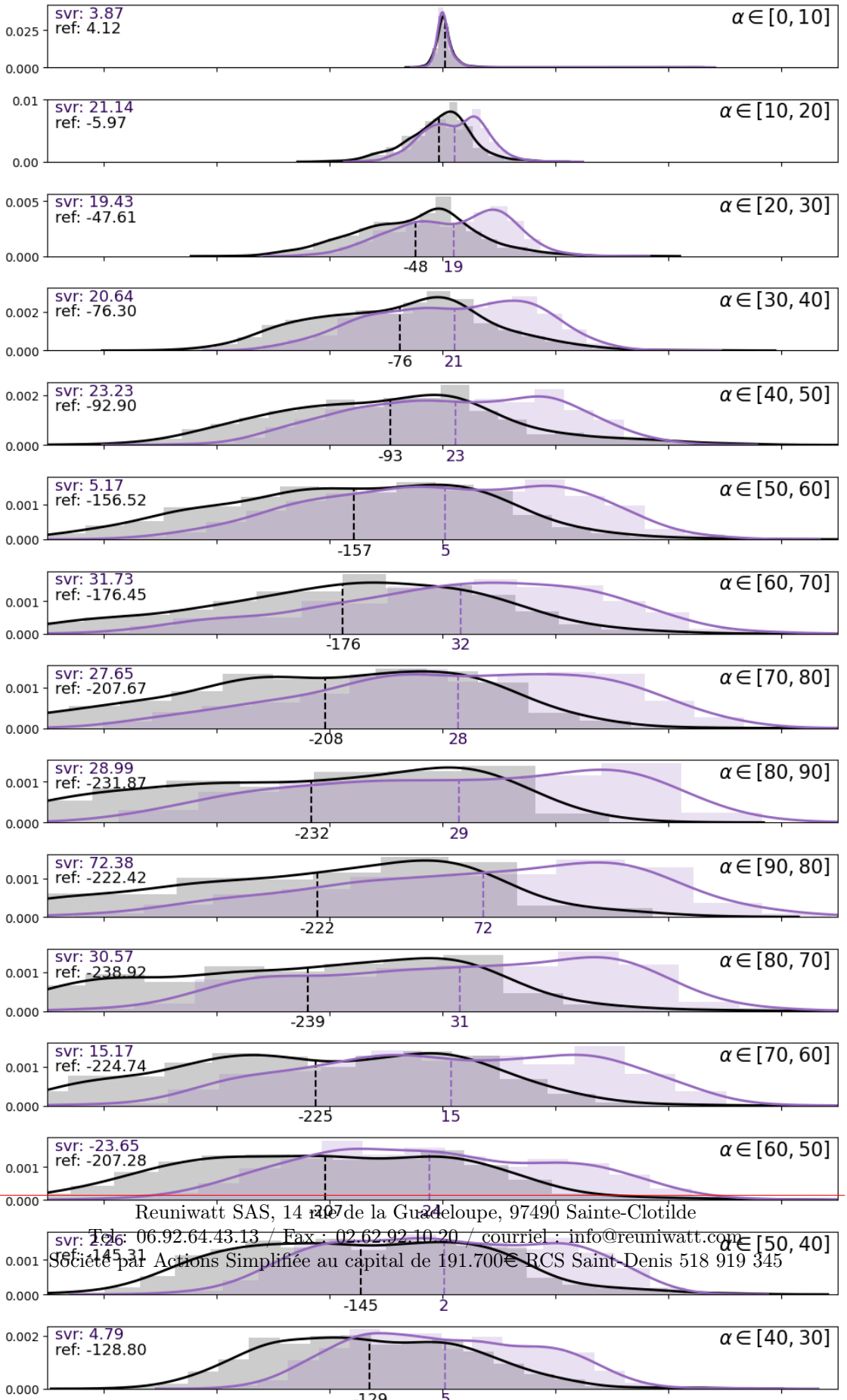
H. Verbois, Y.-M. Saint-Drenan, A. Thiery, and P. Blanc. Statistical learning for NWP post-processing: A benchmark for solar irradiance forecasting. 238:132–149. ISSN 0038092X. doi: 10.1016/j.solener.2022.03.017. URL https://linkinghub.elsevier.com/retrieve/pii/S0038092X22001839.

A Additional results of post-processing a single NWP forecasting model

MAE

RMSE

B Filtering of the measures


svr: 3.87
ref: 4.12
α∈[0, 10]
svr: 21.14
ref: -5.97
α∈[10, 20]
svr: 19.43
ref: -47.61
-48
19
α∈[20, 30]
svr: 20.64
ref: -76.30
-76
21
α∈[30, 40]
svr: 23.23
ref: -92.90
-93
23
α∈[40, 50]
svr: 5.17
ref: -156.52
-157
5
α∈[50, 60]
svr: 31.73
ref: -176.45
-176
32
α∈[60, 70]
svr: 27.65
ref: -207.67
-208
28
α∈[70, 80]
svr: 28.99
ref: -231.87
-232
29
α∈[80, 90]
svr: 72.38
ref: -222.42
-222
72
α∈[90, 80]
svr: 30.57
ref: -238.92
-239
31
α∈[80, 70]
svr: 15.17
ref: -224.74
-225
15
α∈[70, 60]
svr: -23.65
ref: -207.28
α∈[60, 50]
svr: 2.26
ref: -145.31
-145
2
α∈[50, 40]
svr: 4.79
ref: -128.80
α∈[40, 30]

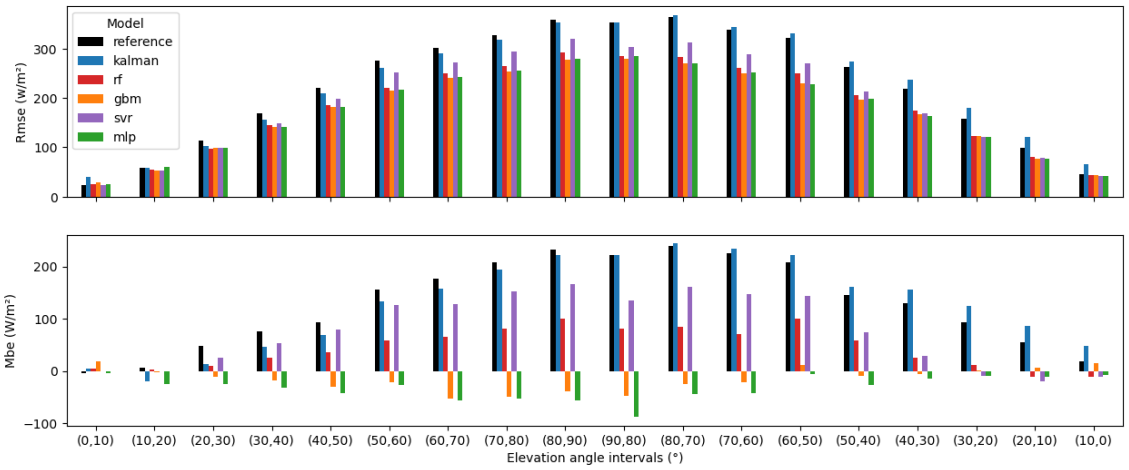
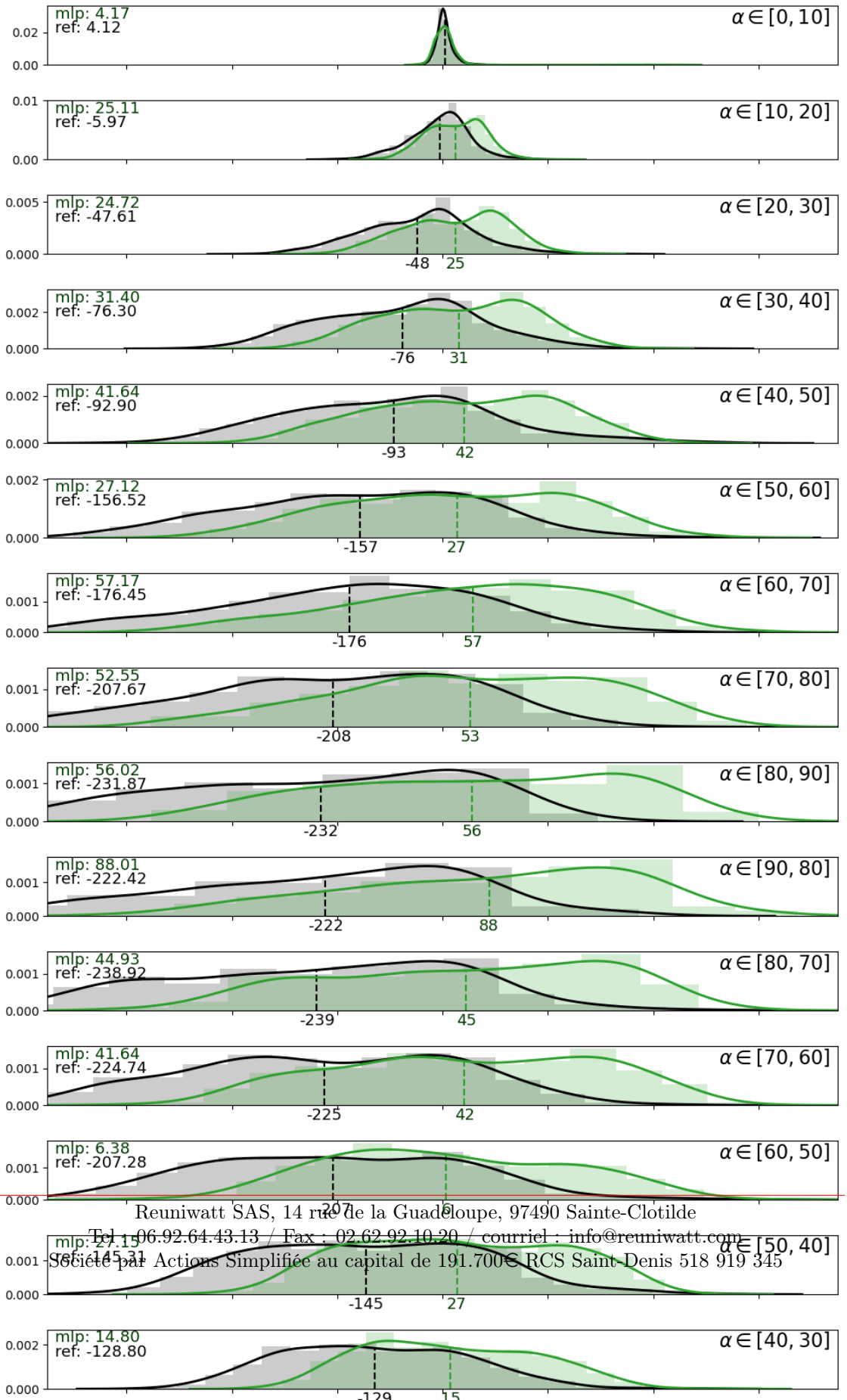


Figure 8: RMSE and MBE levels across all elevation angle intervals of a day, for site 2.


mlp: 4.17
ref: 4.12
α∈[0, 10]
mlp: 25.11
ref: -5.97
α∈[10, 20]
mlp: 24.72
ref: -47.61
-48
25
α∈[20, 30]
mlp: 31.40
ref: -76.30
-76
31
α∈[30, 40]
mlp: 41.64
ref: -92.90
-93
42
α∈[40, 50]
mlp: 27.12
ref: -156.52
-157
27
α∈[50, 60]
mlp: 57.17
ref: -176.45
-176
57
α∈[60, 70]
mlp: 52.55
ref: -207.67
-208
53
α∈[70, 80]
mlp: 56.02
ref: -231.87
-232
56
α∈[80, 90]
mlp: 88.01
ref: -222.42
-222
88
α∈[90, 80]
mlp: 44.93
ref: -238.92
-239
45
α∈[80, 70]
mlp: 41.64
ref: -224.74
-225
42
α∈[70, 60]
mlp: 6.38
ref: -207.28
α∈[60, 50]
mlp: 27.15
ref: -145.31
-145
27
α∈[50, 40]
mlp: 14.80
ref: -128.80
α∈[40, 30]

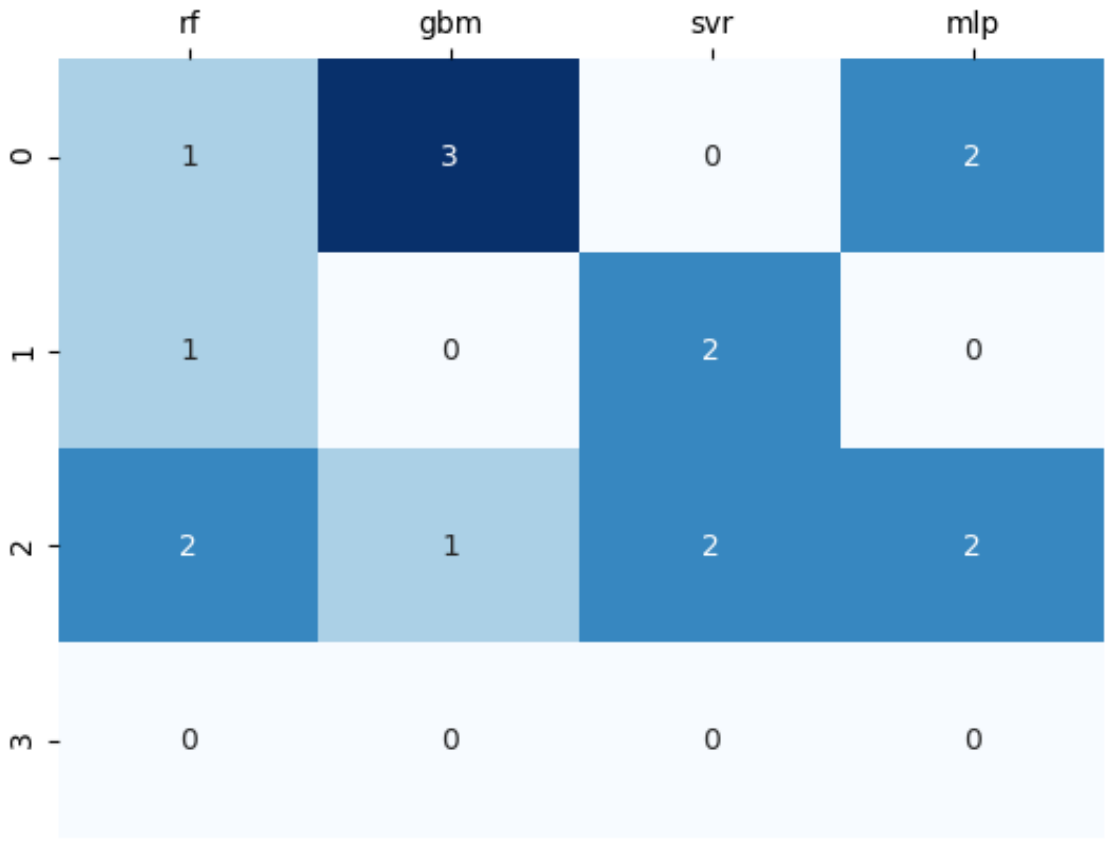


Figure 10: Pairwise systematicity matrix for MAE. The value $V_{i,j}$ of the cell (i, j) indicates how often the configuration of line i is the best one, across the 4 sites, for the model of column j. For example, the configuration 0 is the best one with a GBM post-processing for 3 sites, and the configuration 2 is the best one for 1 site.

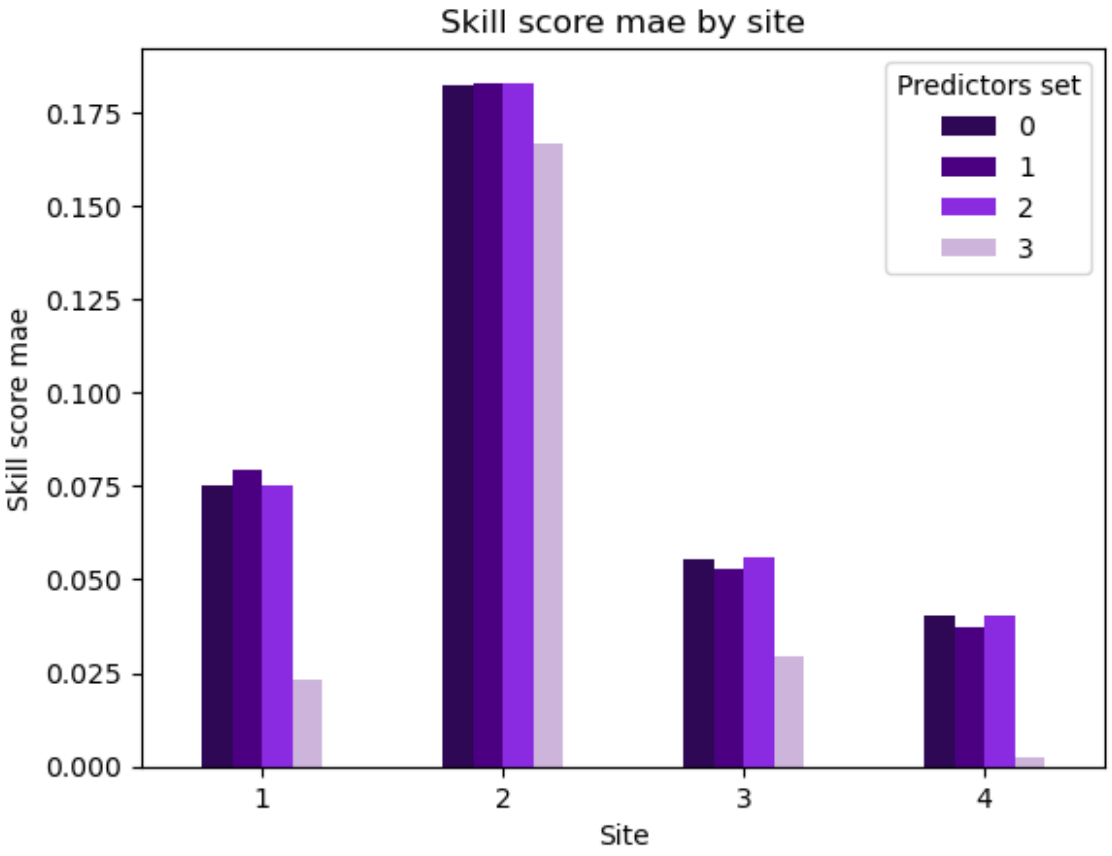


Figure 11: Comparison of the MAE skill scores of the different configurations.

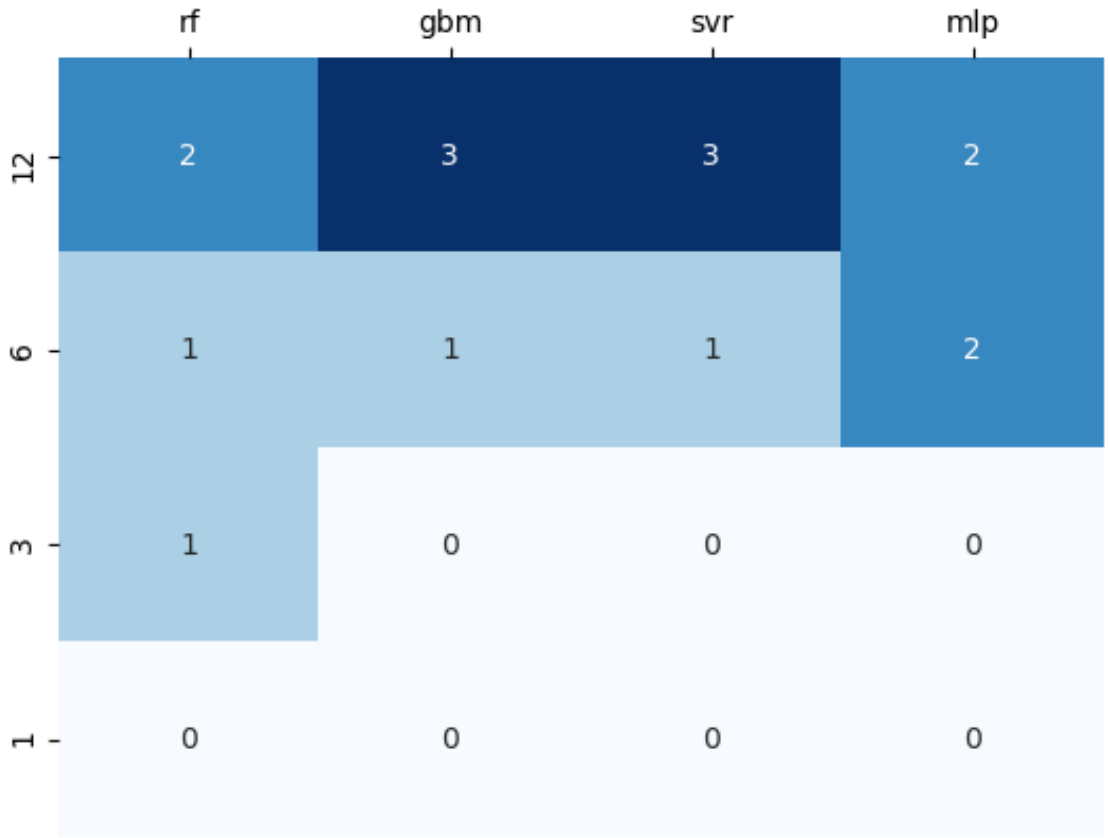


Figure 12: Pairwise systematicity matrix for MAE. The value $V_{i,j}$ of the cell (i, j) indicates how often the learning period duration (in months) of line i performs the best, across the 4 sites, for the model of column j. For example, having a 12-months-long learning period is the best thing in 3 sites out of 4 with a SVR model, the last site performs better with a 6-month-long learning period.

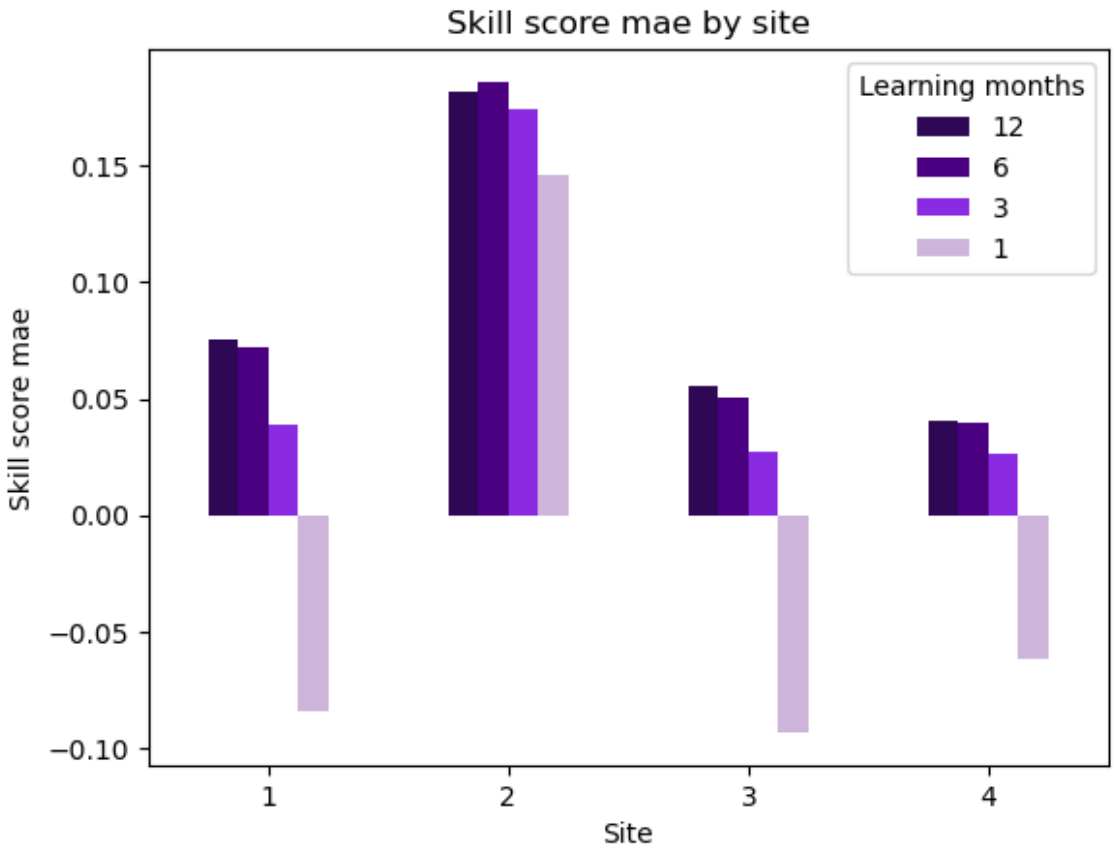


Figure 13: Comparison of the MAE skill scores of the different learning period durations (in months).

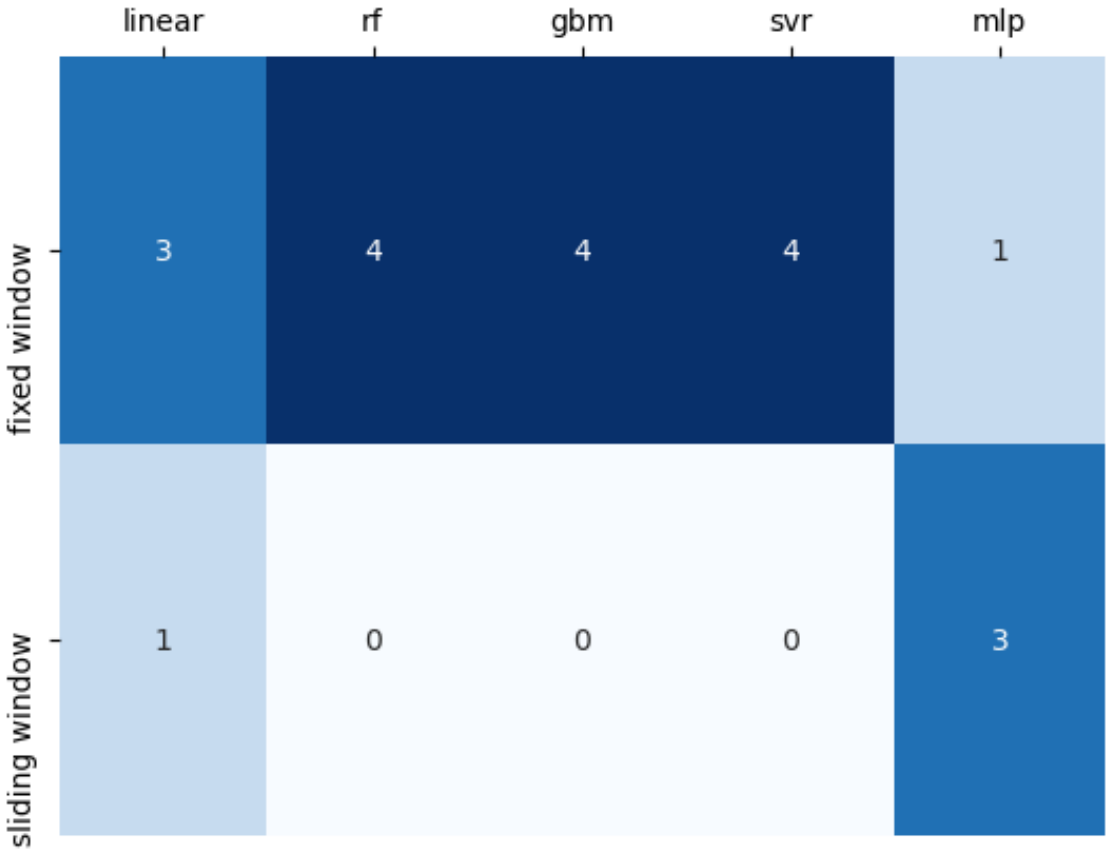


Figure 14: Pairwise systematicity matrix concerning window type for MAE.

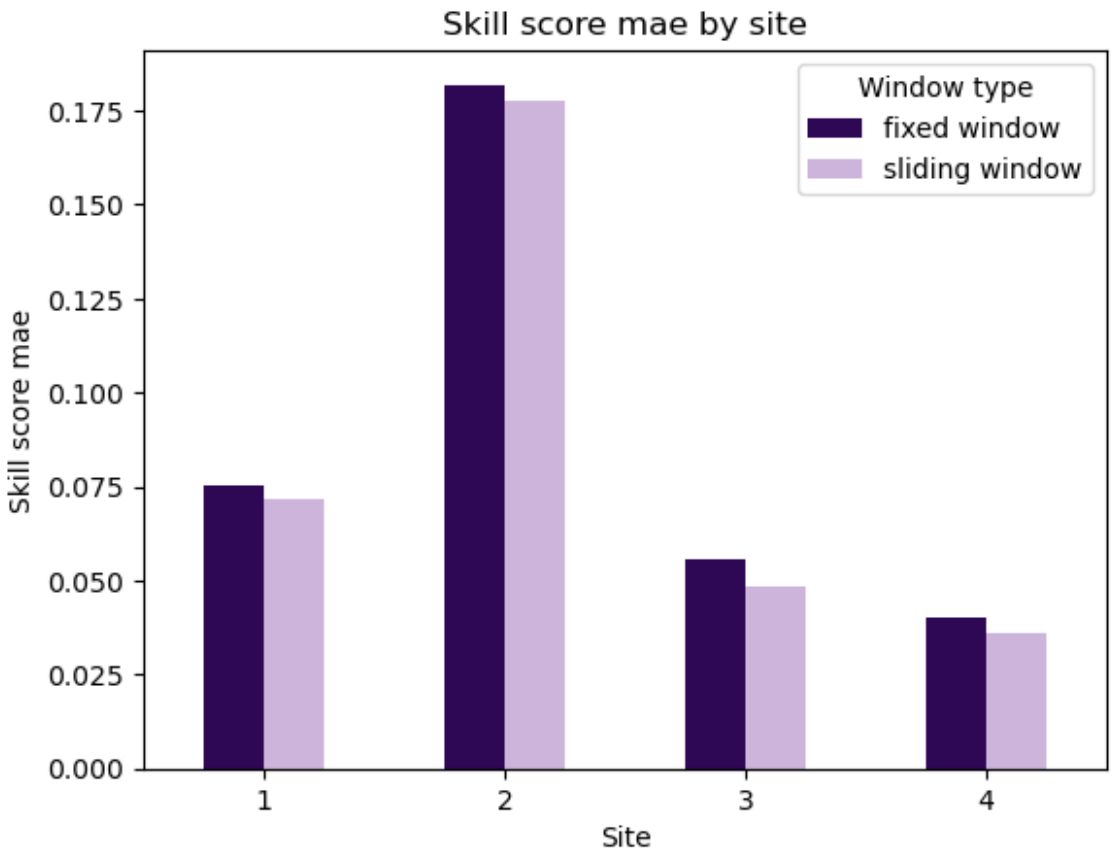


Figure 15: Comparison of the MAE skill scores for a SVR model.

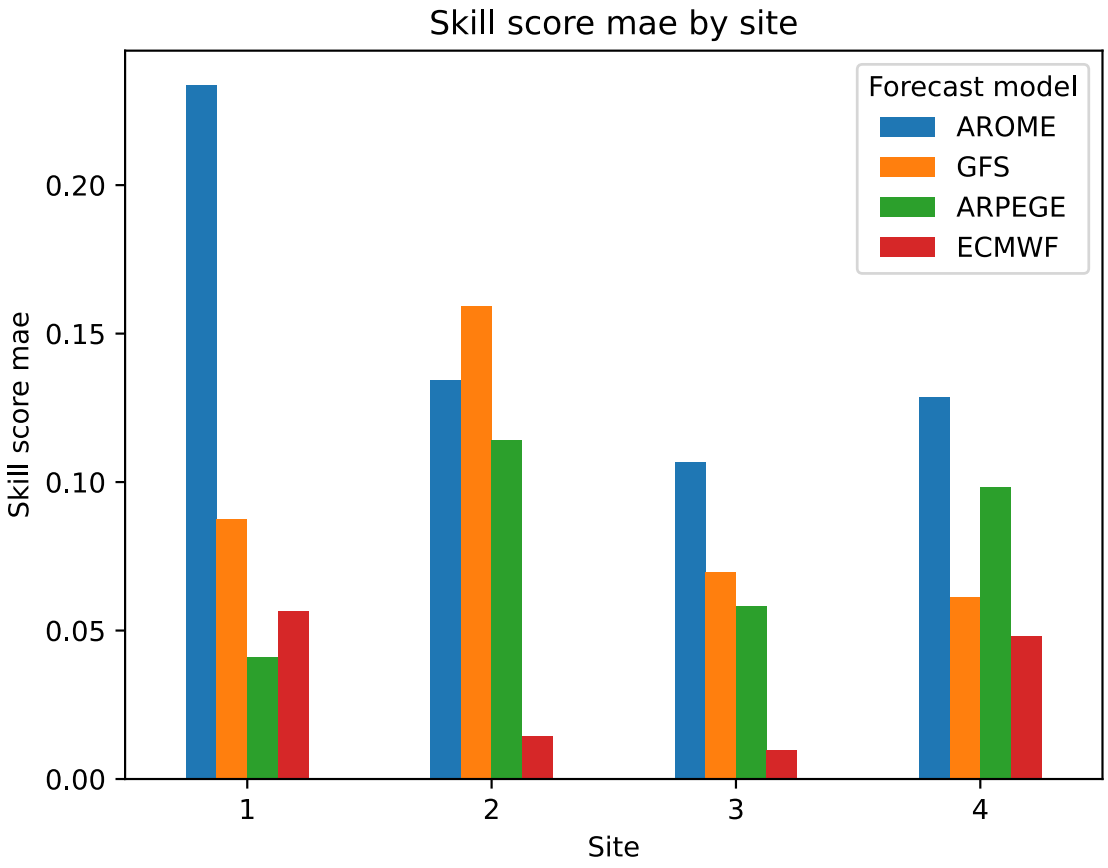


Figure 16: Comparison of the MAE skill scores of the SVR model of four different NWP forecast models.

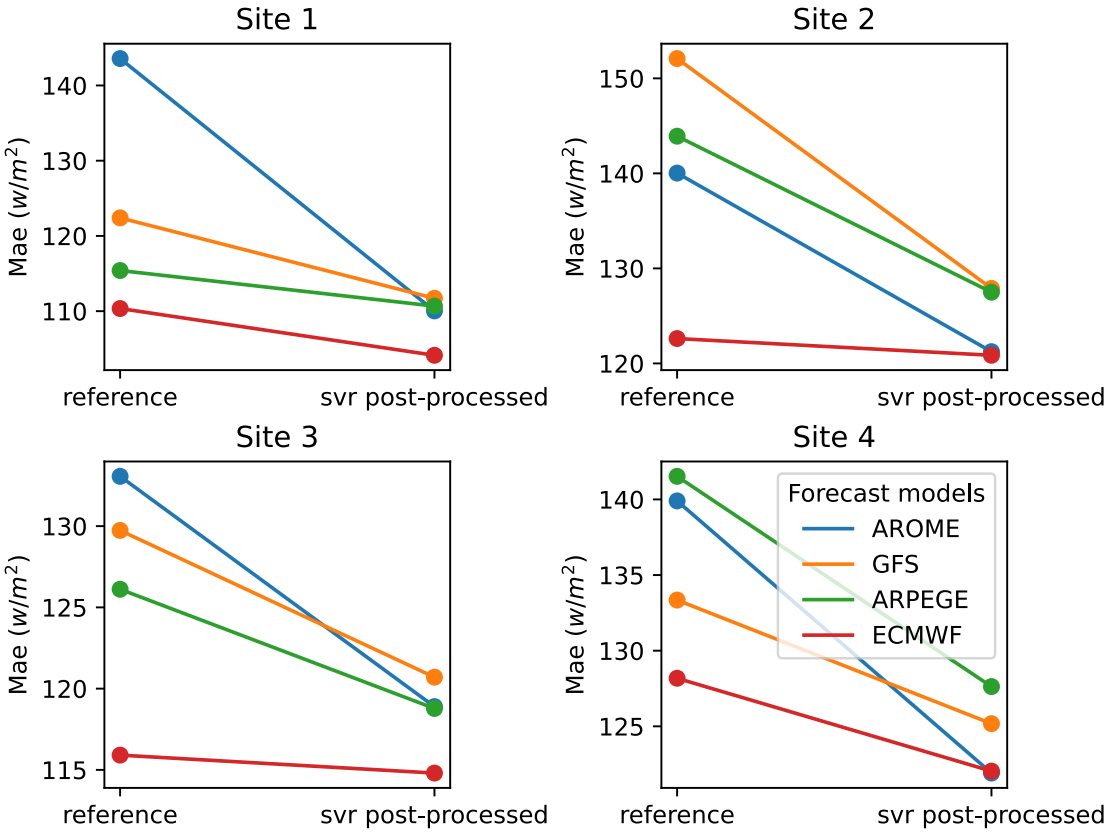


Figure 17: Comparison of the post-processing of four different NWP forecast models on MAE.

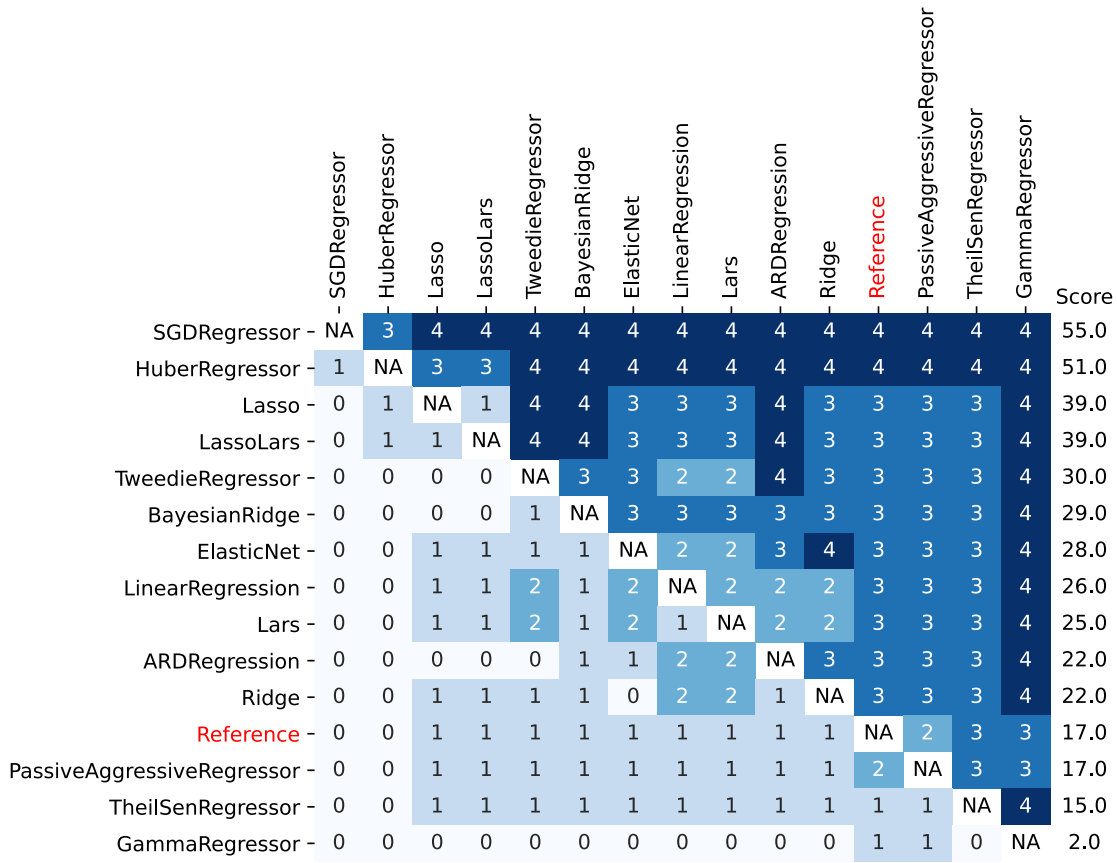


Figure 18: Significance matrix for MAE. Models at the top of the matrix are the most succesful ones, on the criterium of the sum of the values of each line. The value $V_{(i,j)}$ of the (i, j) cell indicates how often the model of line i performs better than the one of column j, across the 4 sites. For example, the MAE of the Lars model post-processed data is 1 times lower than the MAE of the Lasso model, and for 1 site (4 - 1 = 3), it is higher.

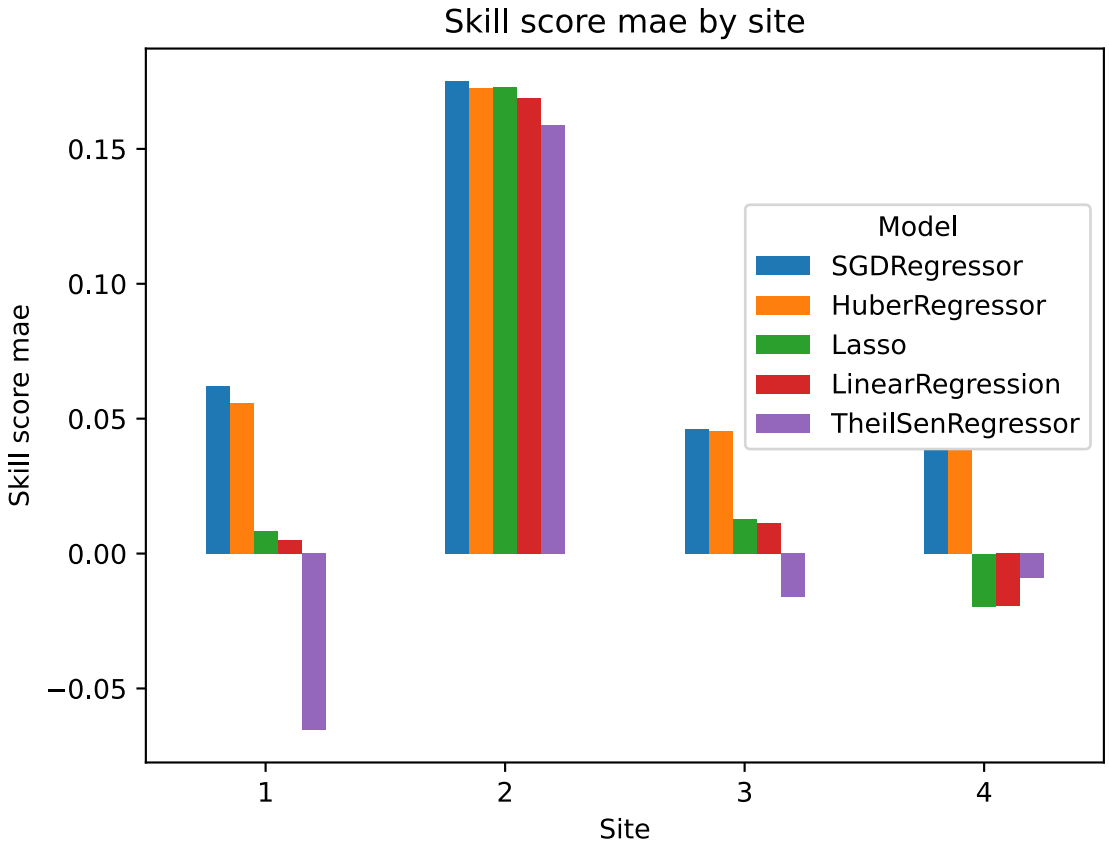


Figure 19: MAE skill score plots of the best linear models.

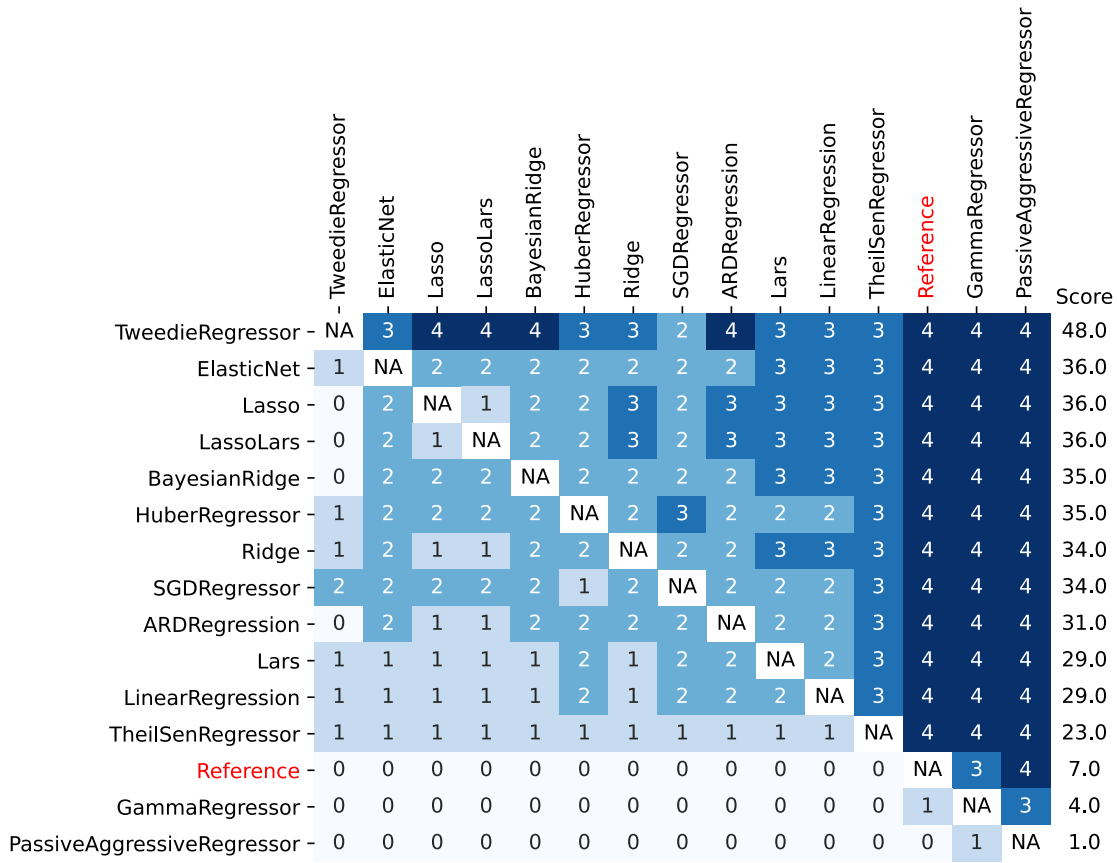


Figure 20: Significance matrix for RMSE. Models at the top of the matrix are the most succesful ones, on the criterium of the sum of the values of each line. The value $V_{(i,j)}$ of the (i, j) cell indicates how often the model of line i performs better than the one of column j, across the 4 sites. For example, the RMSE of the Lars model post-processed data is 1 times lower than the RMSE of the Lasso model, and for 1 site (4 - 1 = 3), it is higher.

Figure 21: RMSE skill score plots of the best linear models.

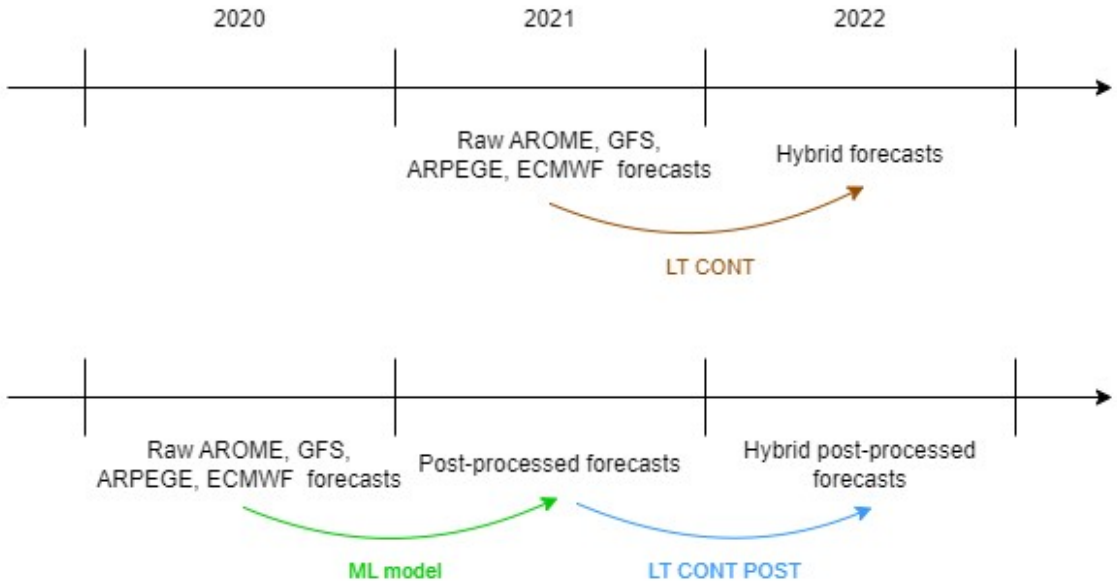


Figure 22: Comparison of the two versions of LT CONT.

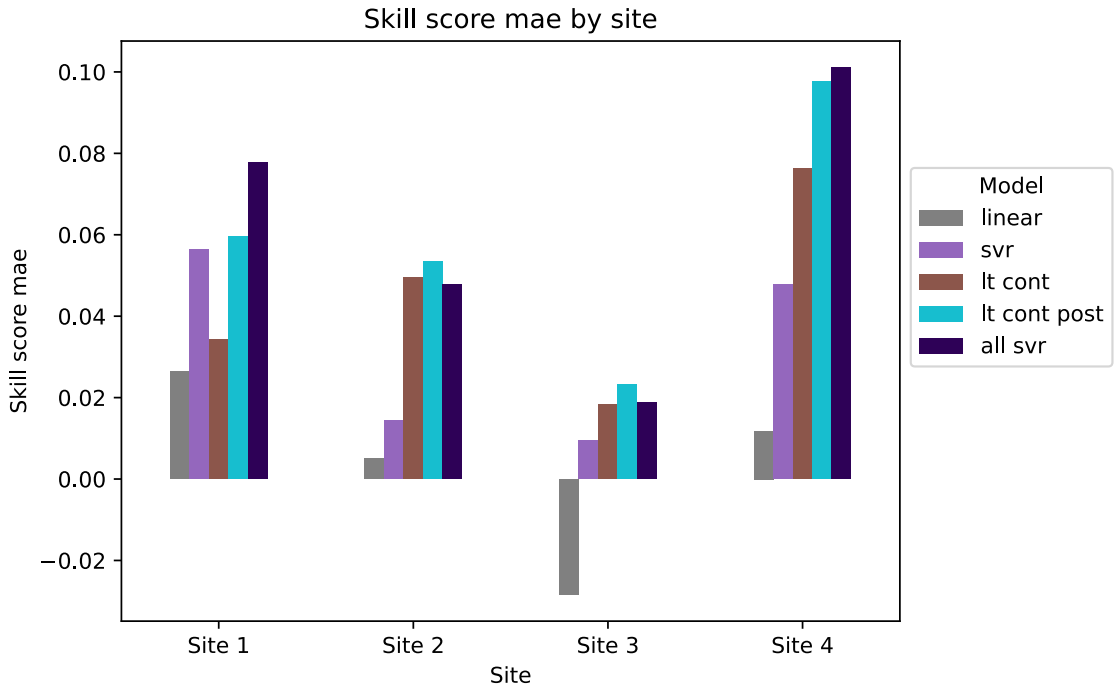


Figure 23: MAE skill score plots of the relevant models, the reference being here the best NWP forecast (in practise, this means ECMWF).

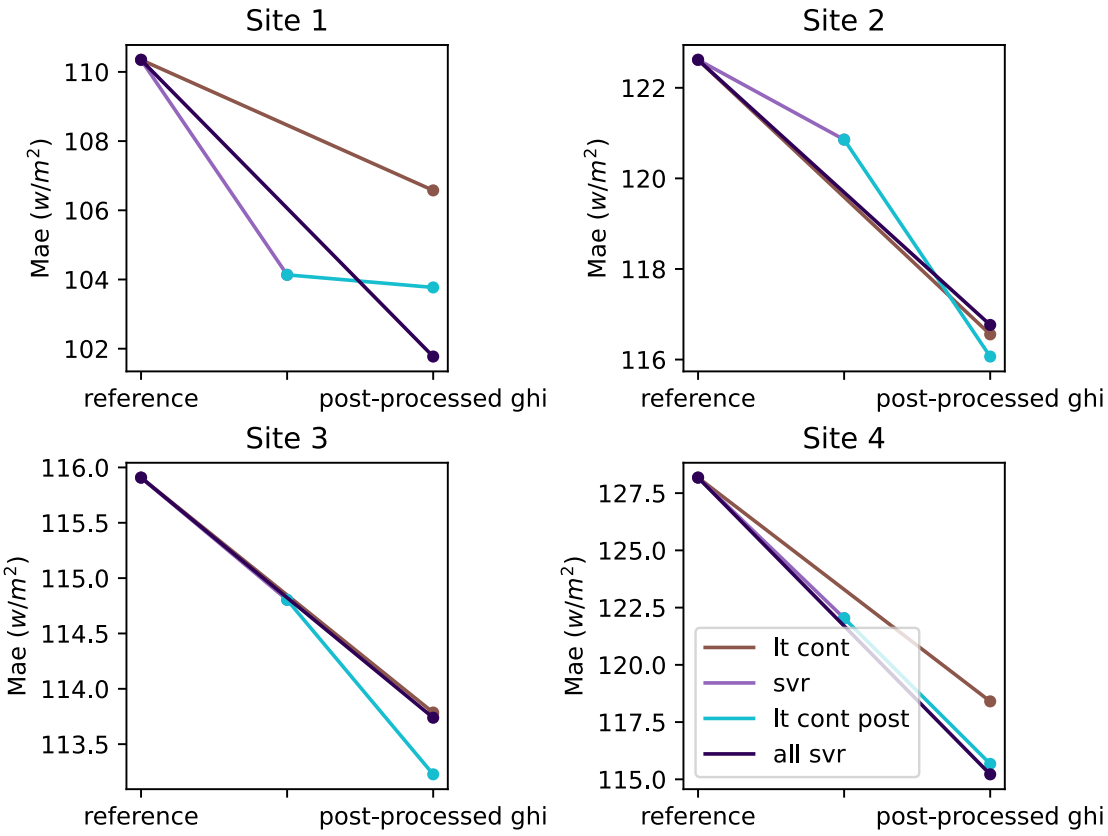


Figure 24: MAE evolution of the relevant models.

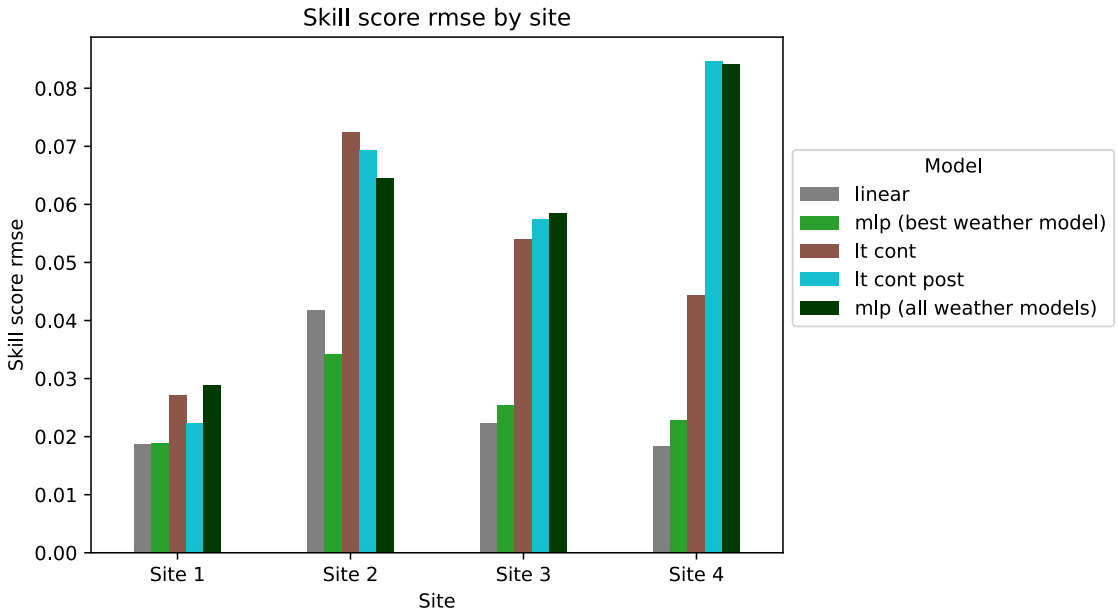


Figure 25: RMSE skill score plots of the relevant models, the reference being here the best NWP forecast (in practise, this means ECMWF).

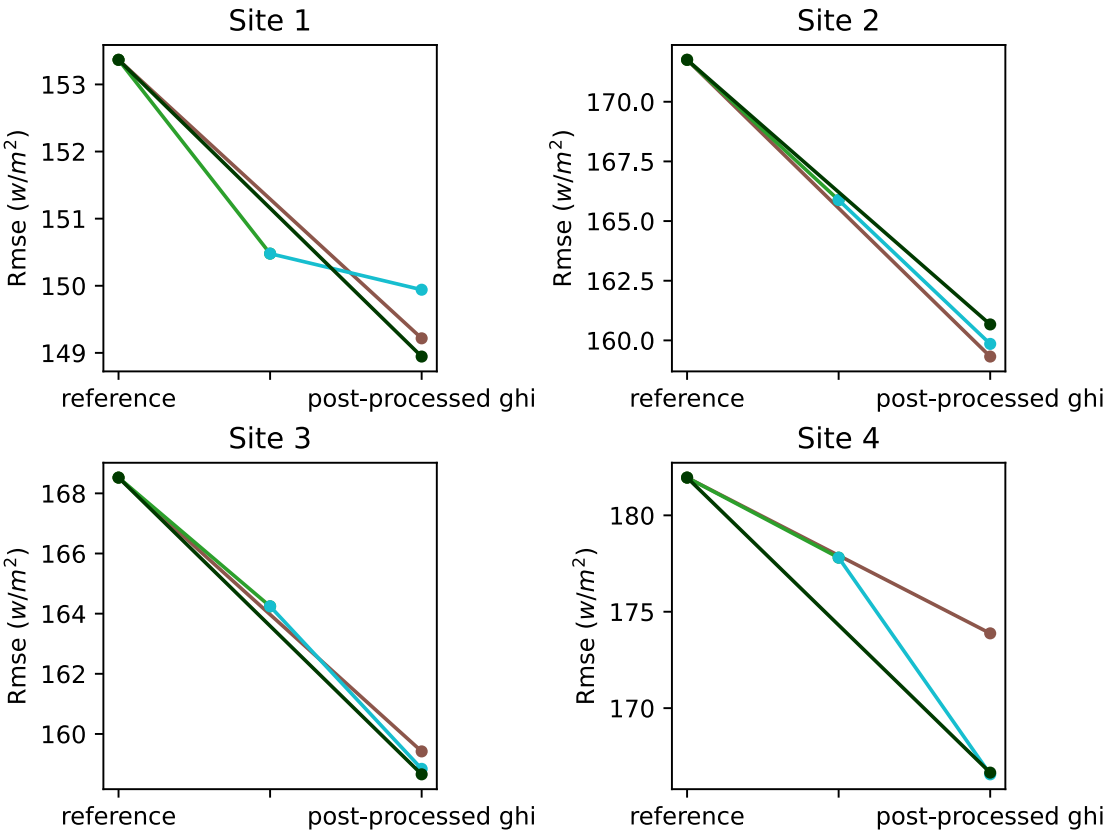


Figure 26: RMSE evolution of the relevant models.

Figure 27: The 25 German sites used for validation.

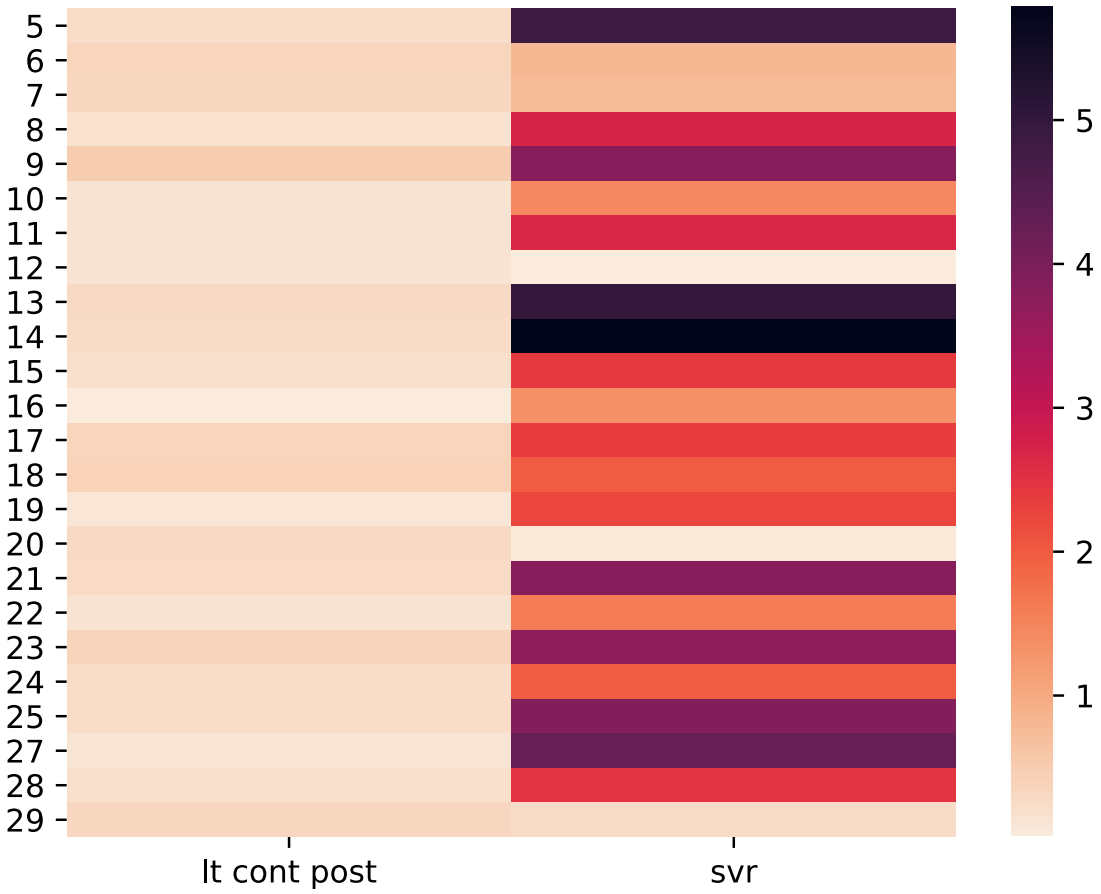


Figure 28: MAE skill score heatmap of the validation sites (reference: LT CONT), skill scores are in percent.

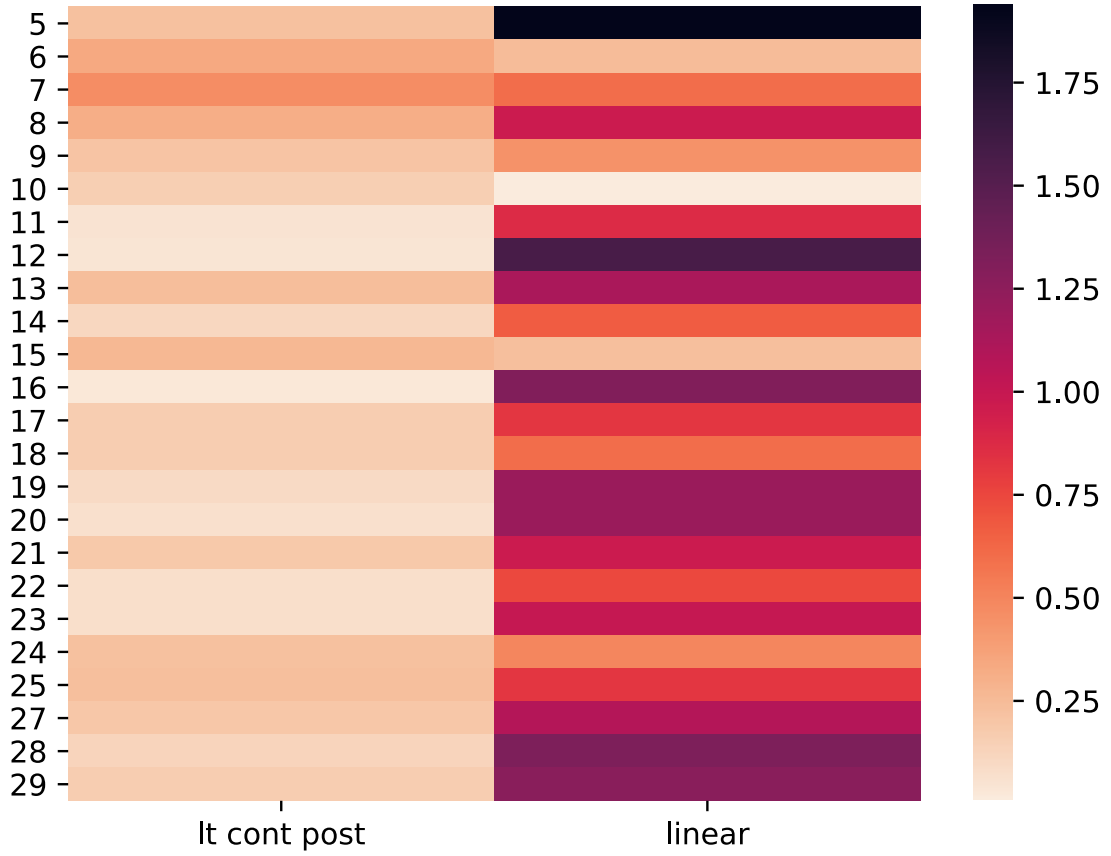


Figure 29: RMSE skill score heatmap of the validation sites (reference: LT CONT), skill scores are in percent.

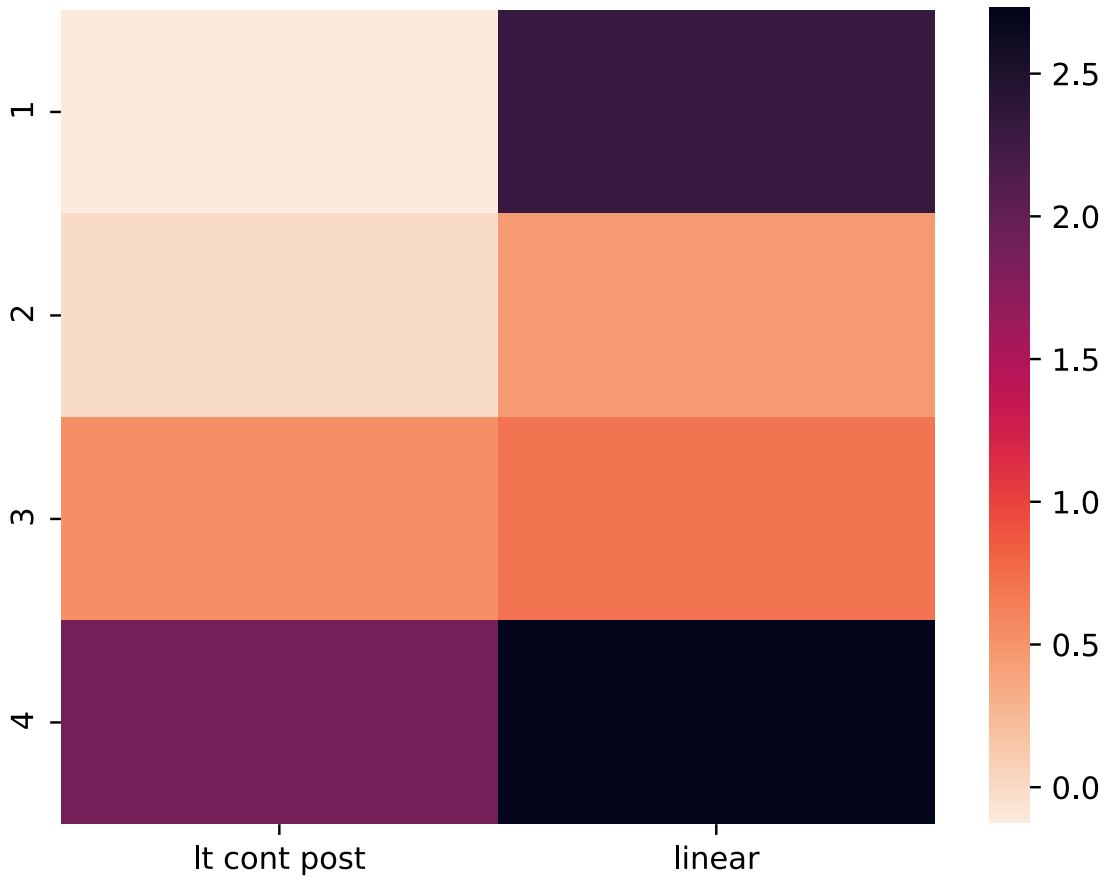
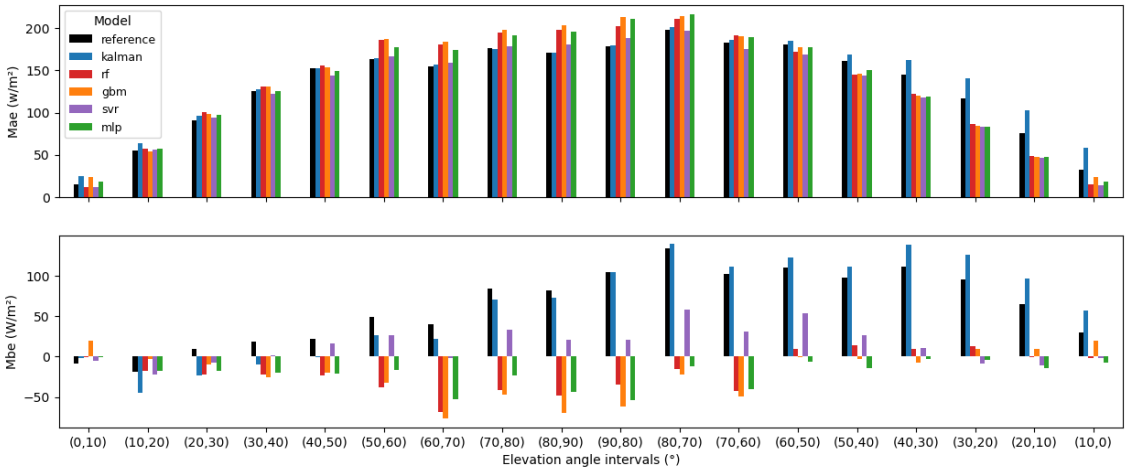
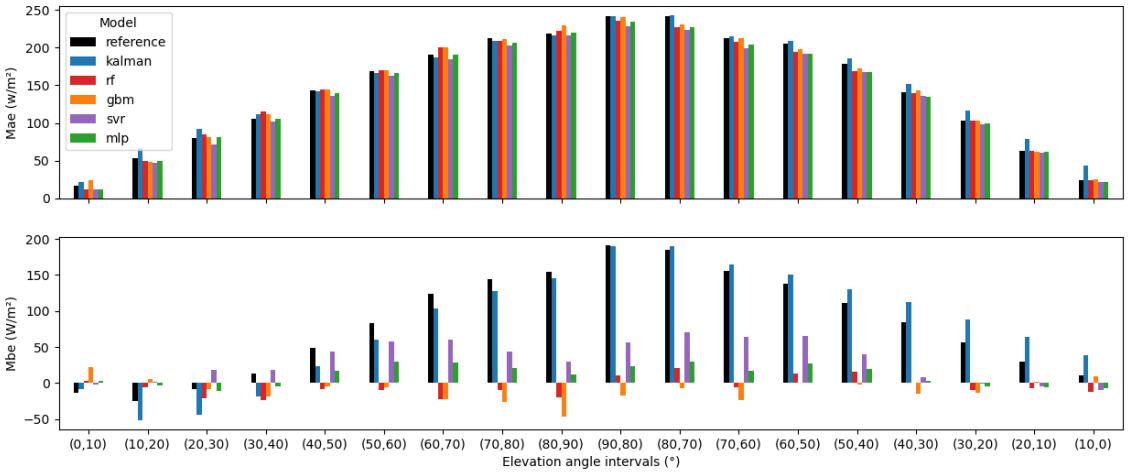


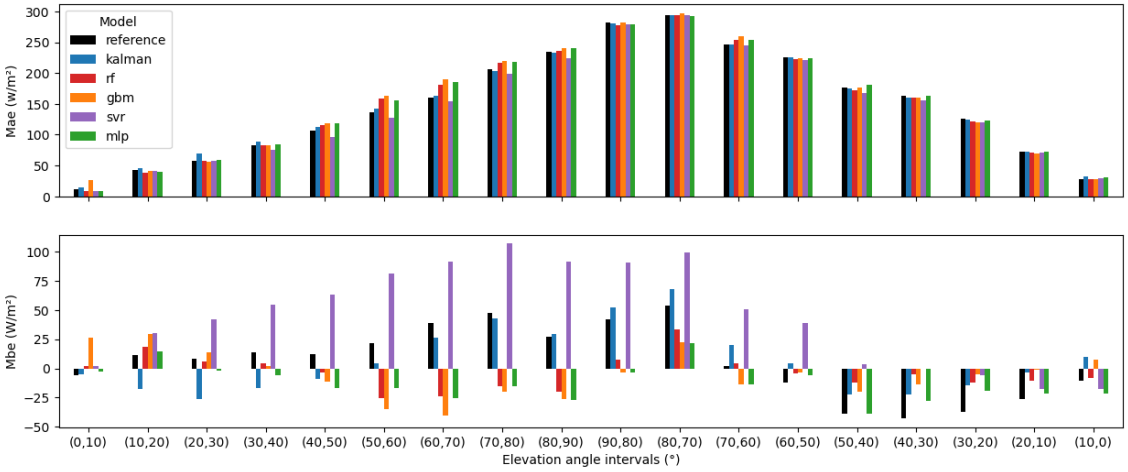
Figure 30: Global linear model verification on the 4 initial sites (reference: LT CONT), skill scores are in percent.



(a) Site 1

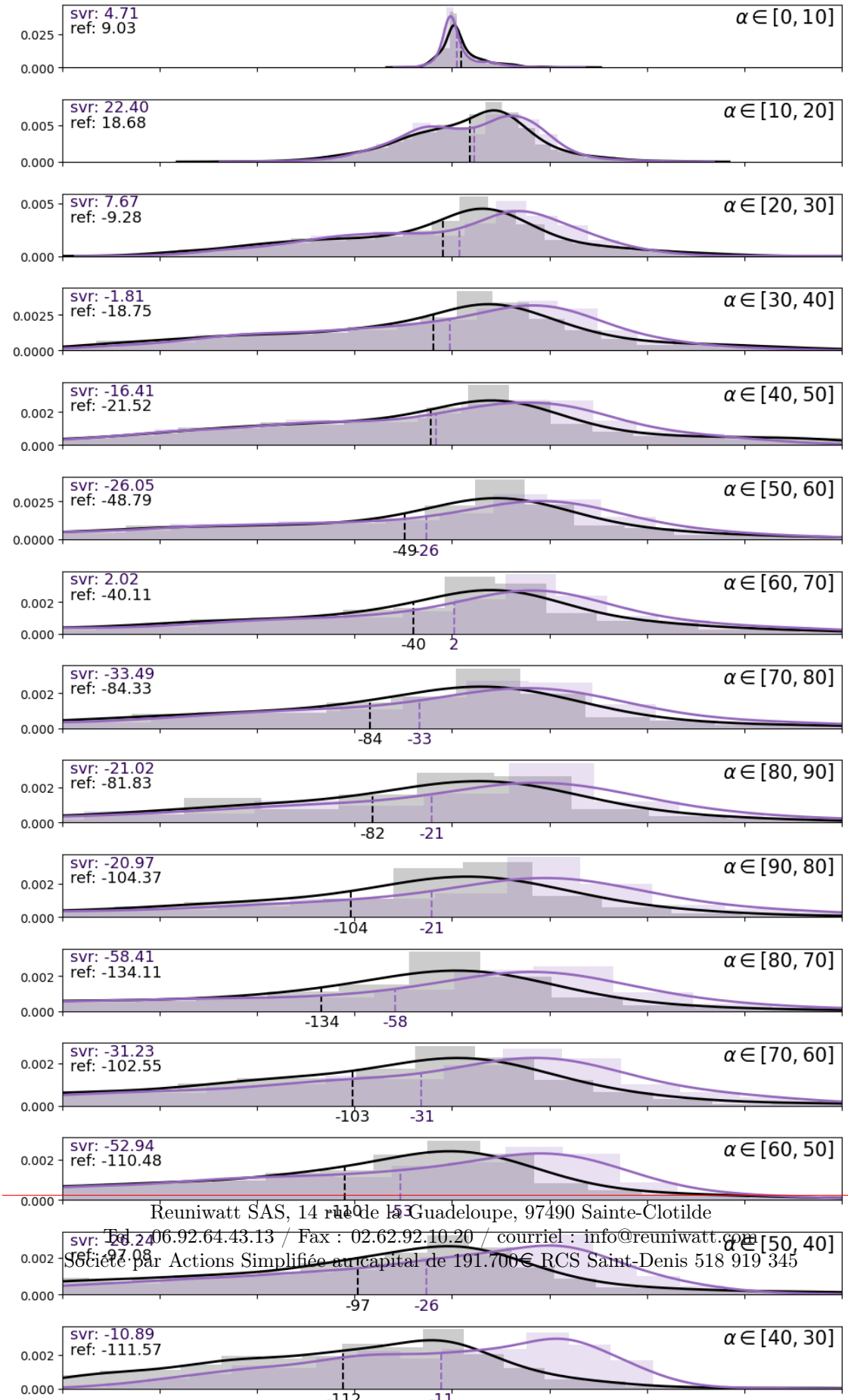


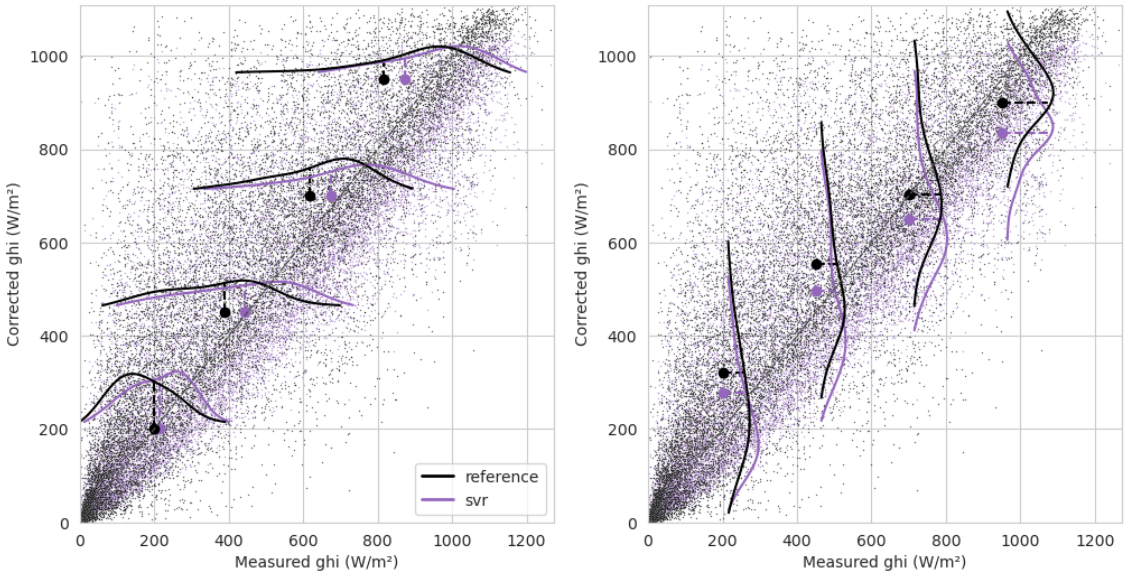
(b) Site 3



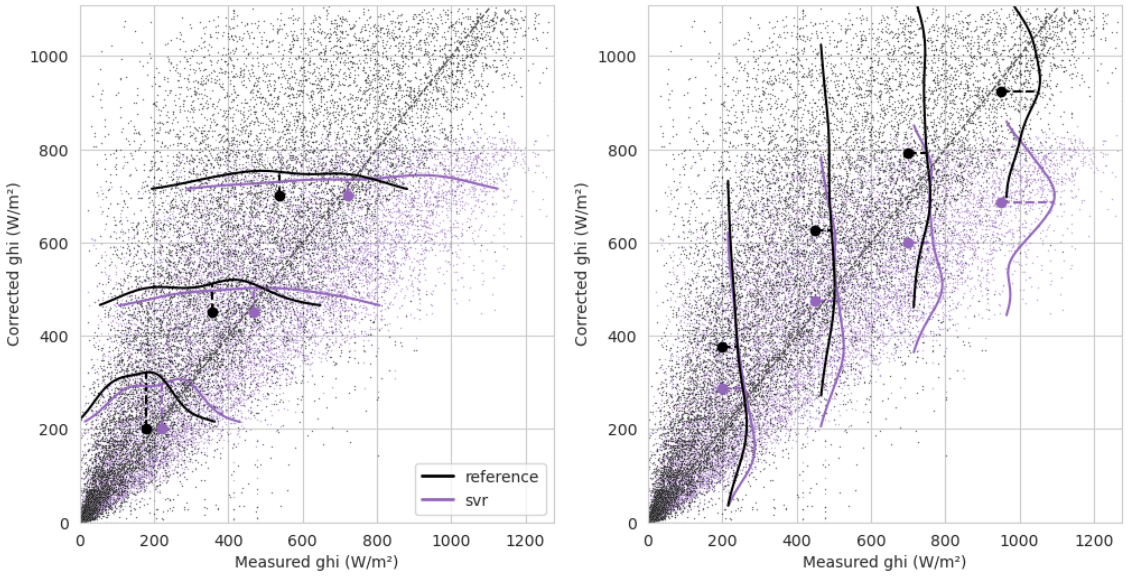
(c) Site 4

Figure 31: MAE and MBE levels across all elevation angle intervals of a day.

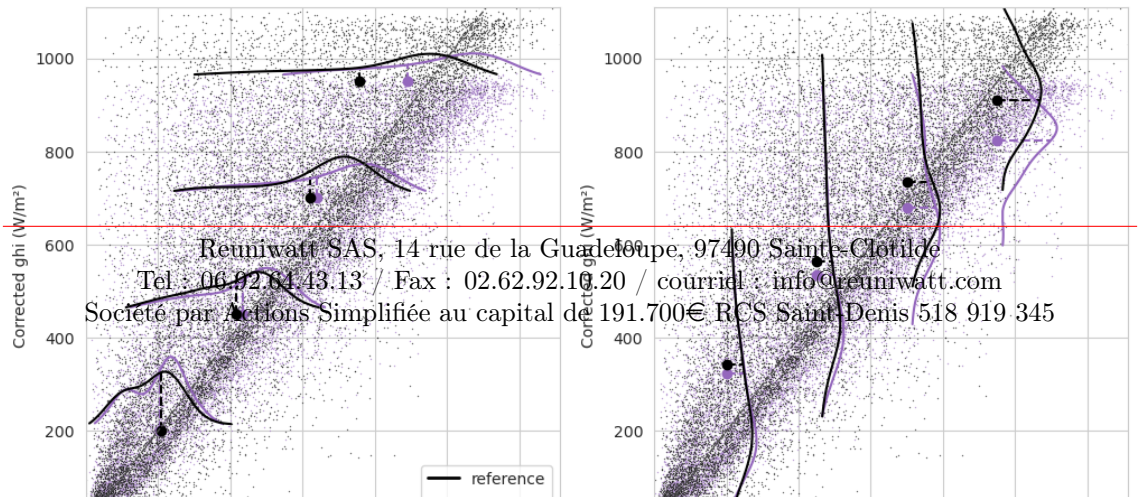

svr: 4.71
ref: 9.03
α∈[0, 10]
svr: 22.40
ref: 18.68
α∈[10, 20]
svr: 7.67
ref: -9.28
α∈[20, 30]
svr: -1.81
ref: -18.75
α∈[30, 40]
svr: -16.41
ref: -21.52
α∈[40, 50]
svr: -26.05
ref: -48.79
-49
-26
α∈[50, 60]
svr: 2.02
ref: -40.11
-40
2
α∈[60, 70]
svr: -33.49
ref: -84.33
-84
-33
α∈[70, 80]
svr: -21.02
ref: -81.83
-82
-21
α∈[80, 90]
svr: -20.97
ref: -104.37
-104
-21
α∈[90, 80]
svr: -58.41
ref: -134.11
-134
-58
α∈[80, 70]
svr: -31.23
ref: -102.55
-103
-31
α∈[70, 60]
svr: -52.94
ref: -110.48
α∈[60, 50]
svr: -26.24
ref: -97.08
-97
-26
α∈[50, 40]
svr: -10.89
ref: -111.57
α∈[40, 30]

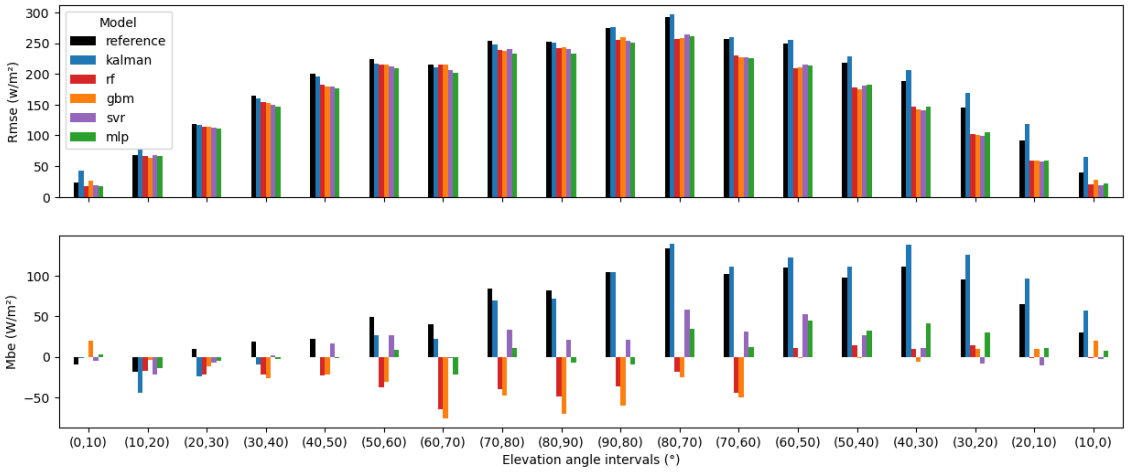


(a) Site 1

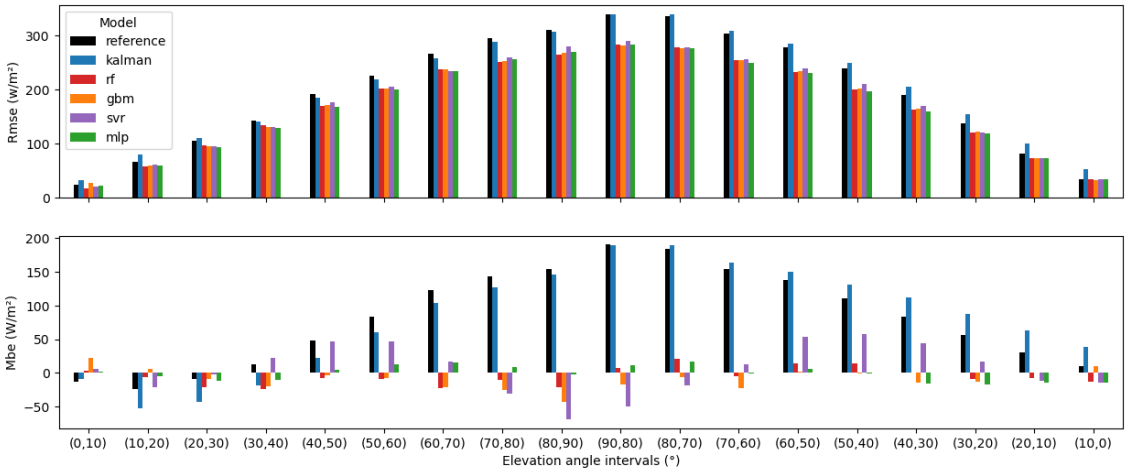


(b) Site 2

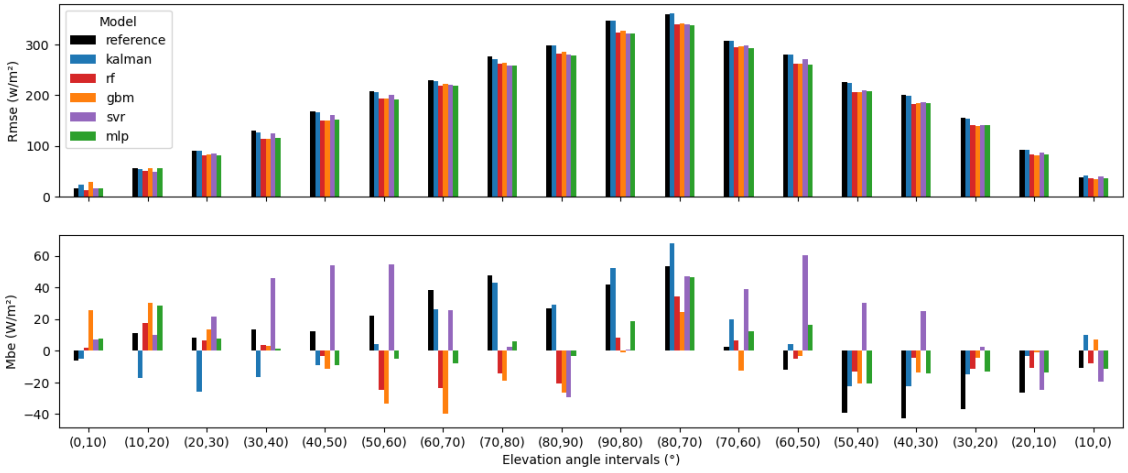




(a) Site 1

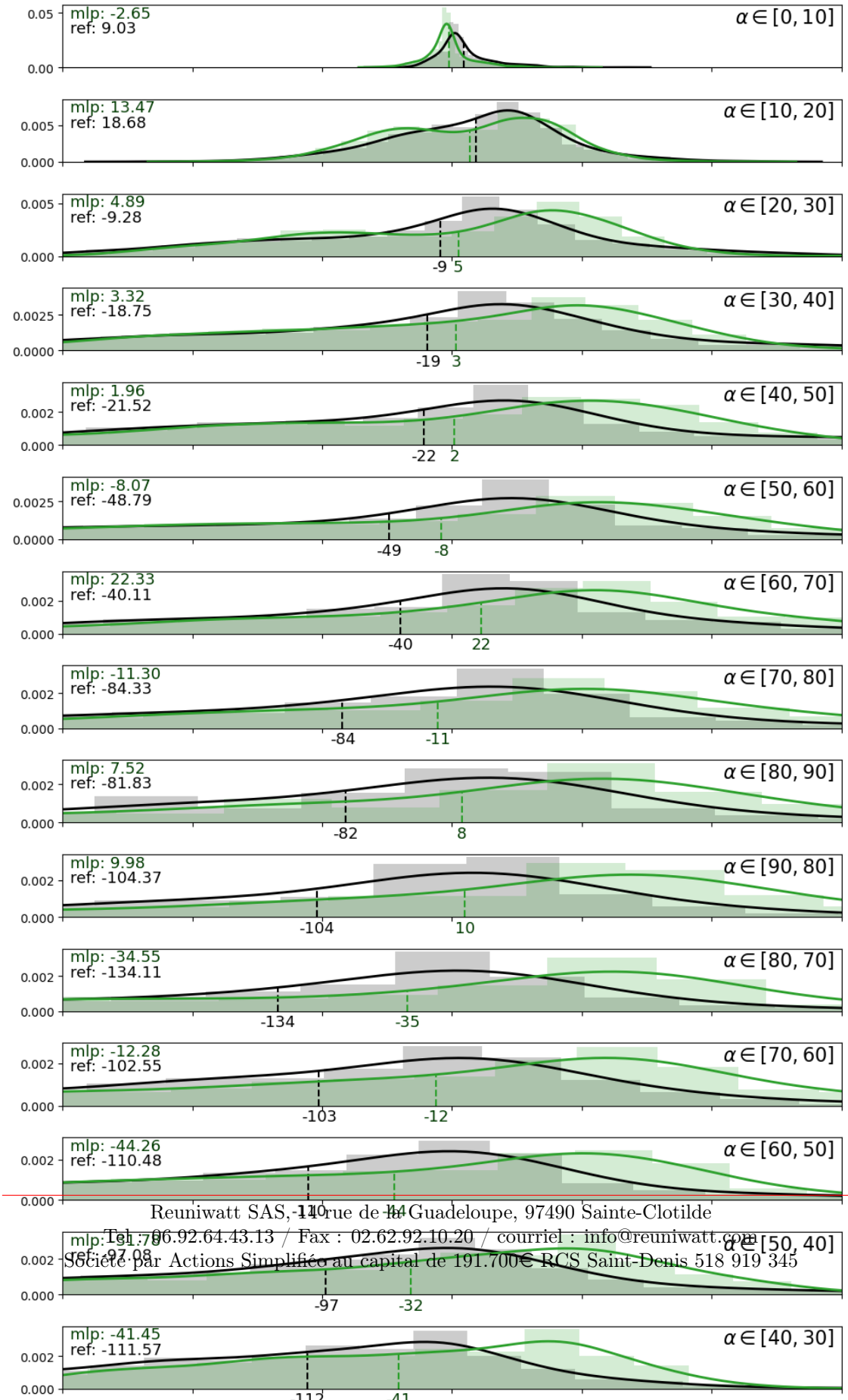


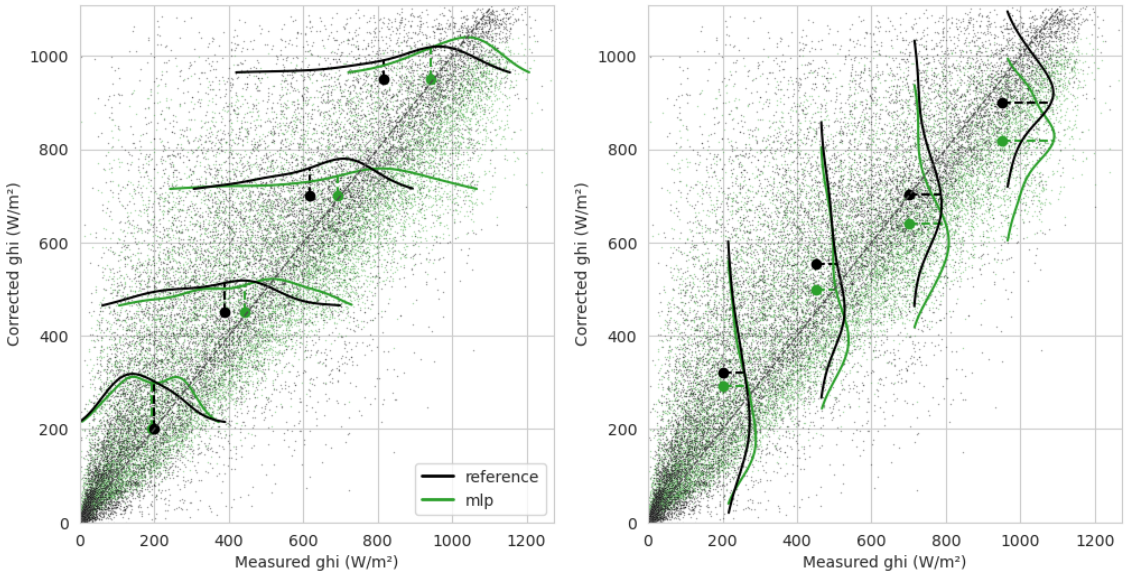
(b) Site 3



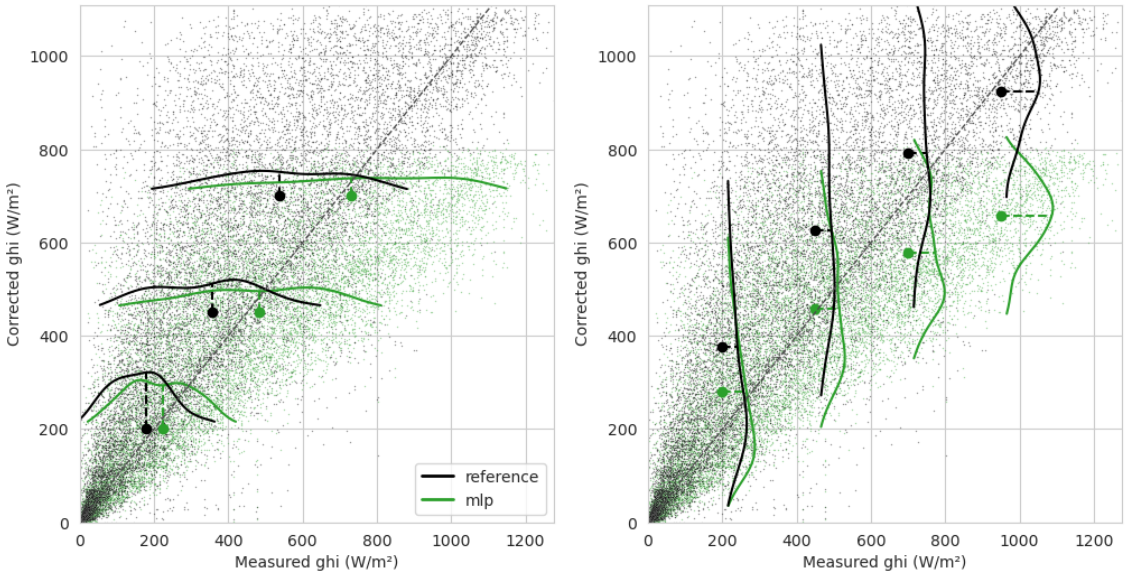
(c) Site 4

Figure 54: RMSE and MBE levels across all elevation angle intervals of a day.

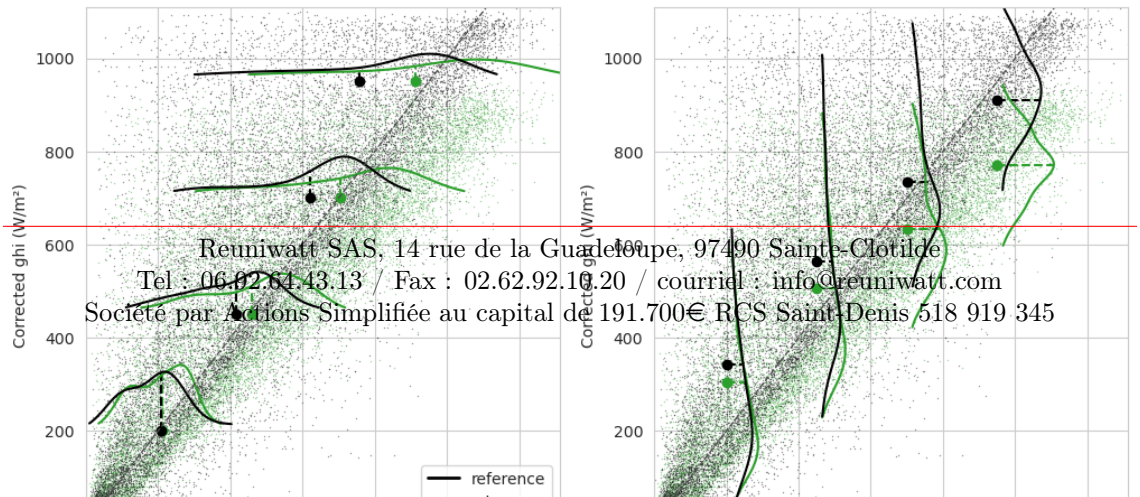

mlp: -2.65
ref: 9.03
α∈[0, 10]
mlp: 13.47
ref: 18.68
α∈[10, 20]
mlp: 4.89
ref: -9.28
-9
5
α∈[20, 30]
mlp: 3.32
ref: -18.75
-19
3
α∈[30, 40]
mlp: 1.96
ref: -21.52
-22
2
α∈[40, 50]
mlp: -8.07
ref: -48.79
-49
-8
α∈[50, 60]
mlp: 22.33
ref: -40.11
-40
22
α∈[60, 70]
mlp: -11.30
ref: -84.33
-84
-11
α∈[70, 80]
mlp: 7.52
ref: -81.83
-82
8
α∈[80, 90]
mlp: 9.98
ref: -104.37
-104
10
α∈[90, 80]
mlp: -34.55
ref: -134.11
-134
-35
α∈[80, 70]
mlp: -12.28
ref: -102.55
-103
-12
α∈[70, 60]
mlp: -44.26
ref: -110.48
α∈[60, 50]
mlp: -31.78
ref: -97.08
-97
-32
α∈[50, 40]
mlp: -41.45
ref: -111.57
α∈[40, 30]



(a) Site 1



(b) Site 2



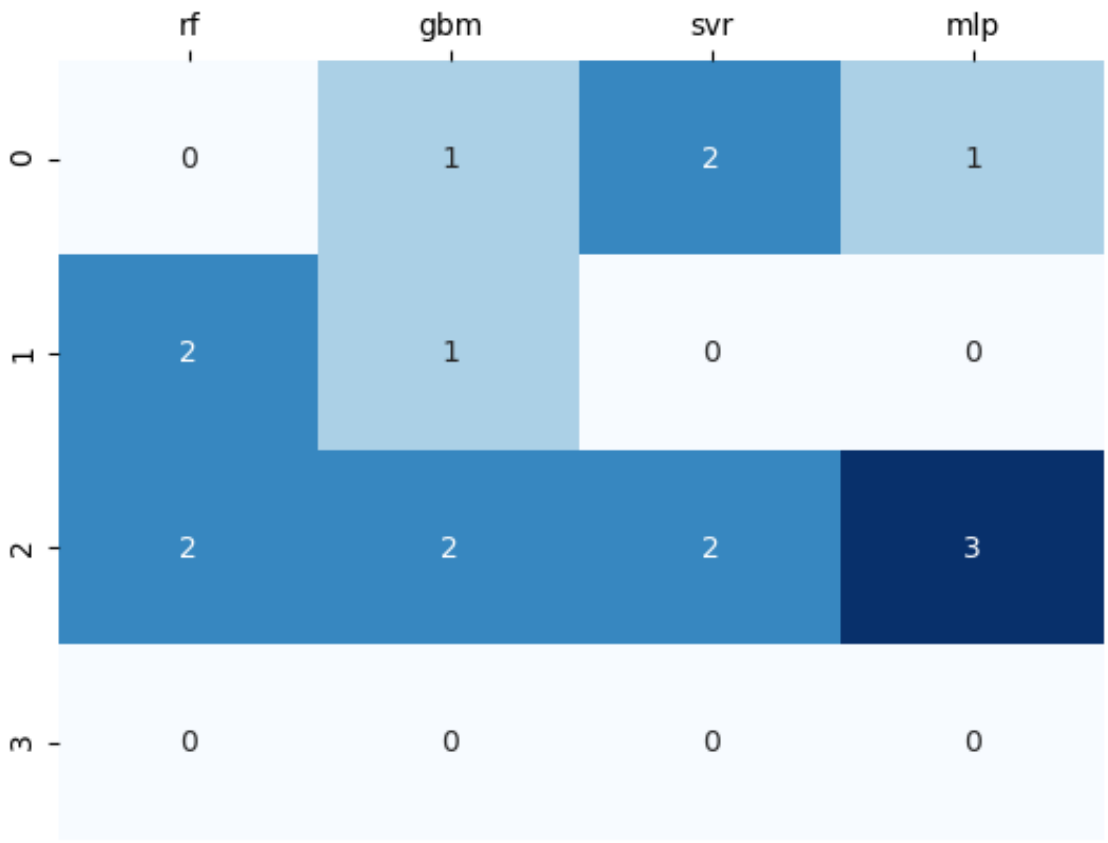


Figure 37: Pairwise systematicity matrix for RMSE. The value $V_{i,j}$ of the cell (i, j) indicates how often the configuration of line i is the best one, across the 4 sites, for the model of column j. For example, the configuration 0 is the best one with a GBM post-processing for 3 sites, and the configuration 2 is the best one for 1 site.

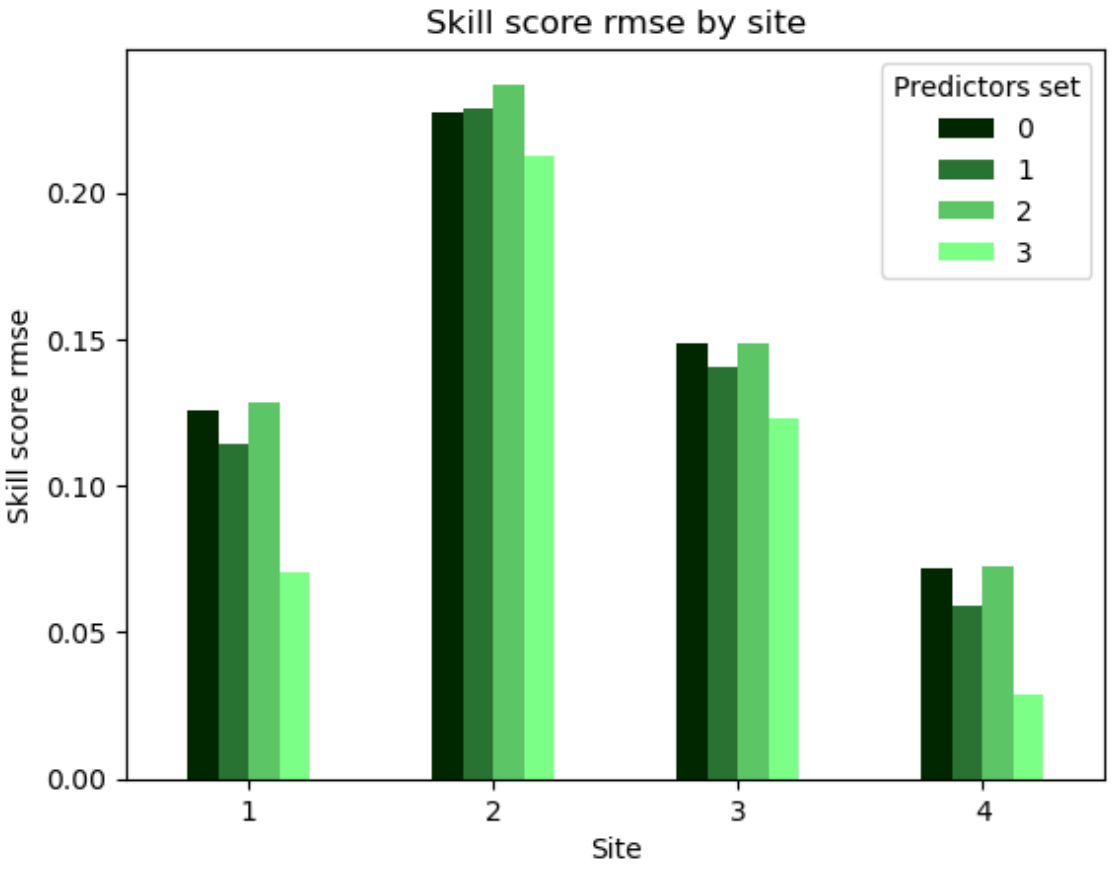


Figure 38: Comparison of the RMSE skill scores of the different configurations.

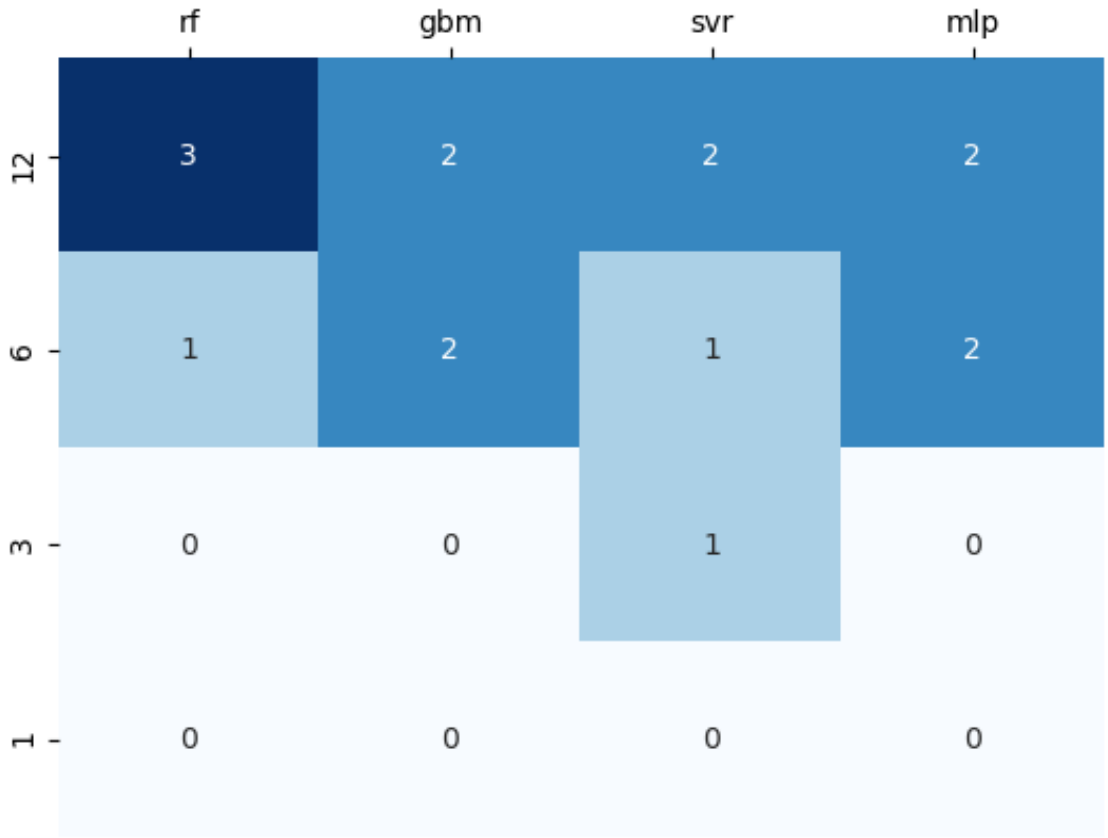


Figure 39: Pairwise systematicity matrix for RMSE. The value $V_{i,j}$ of the cell (i, j) indicates how often the learning period duration (in months) of line i performs the best, across the 4 sites, for the model of column j. For example, having a 12-months-long learning period is the best thing in 3 sites out of 4 with a SVR model, the last site performs better with a 6-month-long learning period.

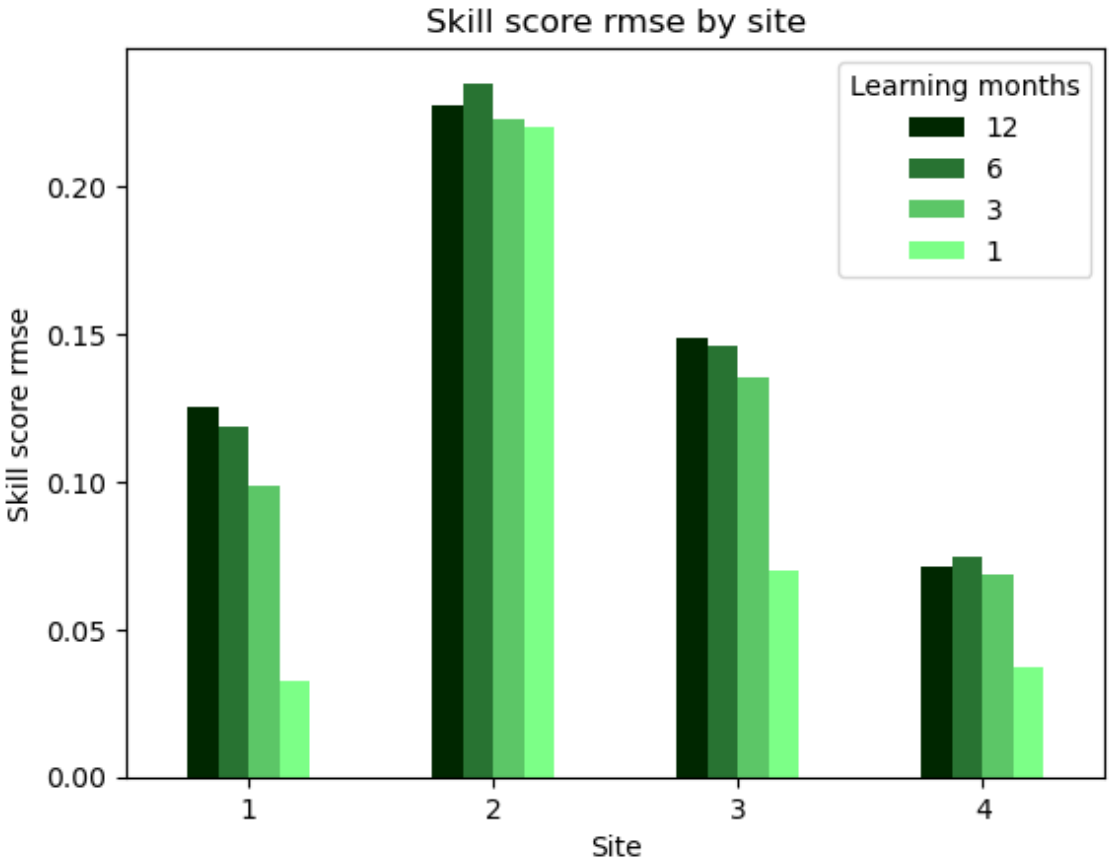


Figure 40: Comparison of the RMSE skill scores of the different learning period durations (in months).

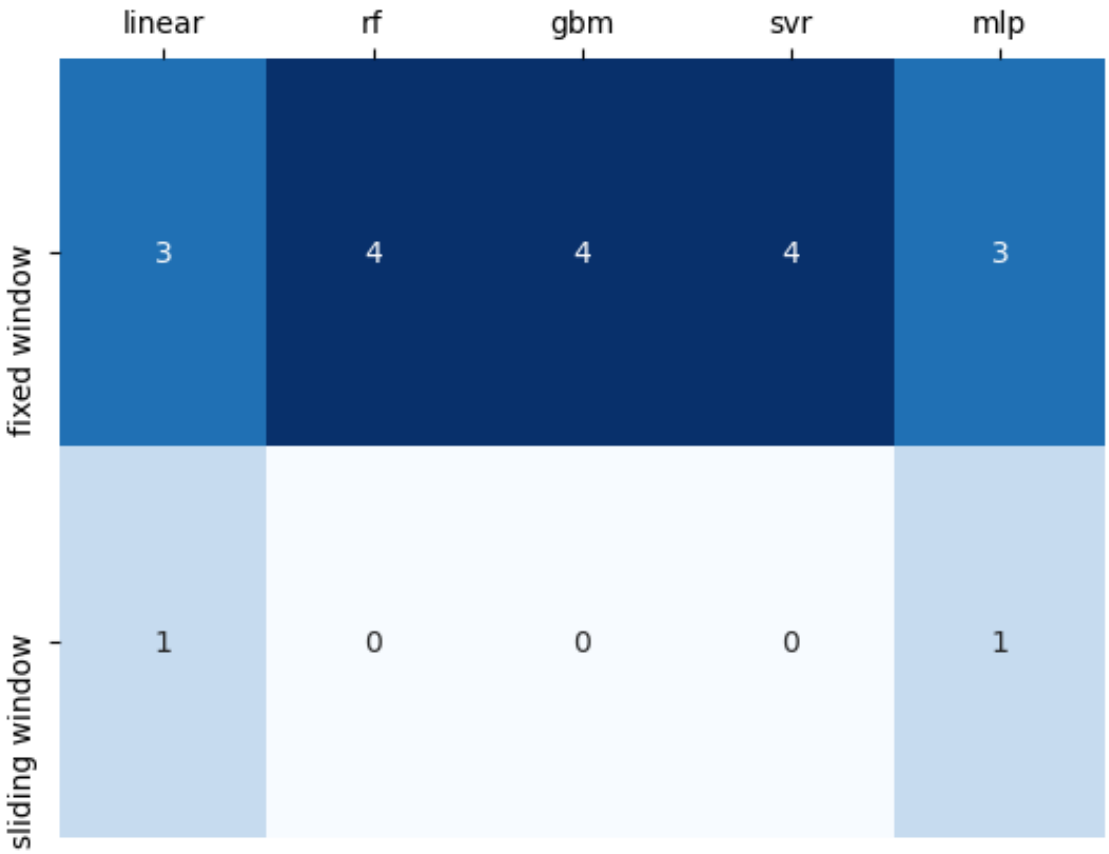


Figure 41: Pairwise systematicity matrix concerning window type for RMSE.

Figure 42: Comparison of the RMSE skill scores for a SVR model.

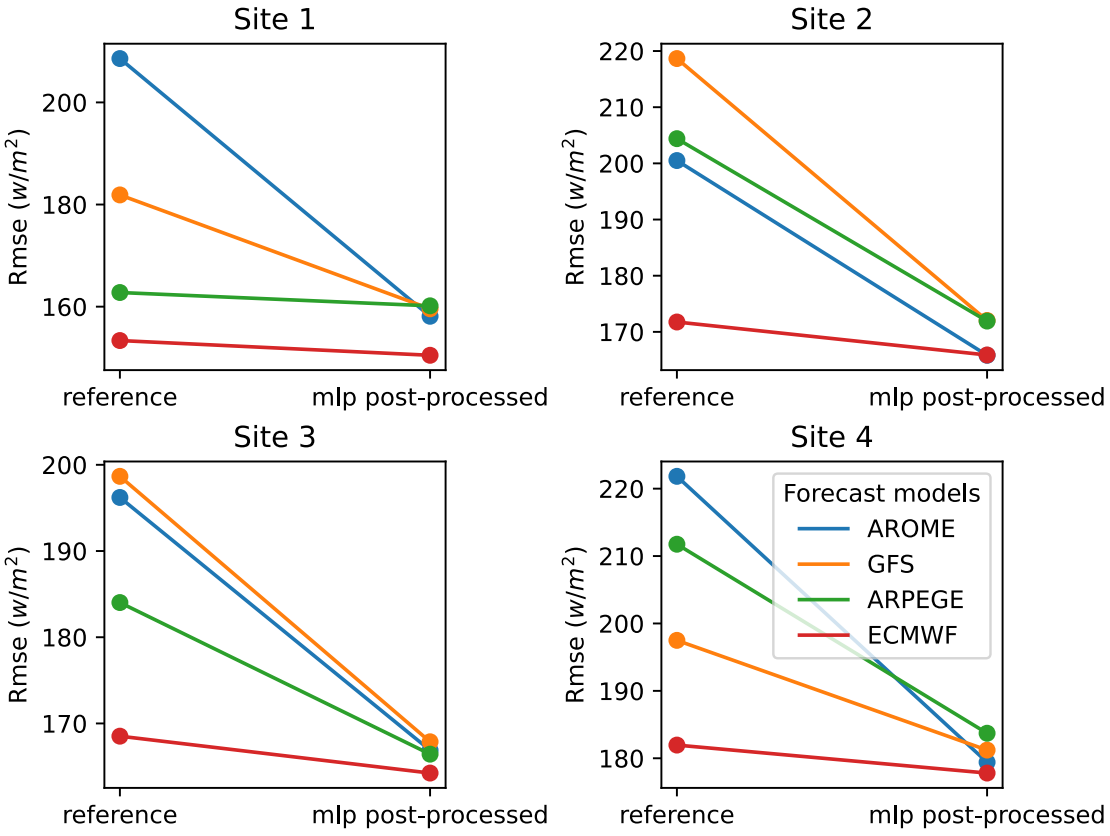


Figure 43: Comparison of the post-processing of four different NWP forecast models on RMSE.