Internship report

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):

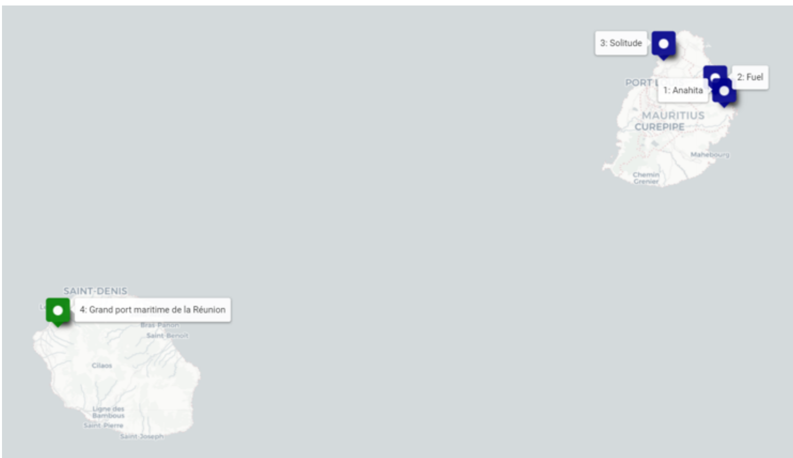


Figure 1: Four initial study sites

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

	reference	kalman	rf	gbm	svr	mlp	Score
reference	NA	4	1	2	0	1	8.0
kalman	0	NA	0	1	0	1	2.0
rf	3	4	NA	4	0	2	13.0
gbm	2	3	0	NA	0	0	5.0
svr	4	4	4	4	NA	4	20.0
mlp	3	3	2	4	0	NA	12.0

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.

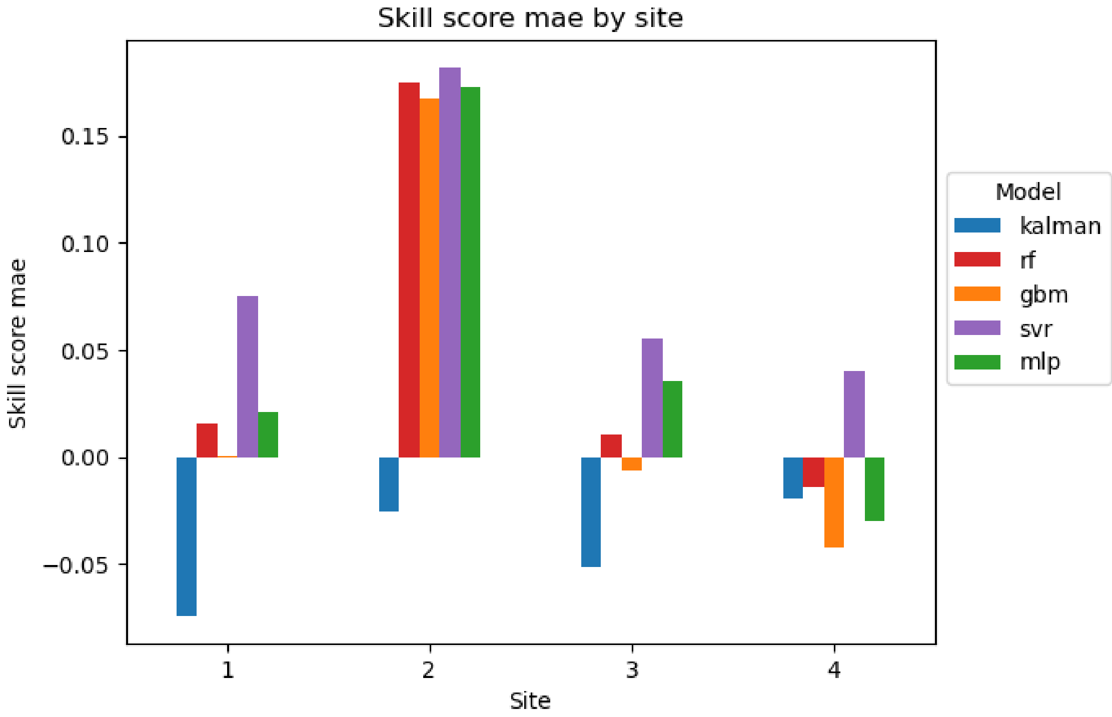


Figure 3: MAE skill score plot across the 4 sites.

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

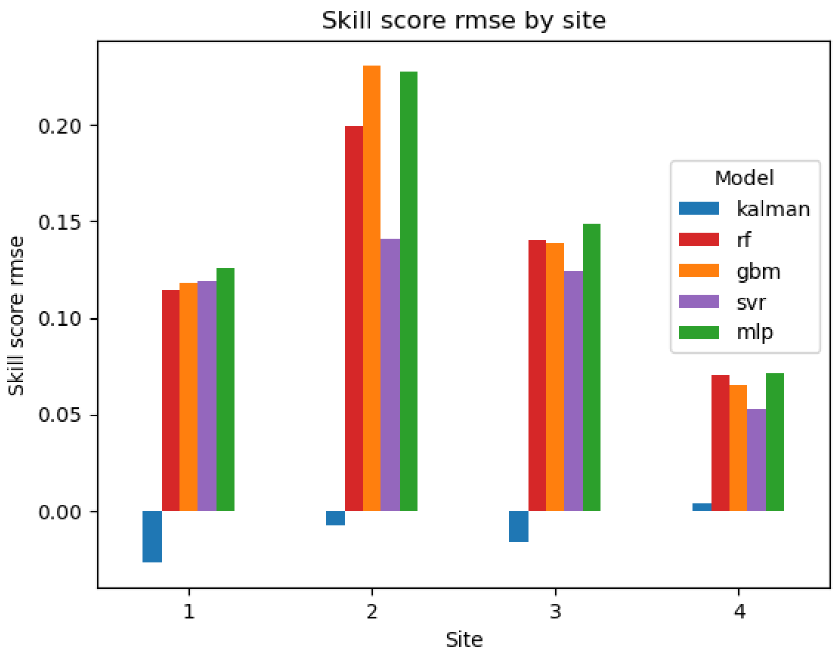


Figure 5: RMSE skill score plot across the 4 sites.

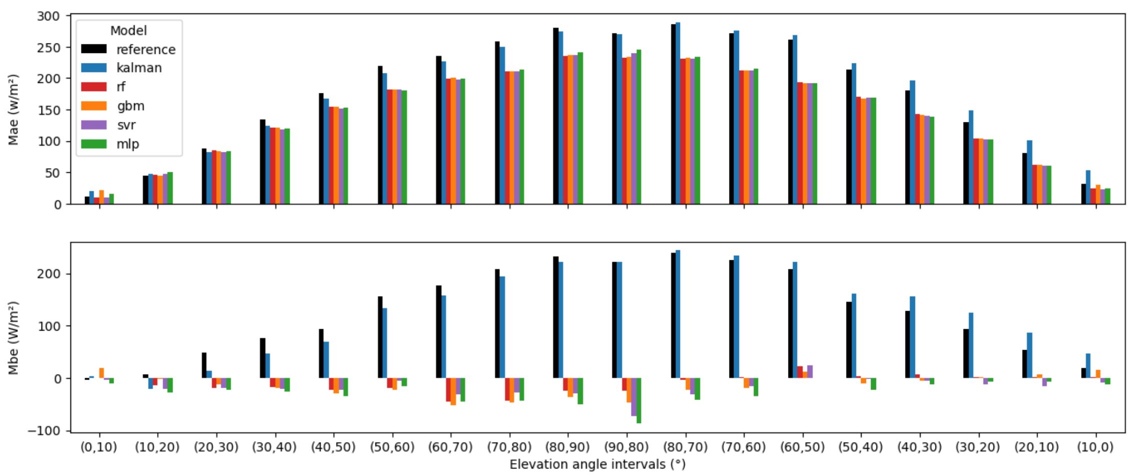


Figure 6: MAE and MBE levels across all elevation angle intervals of a day, for site 2.

RMSE

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 the appendix.

	0	1	2	3
ghi_{GFS}	X	X	X	X
$temperature_{GFS}^{2m}$	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.

Influence of the choice of the predictors

Influence of the learning period

Influence of the window of learning

Influence of the NWP forecasting model

3.2 Benchmarking the linear regression models

MAE

RMSE

3.3 Showcase of the hybrid model

3.3.1 Study on the four inital sites

MAE

RMSE

3.3.2 Study on the German sites

MAE

RMSE

4 Conclusions and perspectives

4.1 Results summary

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.

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/`.

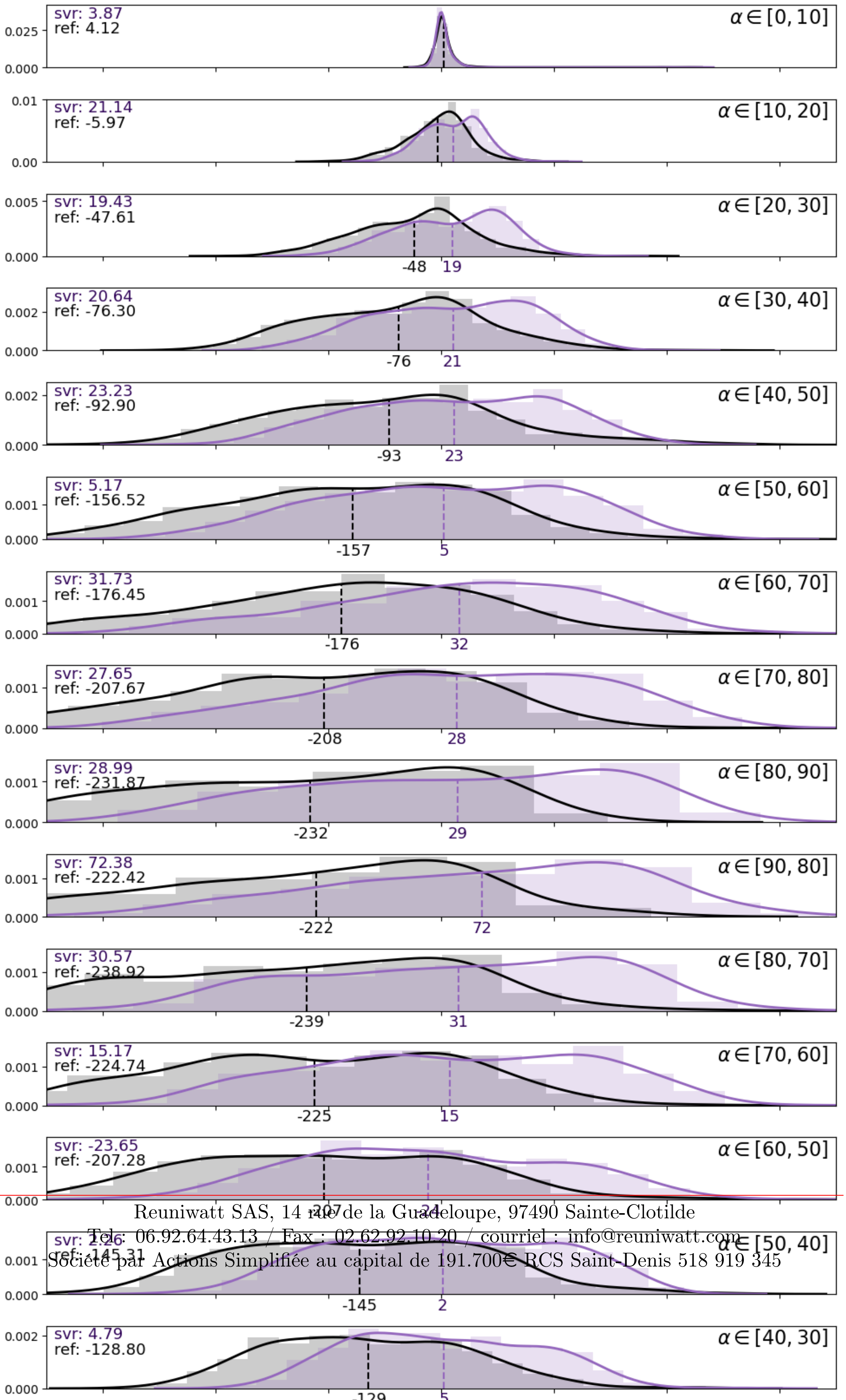
H. Verbois, Y.-M. Saint-Drenan, A. Thiery, and P. Blanc. Statistical learning for NWP post-processing: A benchmark for solar irradiance forecasting. 238:132–149. ISSN 0038092X. doi: 10.1016/j.solener.2022.03.017. URL `https://linkinghub.elsevier.com/retrieve/pii/S0038092X22001839`.

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
α ∈ [20, 30]
-48
19
svr: 20.64
ref: -76.30
α ∈ [30, 40]
-76
21
svr: 23.23
ref: -92.90
α ∈ [40, 50]
-93
23
svr: 5.17
ref: -156.52
α ∈ [50, 60]
-157
5
svr: 31.73
ref: -176.45
α ∈ [60, 70]
-176
32
svr: 27.65
ref: -207.67
α ∈ [70, 80]
-208
28
svr: 28.99
ref: -231.87
α ∈ [80, 90]
-232
29
svr: 72.38
ref: -222.42
α ∈ [90, 80]
-222
72
svr: 30.57
ref: -238.92
α ∈ [80, 70]
-239
31
svr: 15.17
ref: -224.74
α ∈ [70, 60]
-225
15
svr: -23.65
ref: -207.28
α ∈ [60, 50]
-207
-24
svr: 2.26
ref: -145.31
α ∈ [50, 40]
-145
2
svr: 4.79
ref: -128.80
α ∈ [40, 30]
-129
5

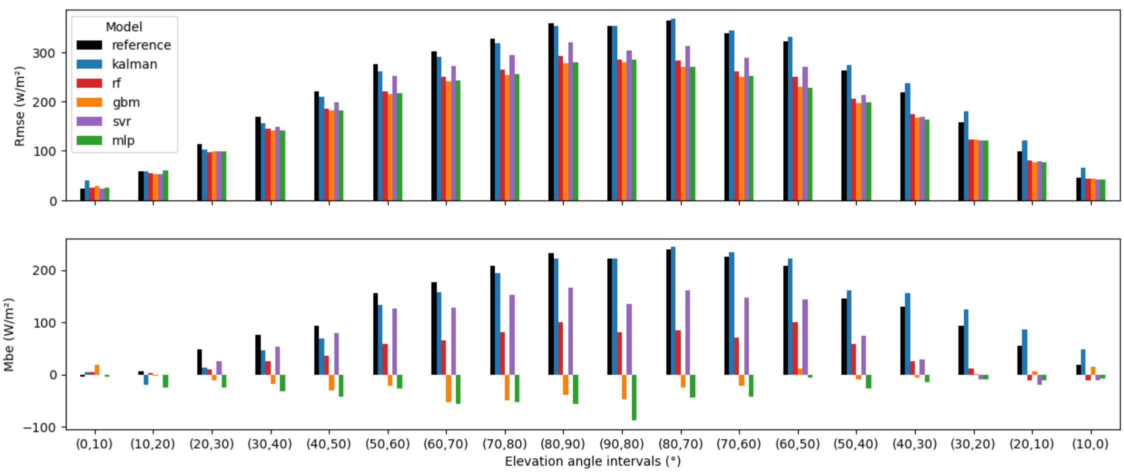
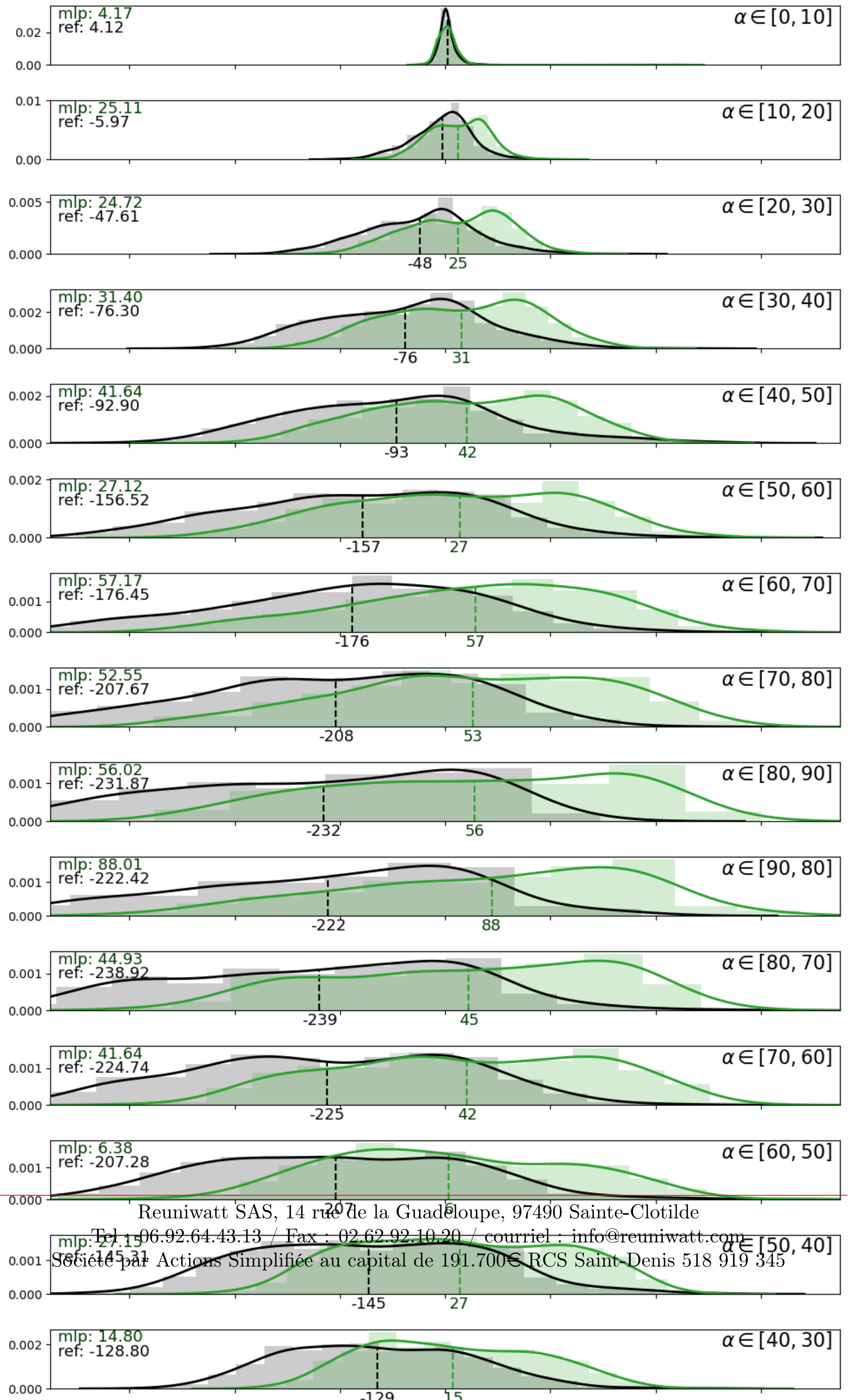


Figure 8: RMSE and MBE levels across all elevation angle intervals of a day, for site 2.


mlp: 4.17
ref: 4.12
$\alpha \in [0, 10]$
mlp: 25.11
ref: -5.97
$\alpha \in [10, 20]$
mlp: 24.72
ref: -47.61
$\alpha \in [20, 30]$
-48
25
mlp: 31.40
ref: -76.30
$\alpha \in [30, 40]$
-76
31
mlp: 41.64
ref: -92.90
$\alpha \in [40, 50]$
-93
42
mlp: 27.12
ref: -156.52
$\alpha \in [50, 60]$
-157
27
mlp: 57.17
ref: -176.45
$\alpha \in [60, 70]$
-176
57
mlp: 52.55
ref: -207.67
$\alpha \in [70, 80]$
-208
53
mlp: 56.02
ref: -231.87
$\alpha \in [80, 90]$
-232
56
mlp: 88.01
ref: -222.42
$\alpha \in [90, 80]$
-222
88
mlp: 44.93
ref: -238.92
$\alpha \in [80, 70]$
-239
45
mlp: 41.64
ref: -224.74
$\alpha \in [70, 60]$
-225
42
mlp: 6.38
ref: -207.28
$\alpha \in [60, 50]$
-207
6
mlp: 27.15
ref: -145.31
$\alpha \in [50, 40]$
-145
27
mlp: 14.80
ref: -128.80
$\alpha \in [40, 30]$
-129
15

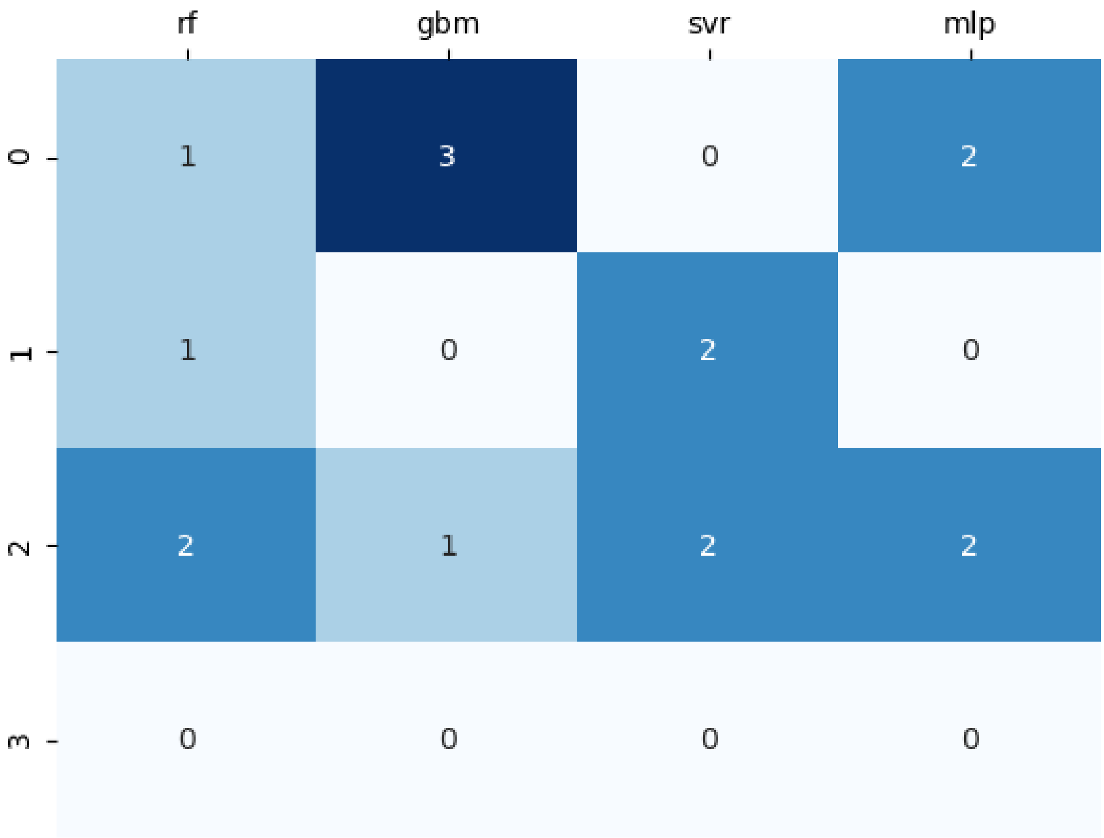


Figure 10: Pairwise systematicity matrix for MAE. The value $V_{i,j}$ of the cell (i, j) indicates how often the configuration of line i is the best one, across the 4 sites, for the model of column j. For example, the configuration 0 is the best one with a GBM post-processing for 3 sites, and the configuration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

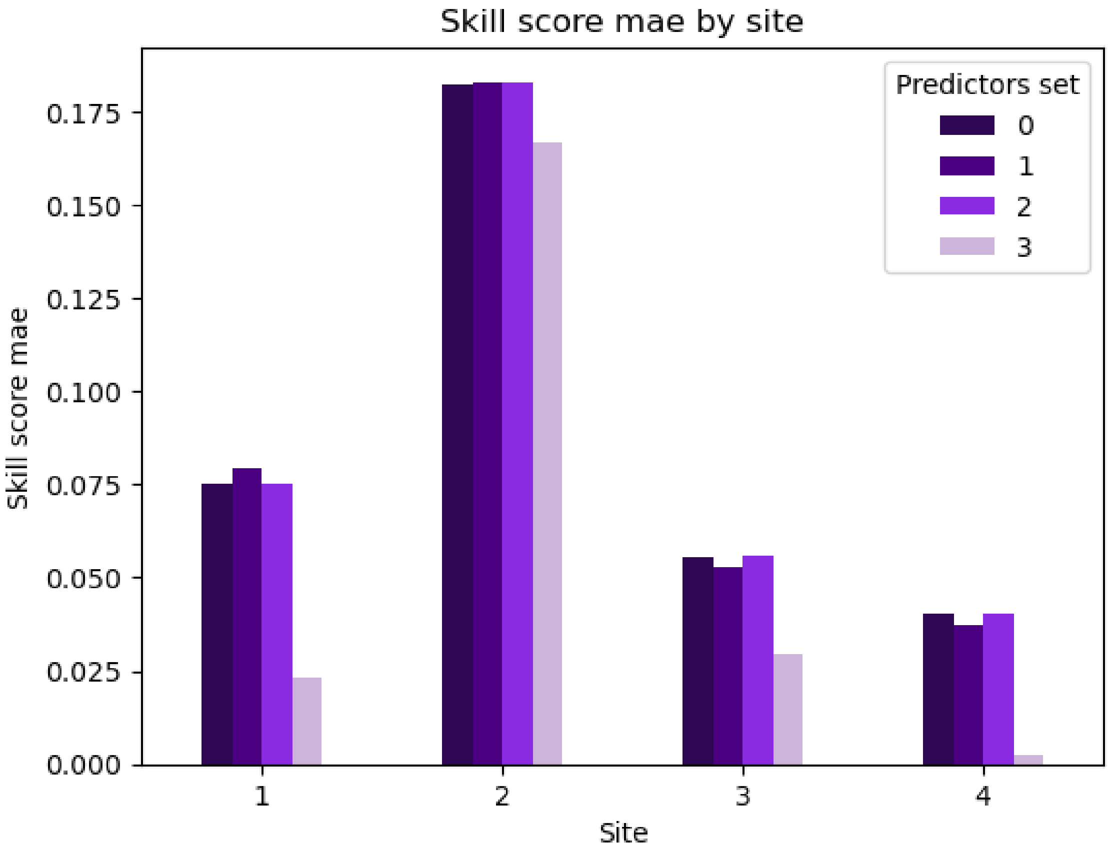


Figure 11: Comparison of the MAE skill scores of the different configurations.

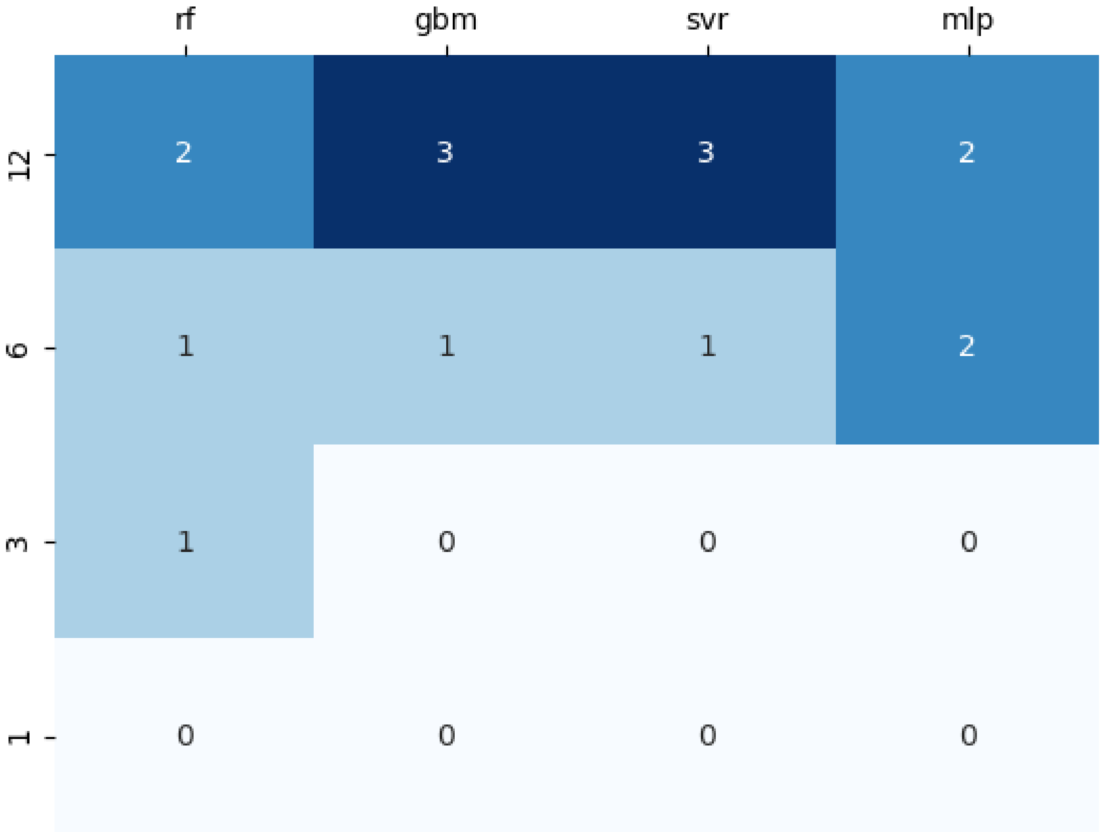


Figure 12: Pairwise systematicity matrix for MAE. The value $V_{i,j}$ of the cell (i, j) indicates how often the learning period duration (in months) of line i performs the best, across the 4 sites, for the model of column j. For example, having a 12-months-long learning period is the best thing in 3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

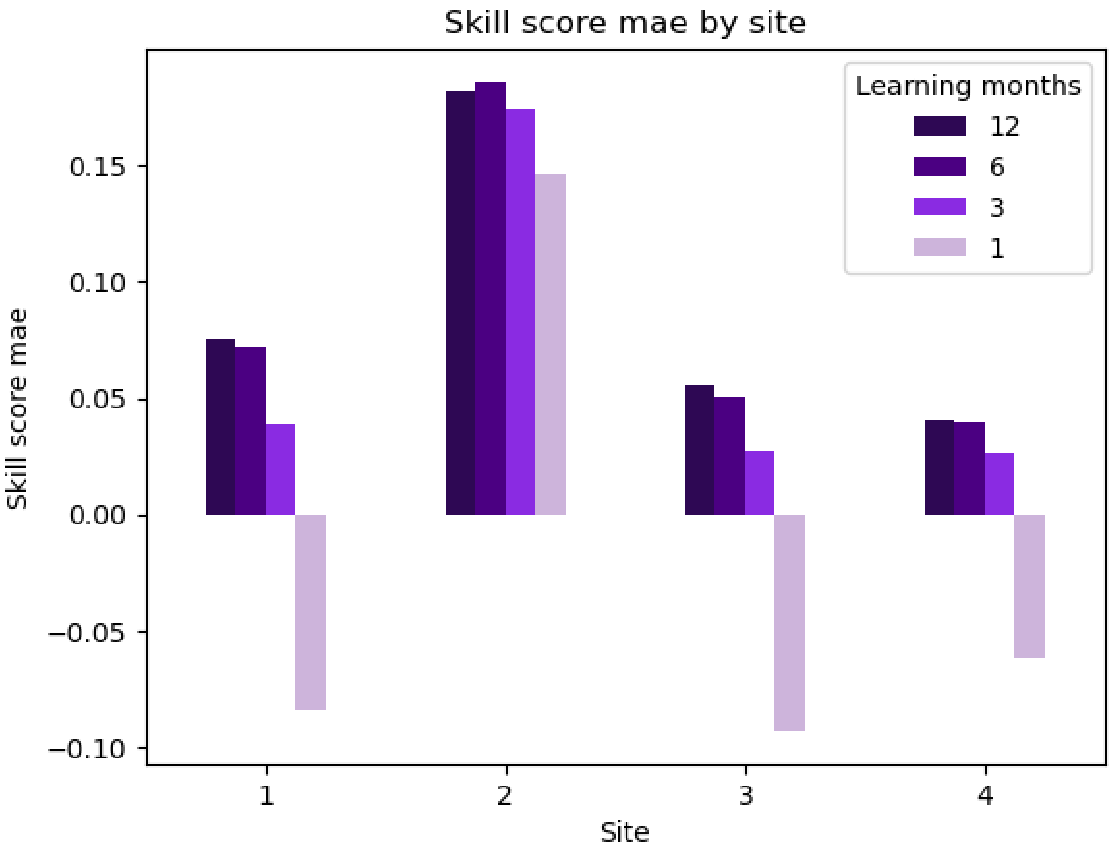


Figure 13: Comparison of the MAE skill scores of the different learning period durations (in months).

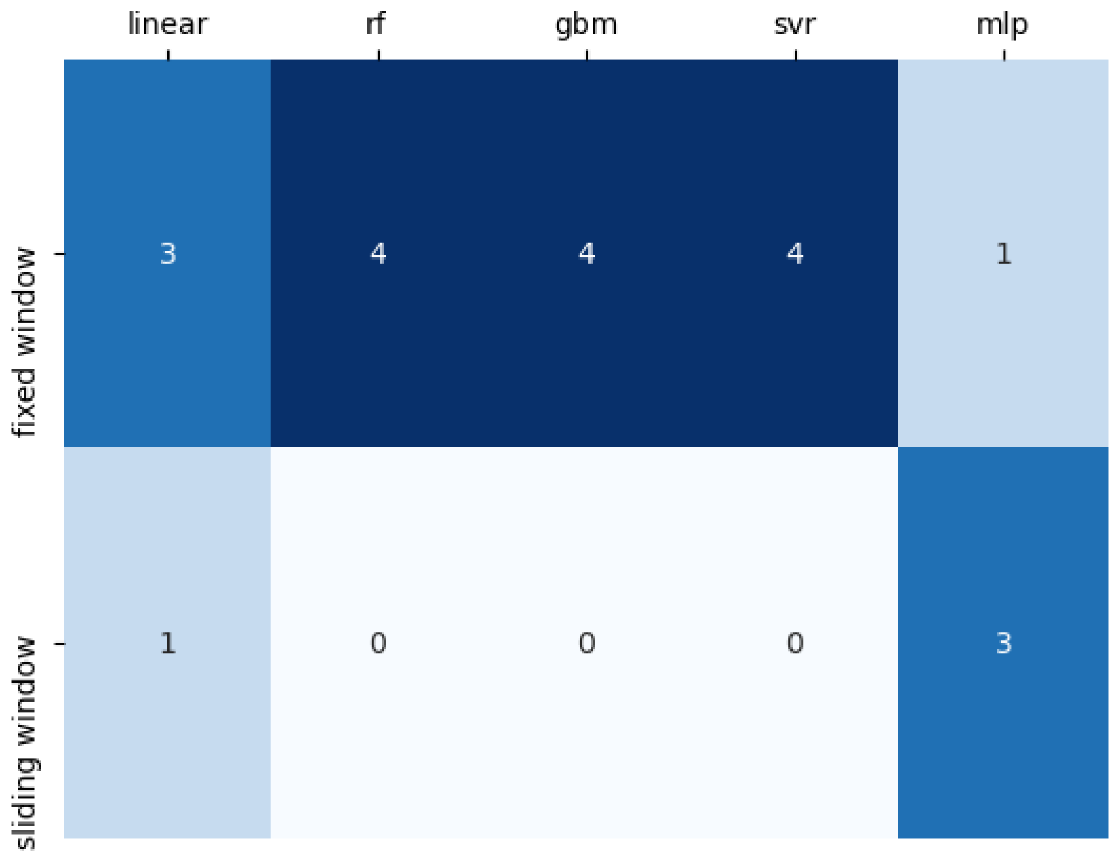


Figure 14: Pairwise systematicity matrix concerning window type for MAE.

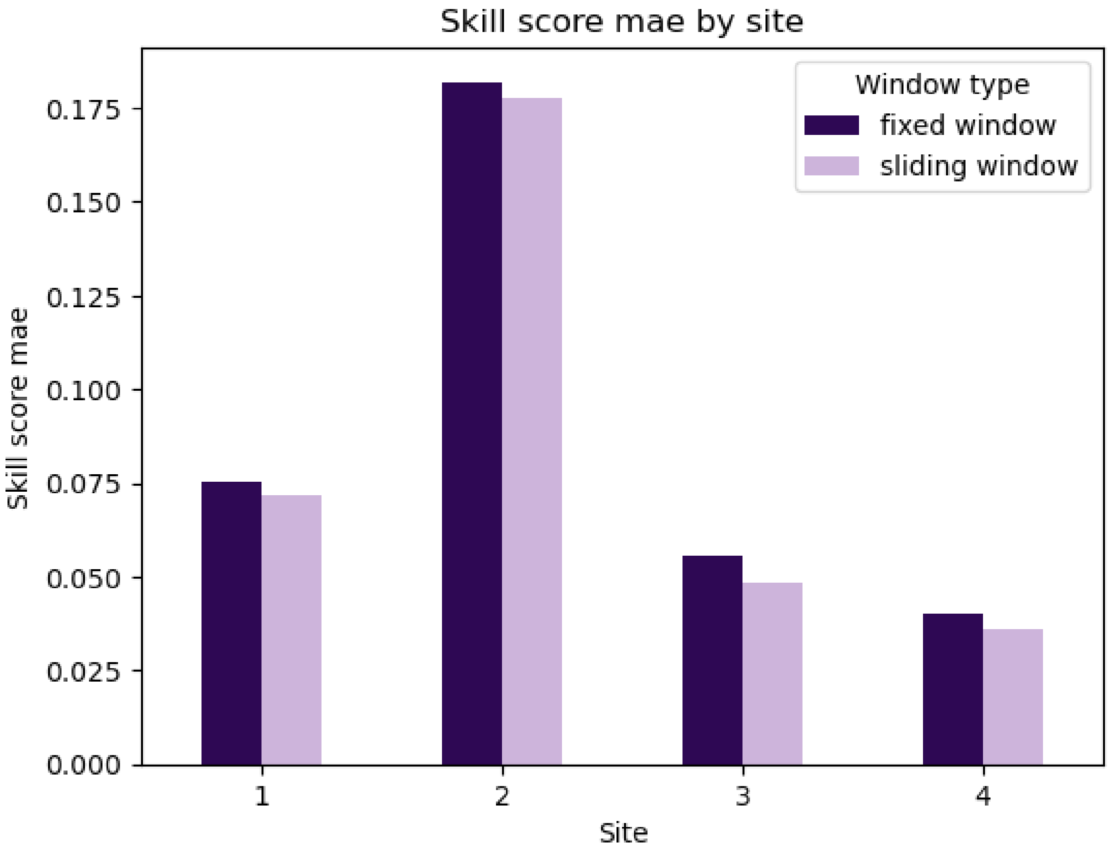


Figure 15: Comparison of the MAE skill scores for a SVR model.

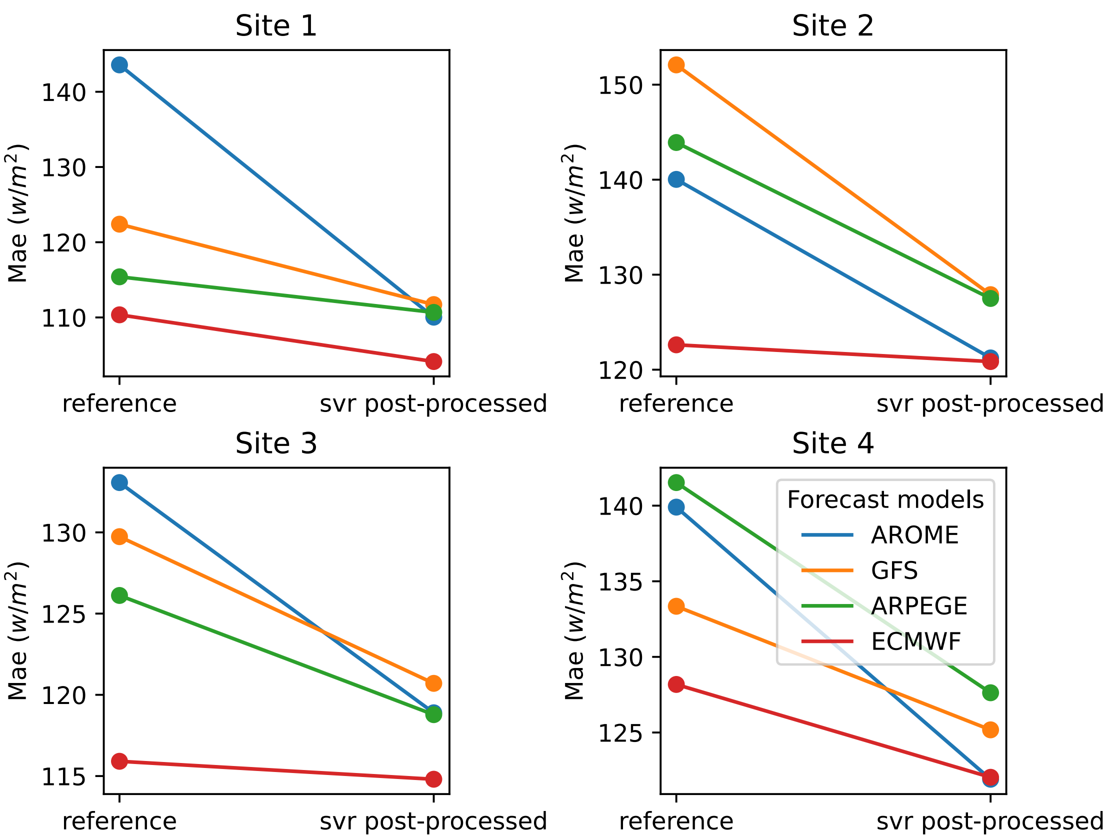


Figure 16: Comparison of the post-processing of four different NWP forecast models on MAE.

	SGDRegressor	HuberRegressor	Lasso	LassoLars	TweedieRegressor	BayesianRidge	ElasticNet	LinearRegression	Lars	ARDRegression	Ridge	Reference	PassiveAggressiveRegressor	TheilSenRegressor	GammaRegressor	Score
SGDRegressor	NA	3	4	4	4	4	4	4	4	4	4	4	4	4	4	55.0
HuberRegressor	1	NA	3	3	4	4	4	4	4	4	4	4	4	4	4	51.0
Lasso	0	1	NA	1	4	4	3	3	3	4	3	3	3	3	4	39.0
LassoLars	0	1	1	NA	4	4	3	3	3	4	3	3	3	3	4	39.0
TweedieRegressor	0	0	0	0	NA	3	3	2	2	4	3	3	3	3	4	30.0
BayesianRidge	0	0	0	0	1	NA	3	3	3	3	3	3	3	3	4	29.0
ElasticNet	0	0	1	1	1	1	NA	2	2	3	4	3	3	3	4	28.0
LinearRegression	0	0	1	1	2	1	2	NA	2	2	2	3	3	3	4	26.0
Lars	0	0	1	1	2	1	2	1	NA	2	2	3	3	3	4	25.0
ARDRegression	0	0	0	0	0	1	1	2	2	NA	3	3	3	3	4	22.0
Ridge	0	0	1	1	1	1	0	2	2	1	NA	3	3	3	4	22.0
Reference	0	0	1	1	1	1	1	1	1	1	1	NA	2	3	3	17.0
PassiveAggressiveRegressor	0	0	1	1	1	1	1	1	1	1	1	2	NA	3	3	17.0
TheilSenRegressor	0	0	1	1	1	1	1	1	1	1	1	1	1	NA	4	15.0
GammaRegressor	0	0	0	0	0	0	0	0	0	0	0	1	1	0	NA	2.0

Figure 17: 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.

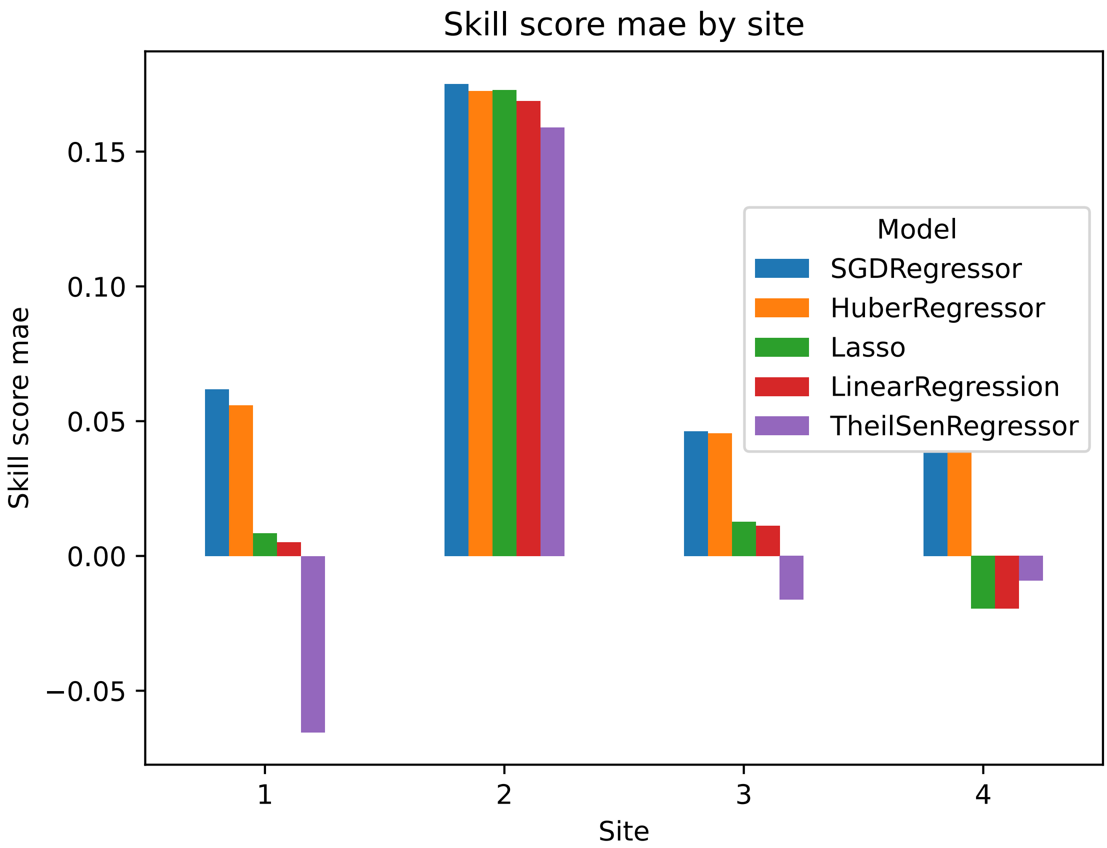


Figure 18: MAE skill score plots of the best linear models.

	TweedieRegressor	ElasticNet	Lasso	LassoLars	BayesianRidge	HuberRegressor	Ridge	SGDRegressor	ARDRegression	Lars	LinearRegression	TheilSenRegressor	Reference	GammaRegressor	PassiveAggressiveRegressor	Score
TweedieRegressor	NA	3	4	4	4	3	3	2	4	3	3	3	4	4	4	48.0
ElasticNet	1	NA	2	2	2	2	2	2	2	3	3	3	4	4	4	36.0
Lasso	0	2	NA	1	2	2	3	2	3	3	3	3	4	4	4	36.0
LassoLars	0	2	1	NA	2	2	3	2	3	3	3	3	4	4	4	36.0
BayesianRidge	0	2	2	2	NA	2	2	2	2	3	3	3	4	4	4	35.0
HuberRegressor	1	2	2	2	2	NA	2	3	2	2	2	3	4	4	4	35.0
Ridge	1	2	1	1	2	2	NA	2	2	3	3	3	4	4	4	34.0
SGDRegressor	2	2	2	2	2	1	2	NA	2	2	2	3	4	4	4	34.0
ARDRegression	0	2	1	1	2	2	2	2	NA	2	2	3	4	4	4	31.0
Lars	1	1	1	1	1	2	1	2	2	NA	2	3	4	4	4	29.0
LinearRegression	1	1	1	1	1	2	1	2	2	2	NA	3	4	4	4	29.0
TheilSenRegressor	1	1	1	1	1	1	1	1	1	1	1	NA	4	4	4	23.0
Reference	0	0	0	0	0	0	0	0	0	0	0	0	NA	3	4	7.0
GammaRegressor	0	0	0	0	0	0	0	0	0	0	0	0	1	NA	3	4.0
PassiveAggressiveRegressor	0	0	0	0	0	0	0	0	0	0	0	0	0	1	NA	1.0

Figure 19: 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.

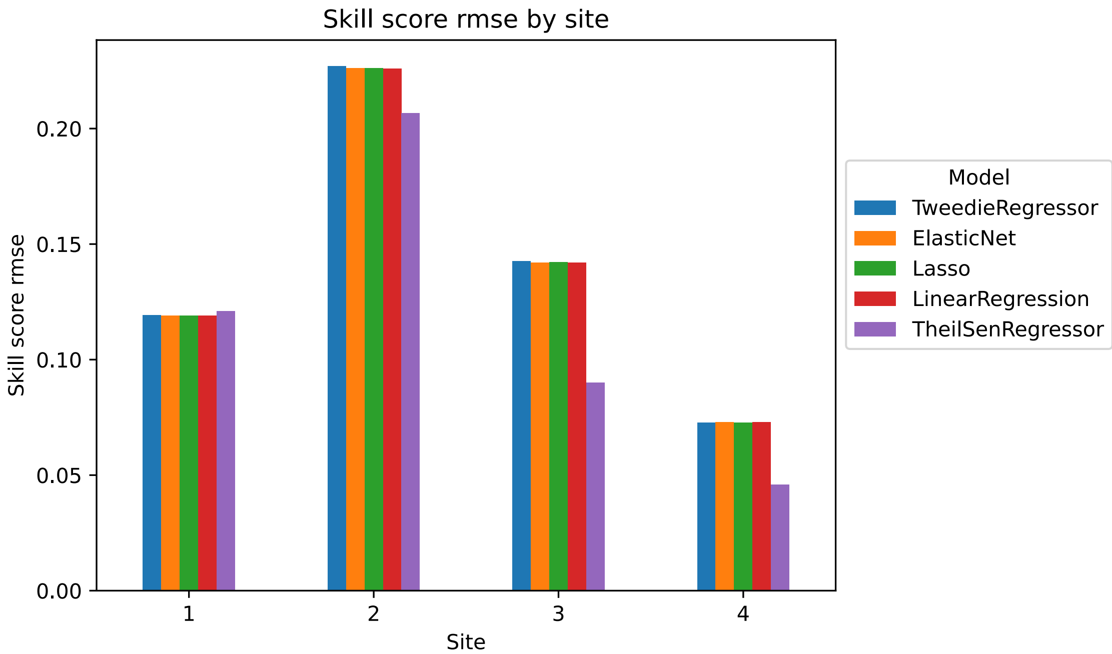


Figure 20: RMSE skill score plots of the best linear models.

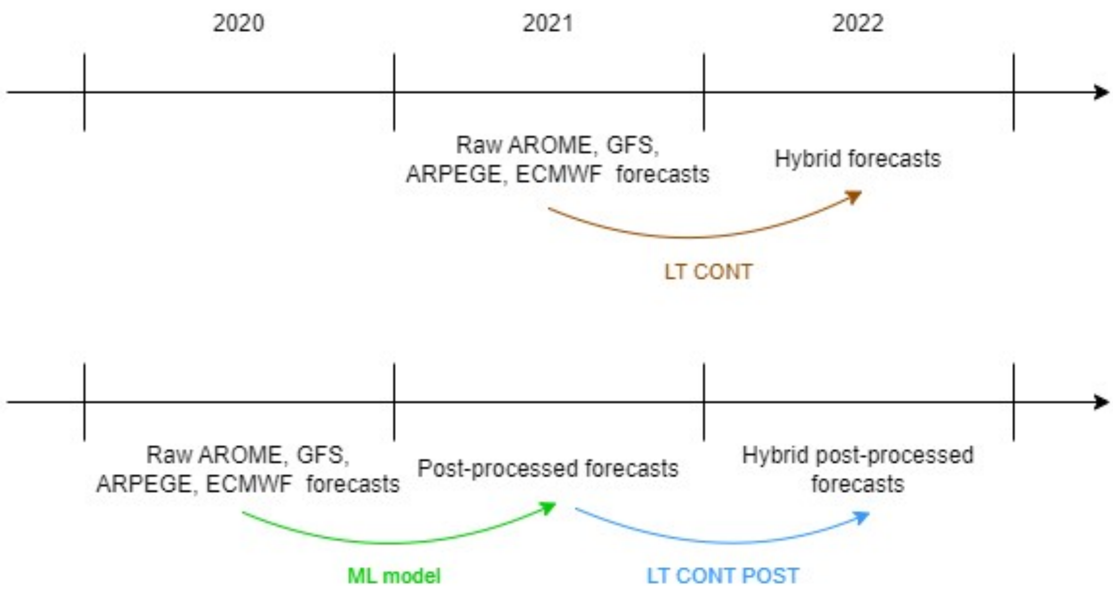


Figure 21: Comparison of the two versions of LT CONT.

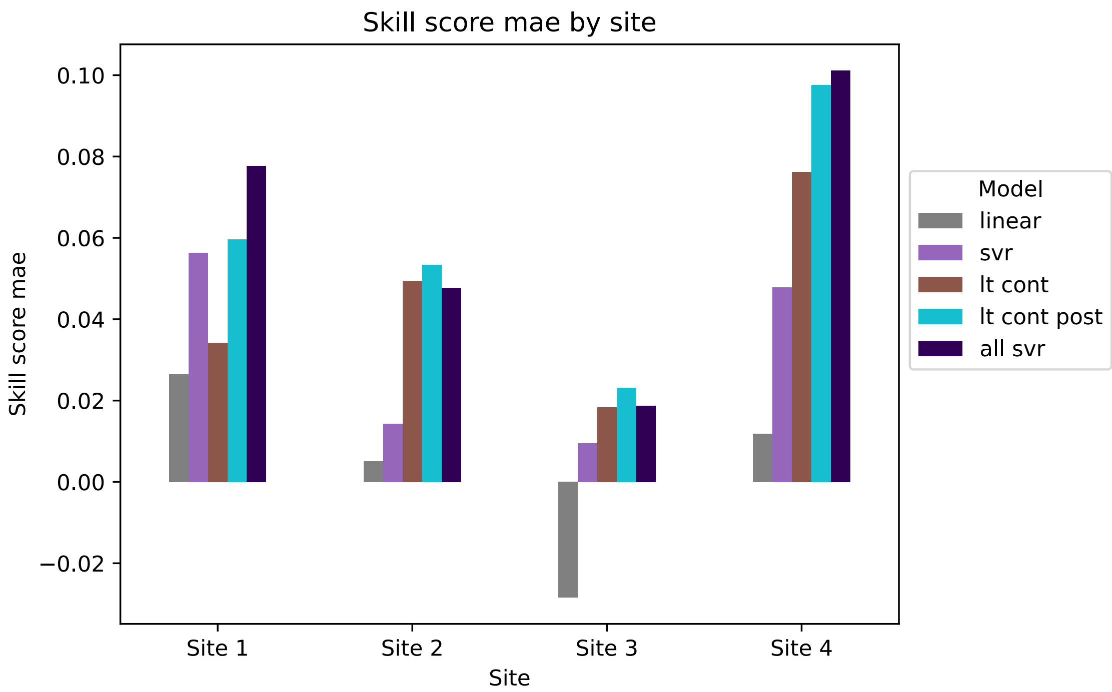


Figure 22: MAE skill score plots of the relevant models, the reference being here the best NWP forecast (in practise, this means ECMWF).

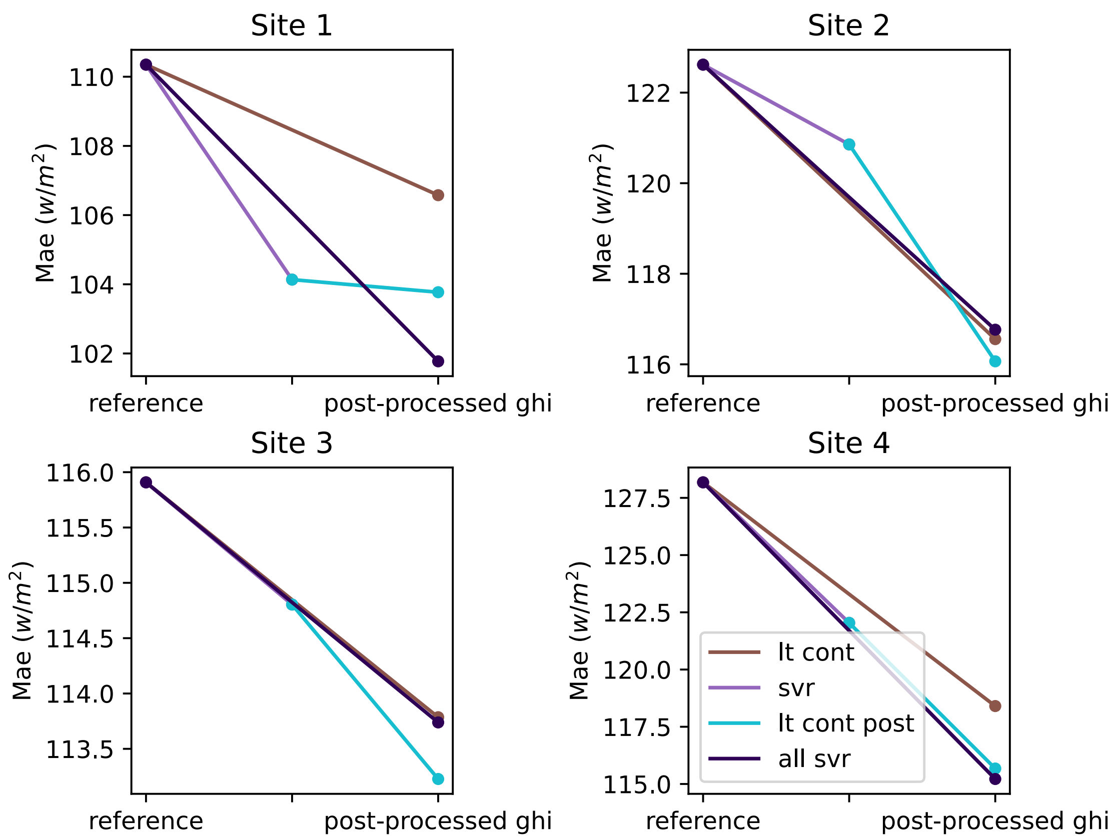


Figure 23: MAE evolution of the relevant models.

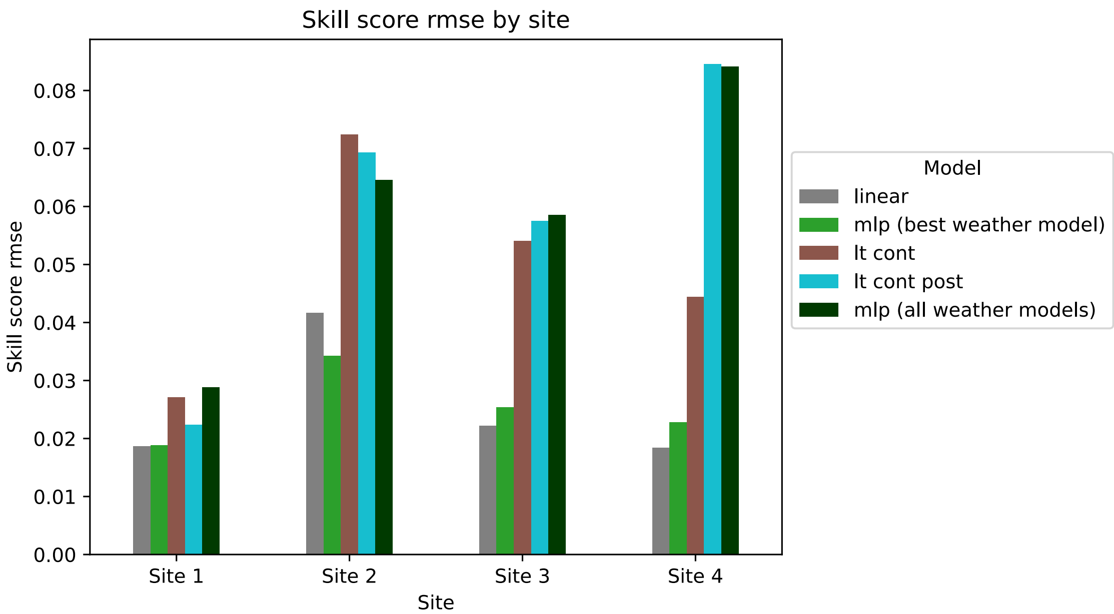


Figure 24: RMSE skill score plots of the relevant models, the reference being here the best NWP forecast (in practise, this means ECMWF).

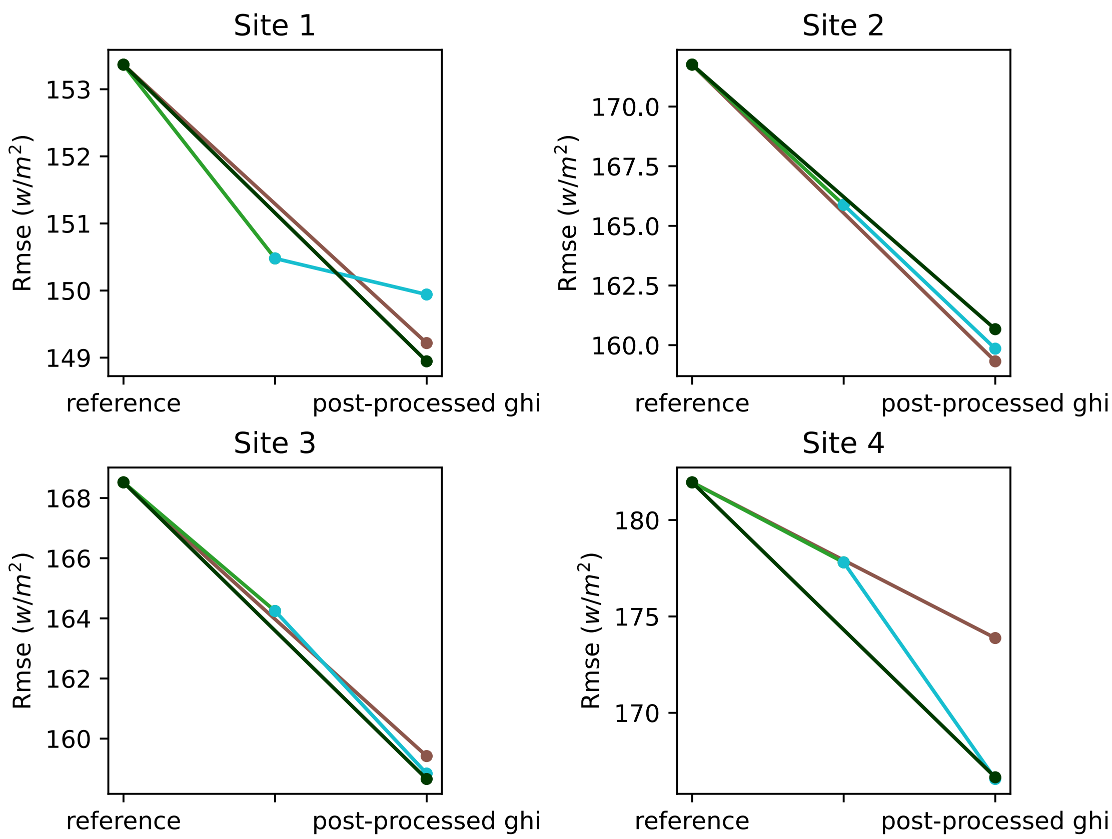


Figure 25: RMSE evolution of the relevant models.

Figure 26: The 25 German sites used for validation.

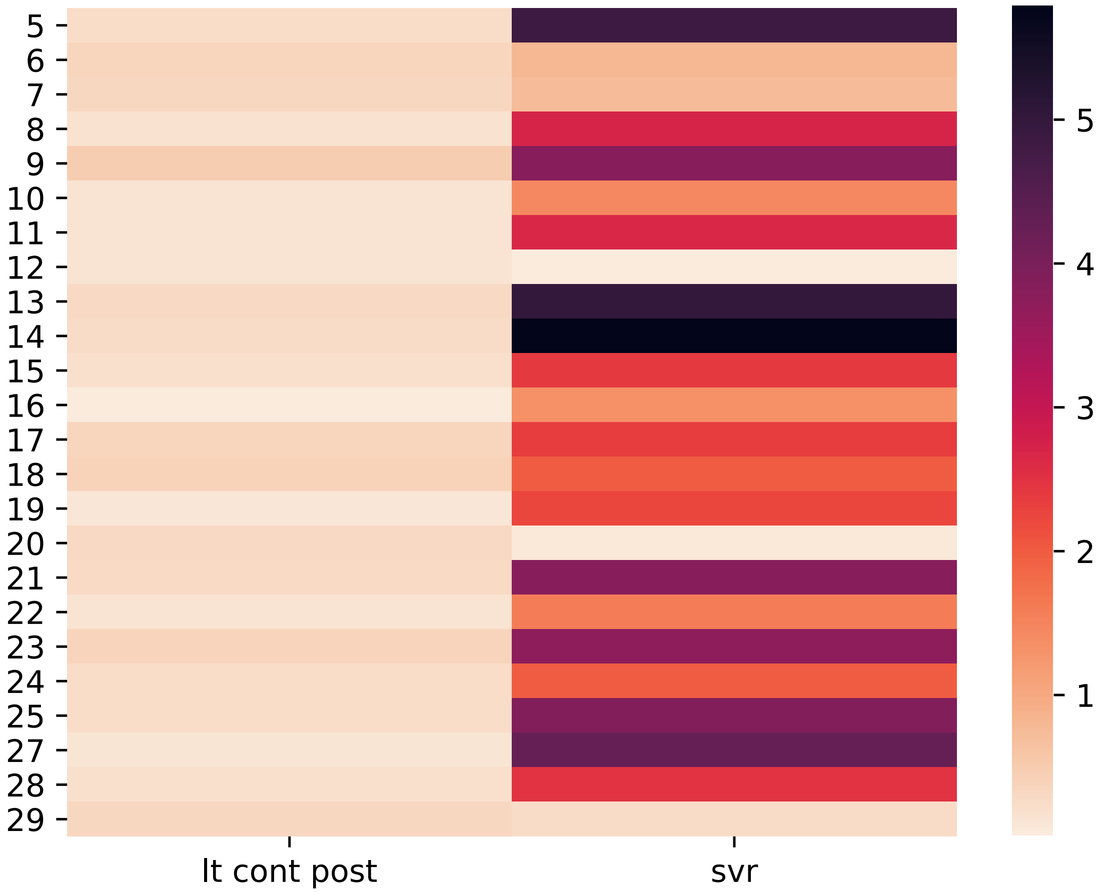


Figure 27: MAE skill score heatmap of the validation sites.

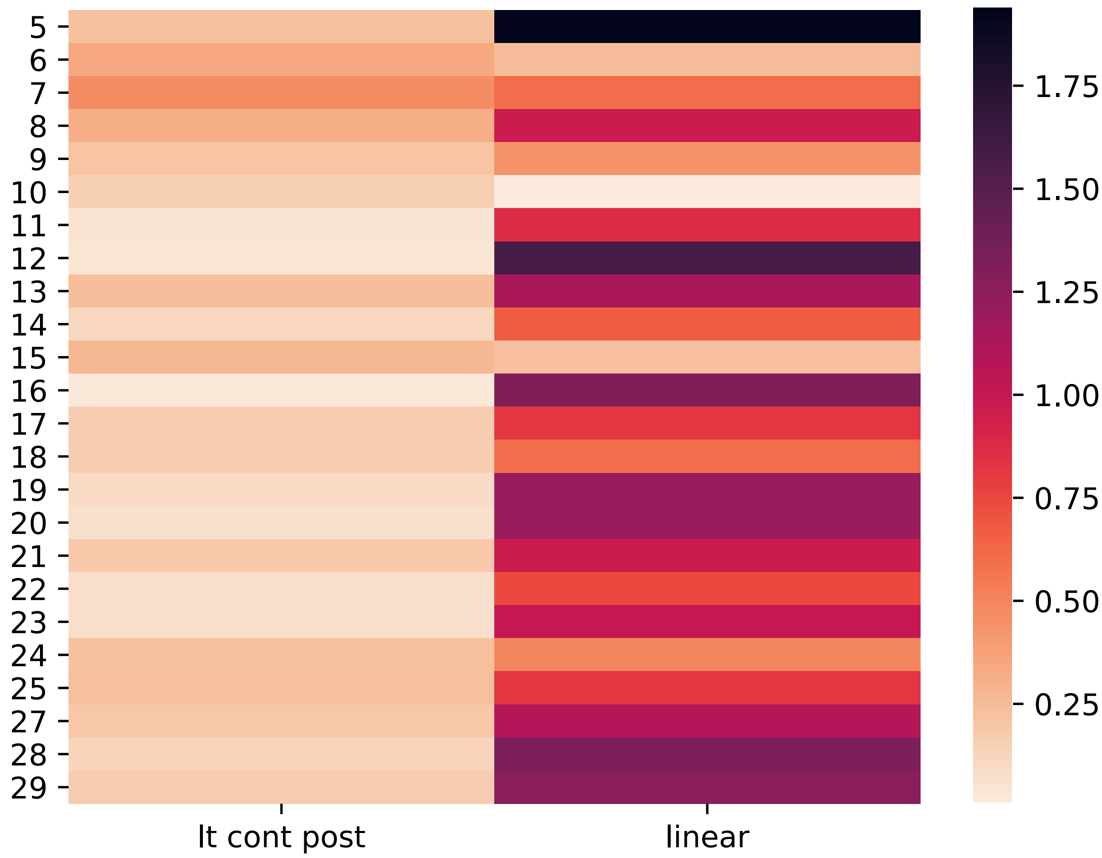


Figure 28: RMSE skill score heatmap of the validation sites.

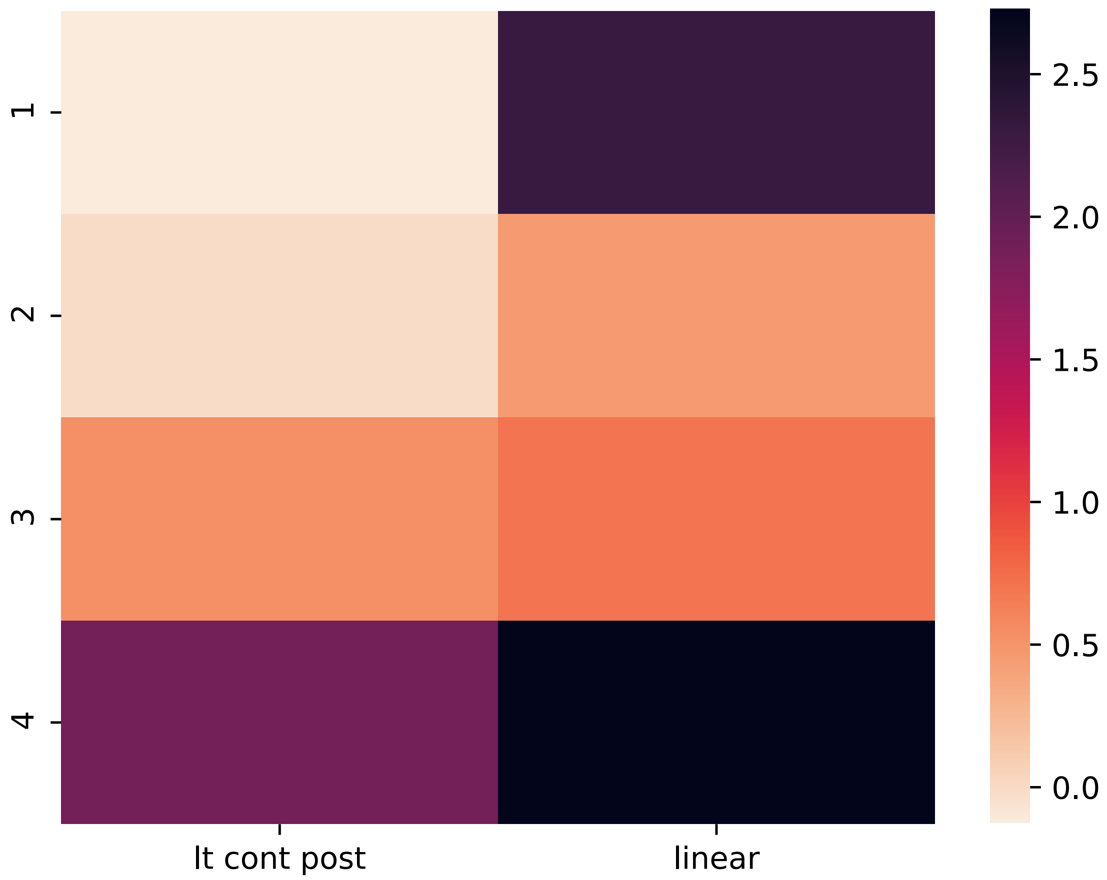
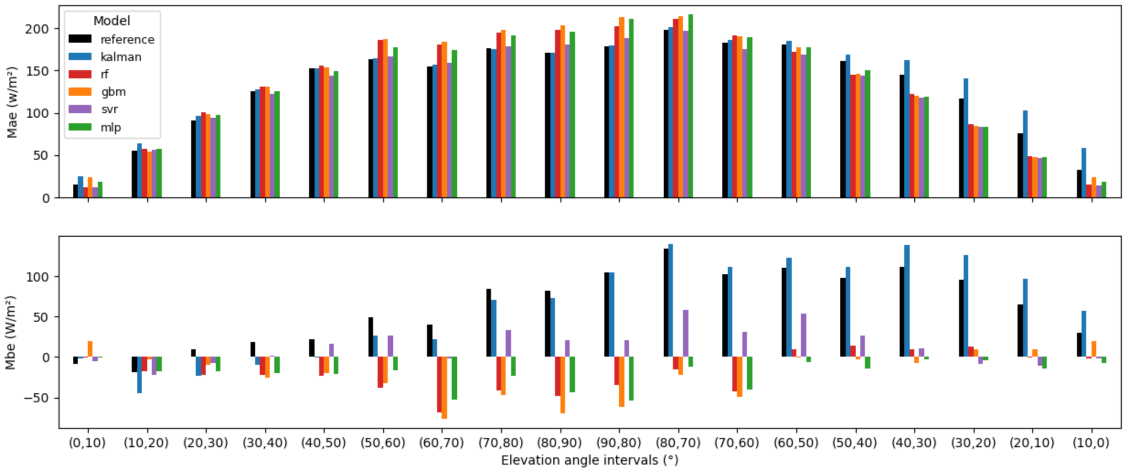
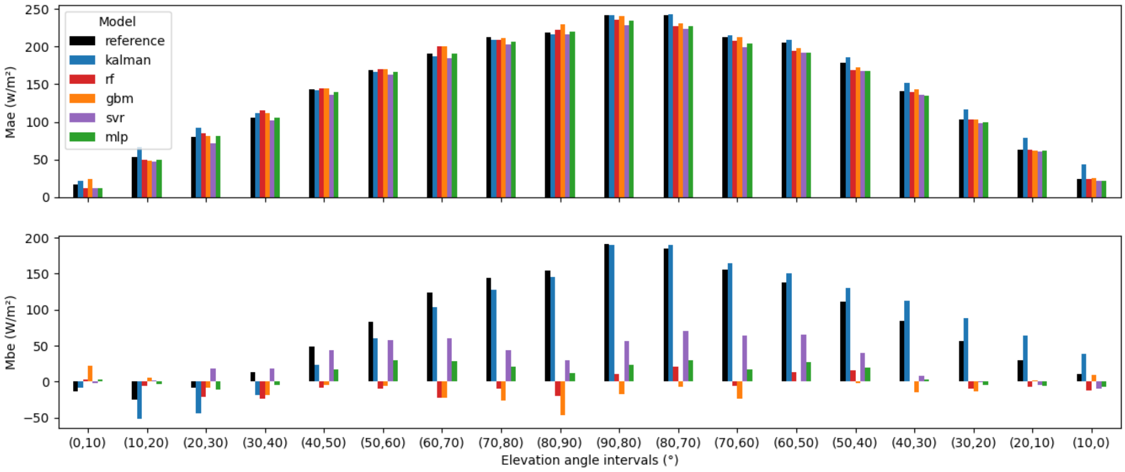


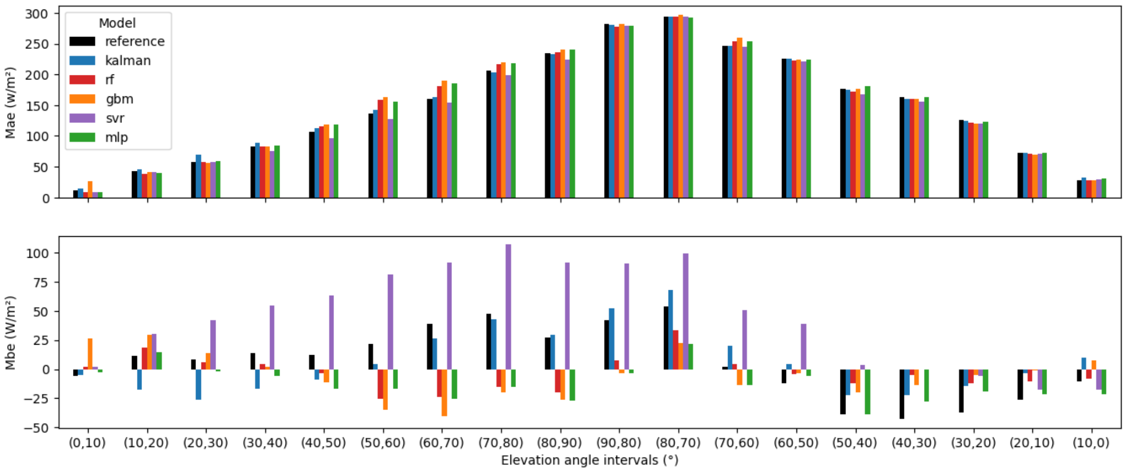
Figure 29: Global linear model verification on the 4 initial sites.



(a) Site 1

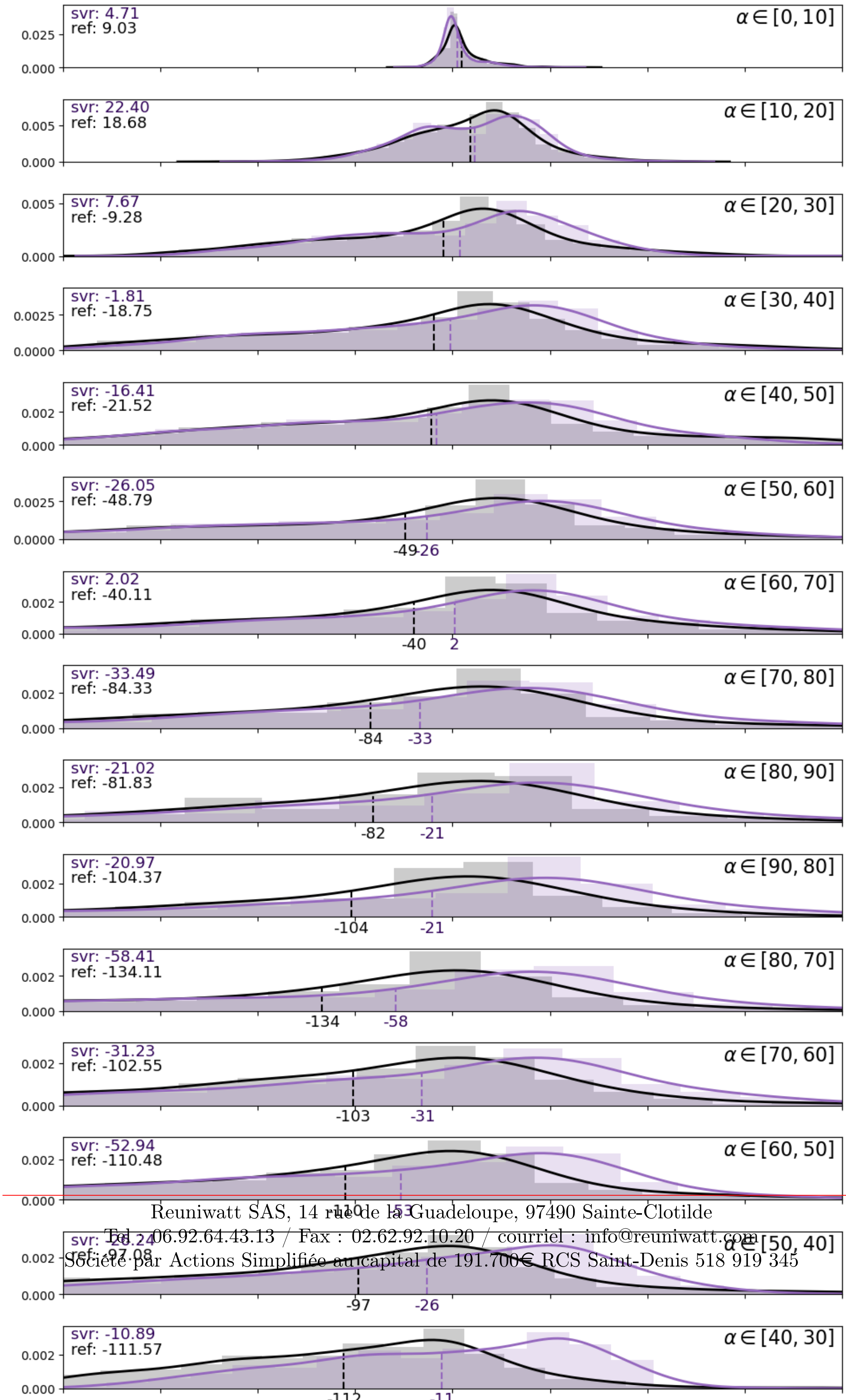


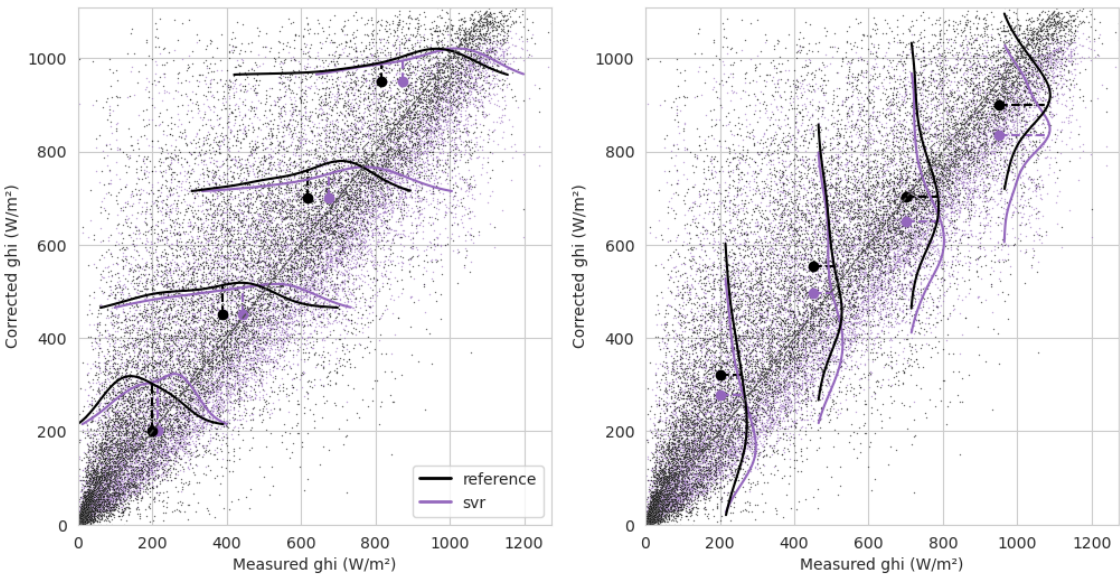
(b) Site 3



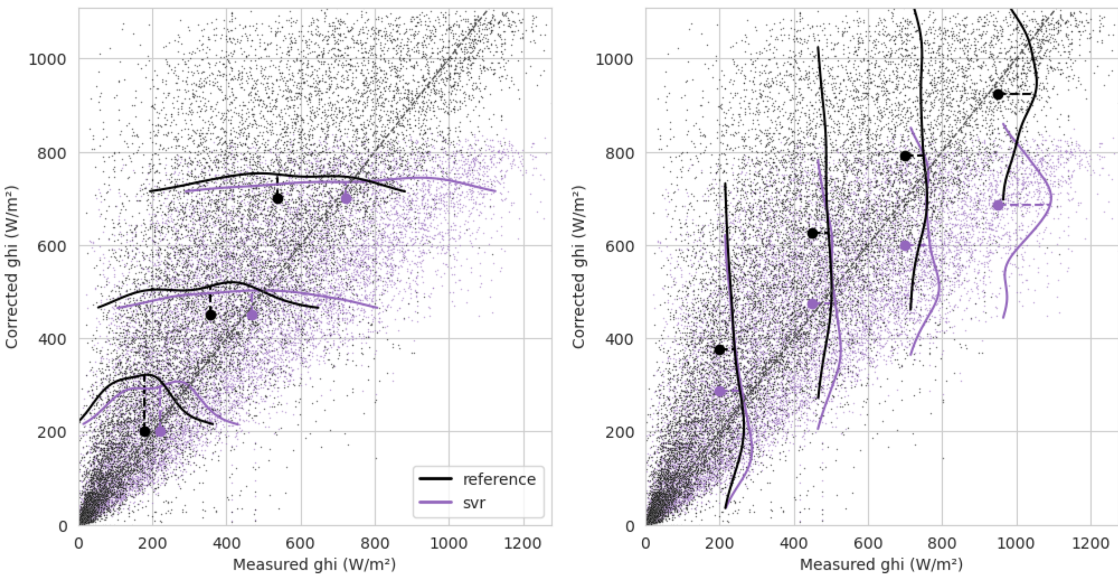
(c) Site 4

Figure 30: MAE and MBE levels across all elevation angle intervals of a day.

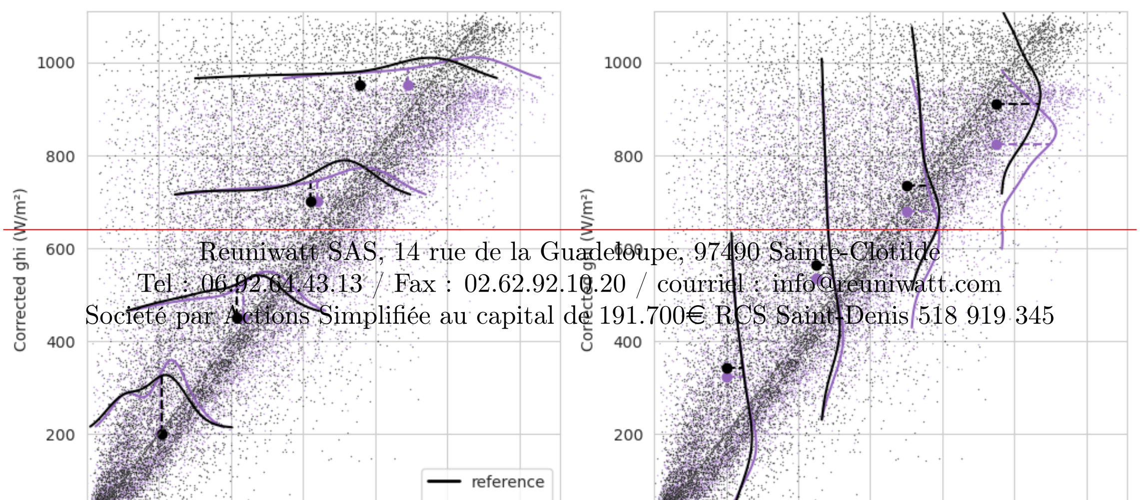

svr: 4.71
ref: 9.03
α ∈ [0, 10]
svr: 22.40
ref: 18.68
α ∈ [10, 20]
svr: 7.67
ref: -9.28
α ∈ [20, 30]
svr: -1.81
ref: -18.75
α ∈ [30, 40]
svr: -16.41
ref: -21.52
α ∈ [40, 50]
svr: -26.05
ref: -48.79
α ∈ [50, 60]
-49
-26
svr: 2.02
ref: -40.11
α ∈ [60, 70]
-40
2
svr: -33.49
ref: -84.33
α ∈ [70, 80]
-84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

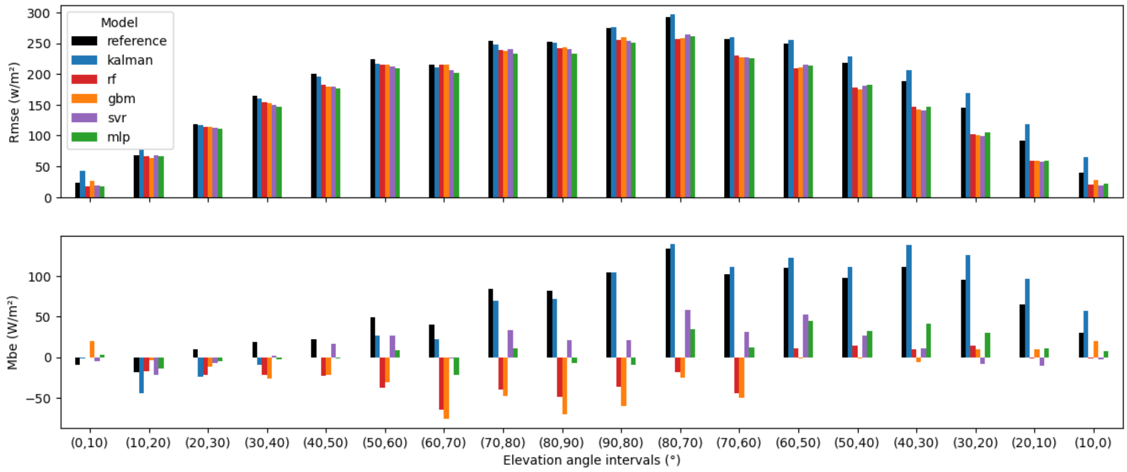


(a) Site 1

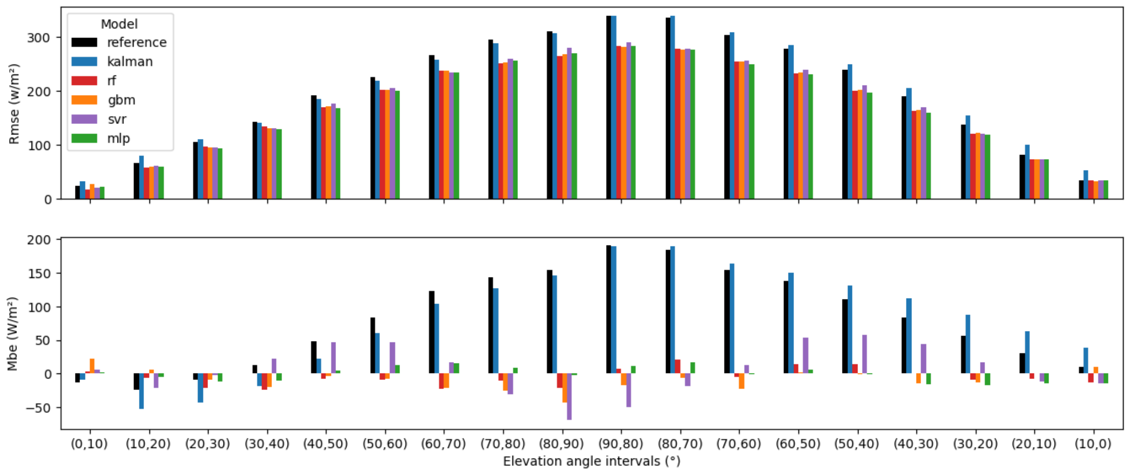


(b) Site 2

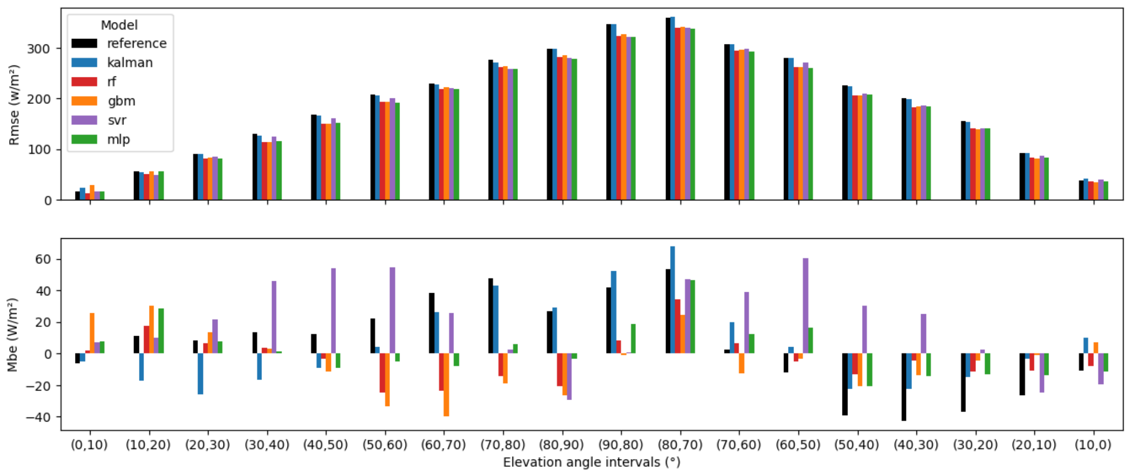




(a) Site 1

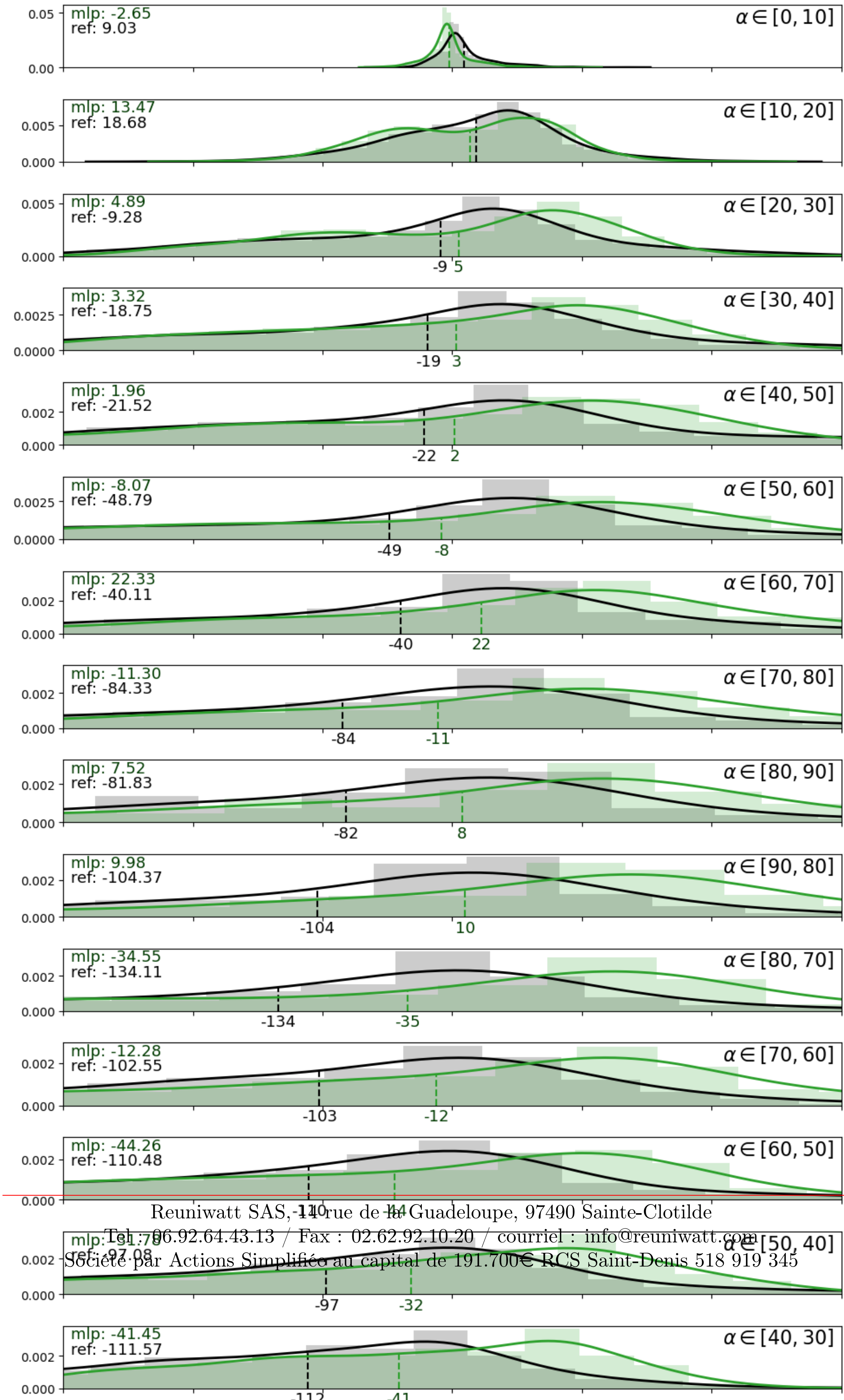


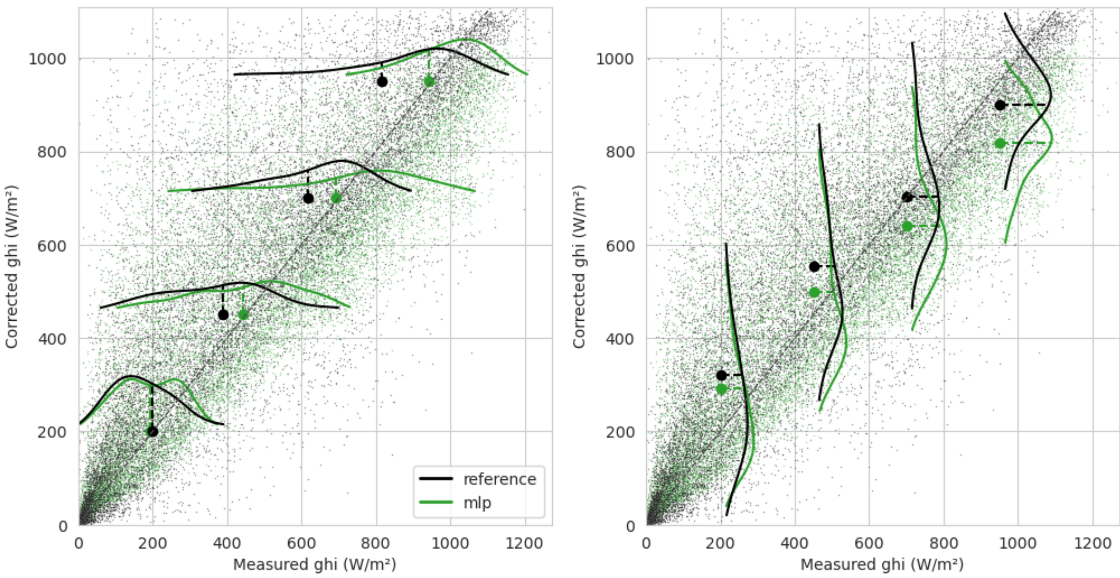
(b) Site 3



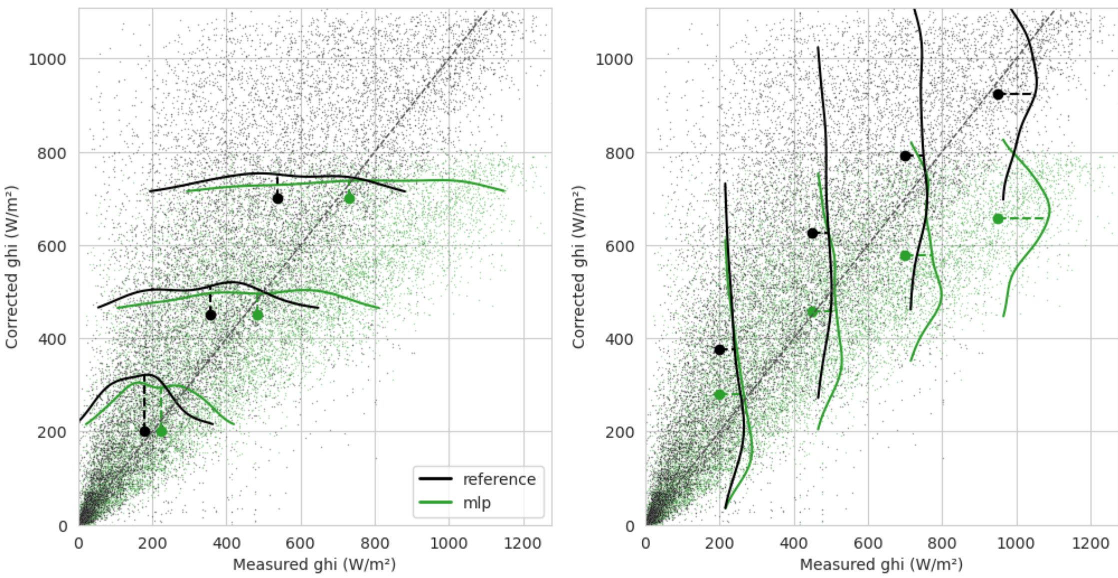
(c) Site 4

Figure 33: RMSE and MBE levels across all elevation angle intervals of a day.

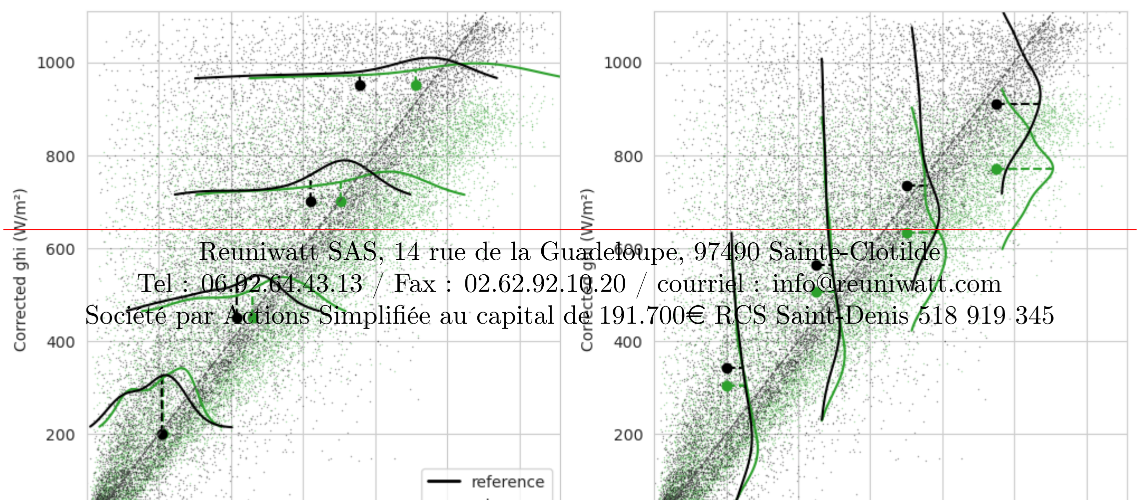

mlp: -2.65
ref: 9.03
α ∈ [0, 10]
mlp: 13.47
ref: 18.68
α ∈ [10, 20]
mlp: 4.89
ref: -9.28
α ∈ [20, 30]
-9
5
mlp: 3.32
ref: -18.75
α ∈ [30, 40]
-19
3
mlp: 1.96
ref: -21.52
α ∈ [40, 50]
-22
2
mlp: -8.07
ref: -48.79
α ∈ [50, 60]
-49
-8
mlp: 22.33
ref: -40.11
α ∈ [60, 70]
-40
22
mlp: -11.30
ref: -84.33
α ∈ [70, 80]
-84
-11
mlp: 7.52
ref: -81.83
α ∈ [80, 90]
-82
8
mlp: 9.98
ref: -104.37
α ∈ [90, 80]
-104
10
mlp: -34.55
ref: -134.11
α ∈ [80, 70]
-134
-35
mlp: -12.28
ref: -102.55
α ∈ [70, 60]
-103
-12
mlp: -44.26
ref: -110.48
α ∈ [60, 50]
-110
-44
ref: -97.08
α ∈ [50, 40]
-97
-32
mlp: -41.45
ref: -111.57
α ∈ [40, 30]



(a) Site 1



(b) Site 2



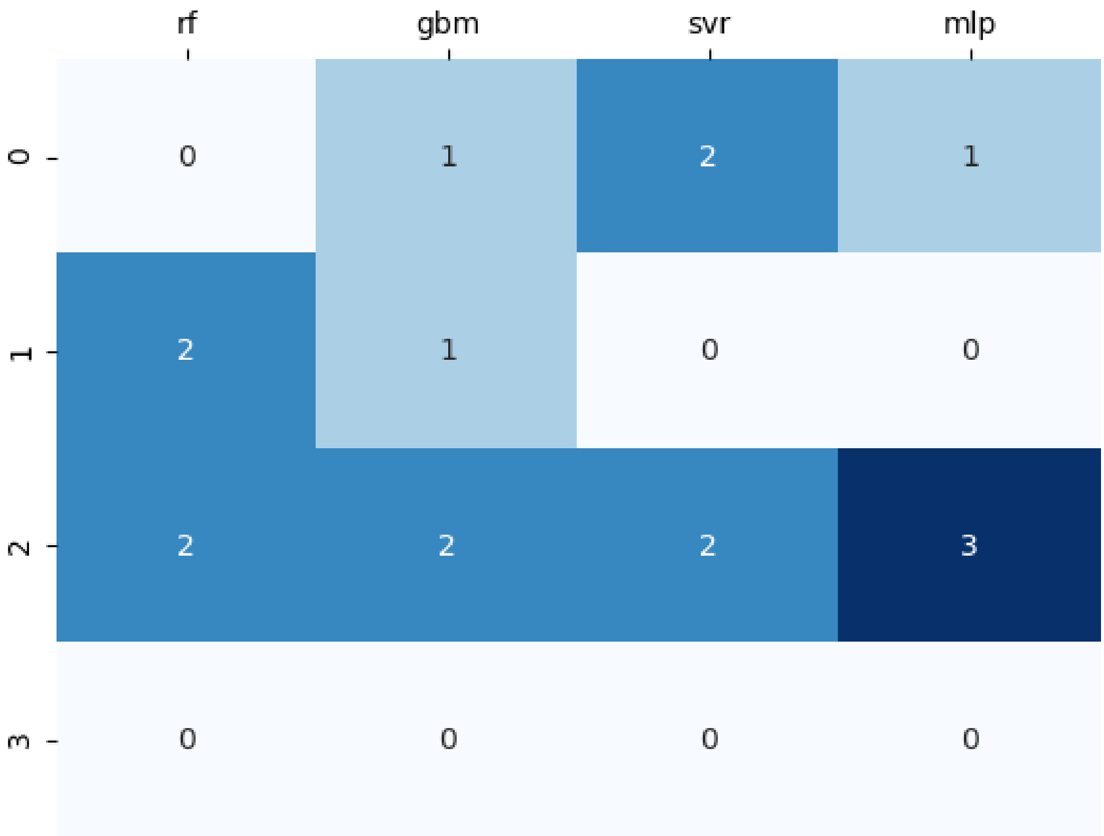


Figure 36: Pairwise systematicity matrix for RMSE. The value $V_{i,j}$ of the cell (i, j) indicates how often the configuration of line i is the best one, across the 4 sites, for the model of column j. For example, the configuration 0 is the best one with a GBM post-processing for 3 sites, and the configuration 2 is the best one for 1 site.

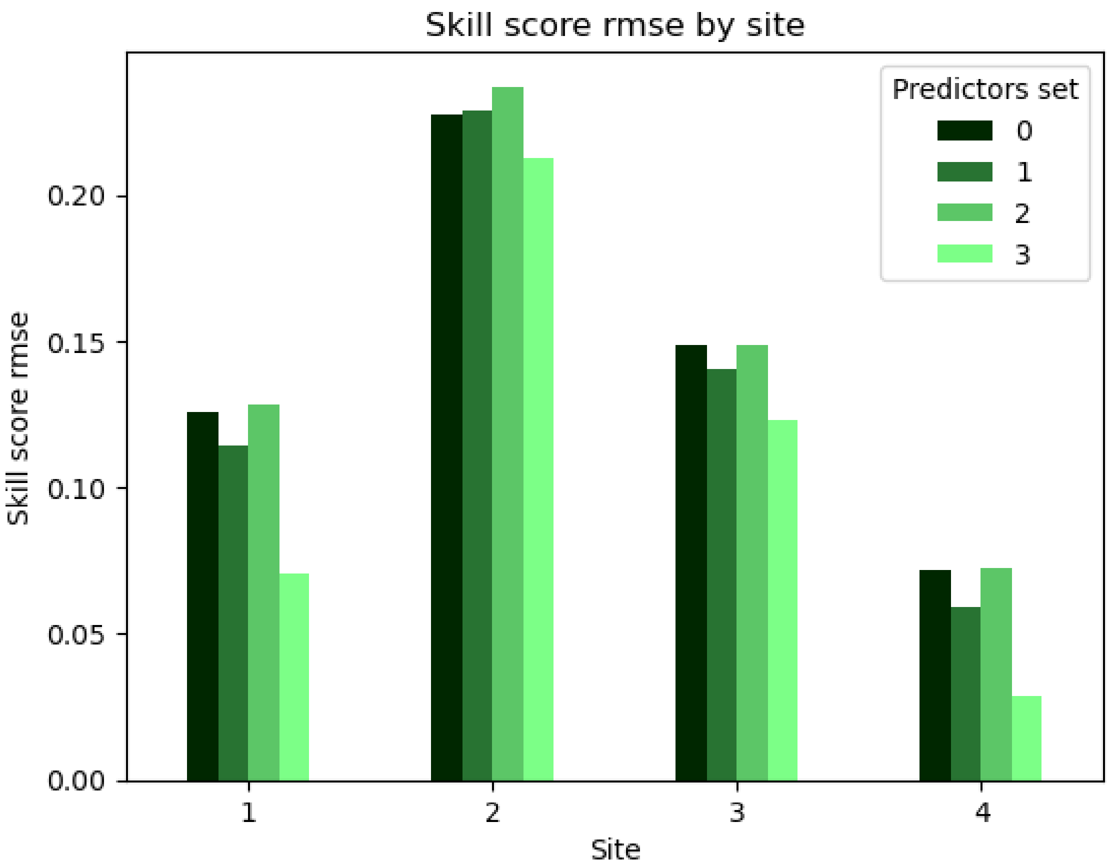


Figure 37: Comparison of the RMSE skill scores of the different configurations.

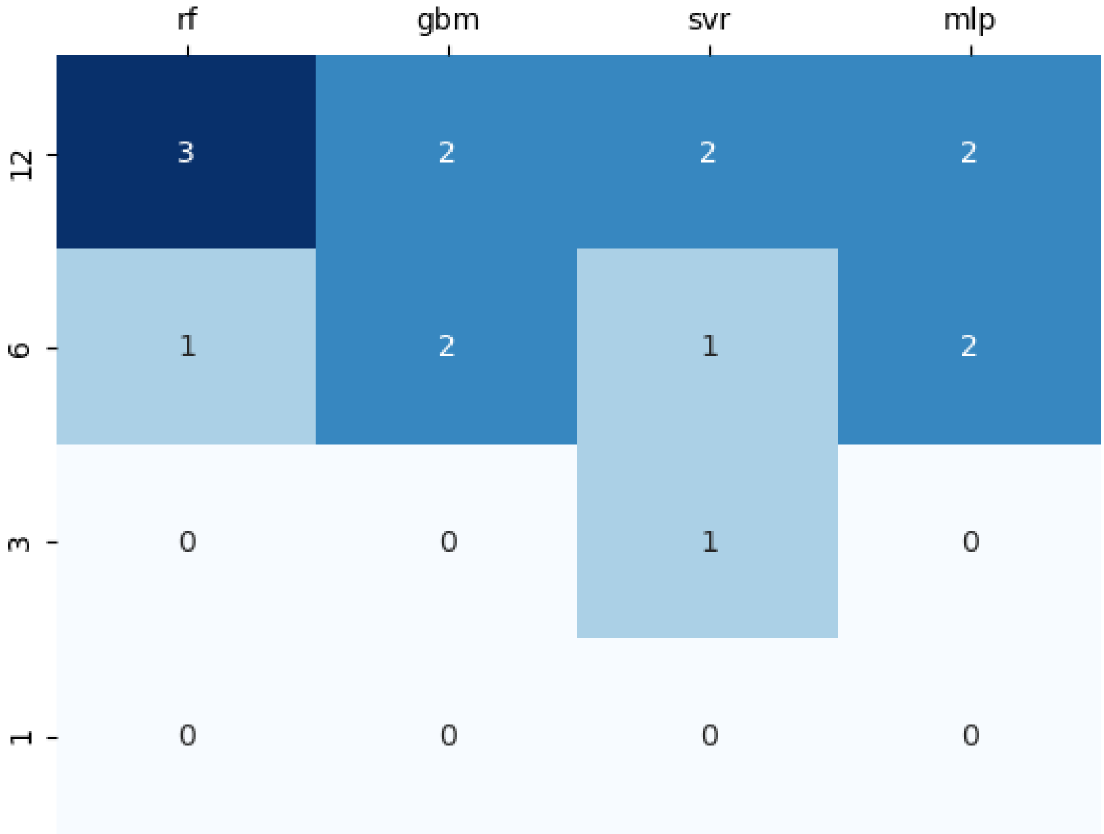


Figure 38: Pairwise systematicity matrix for RMSE. The value $V_{i,j}$ of the cell (i, j) indicates how often the learning period duration (in months) of line i performs the best, across the 4 sites, for the model of column j. For example, having a 12-months-long learning period is the best thing in 3 sites out of 4 with a SVR model, the last site performs better with a 6-month-long learning period.

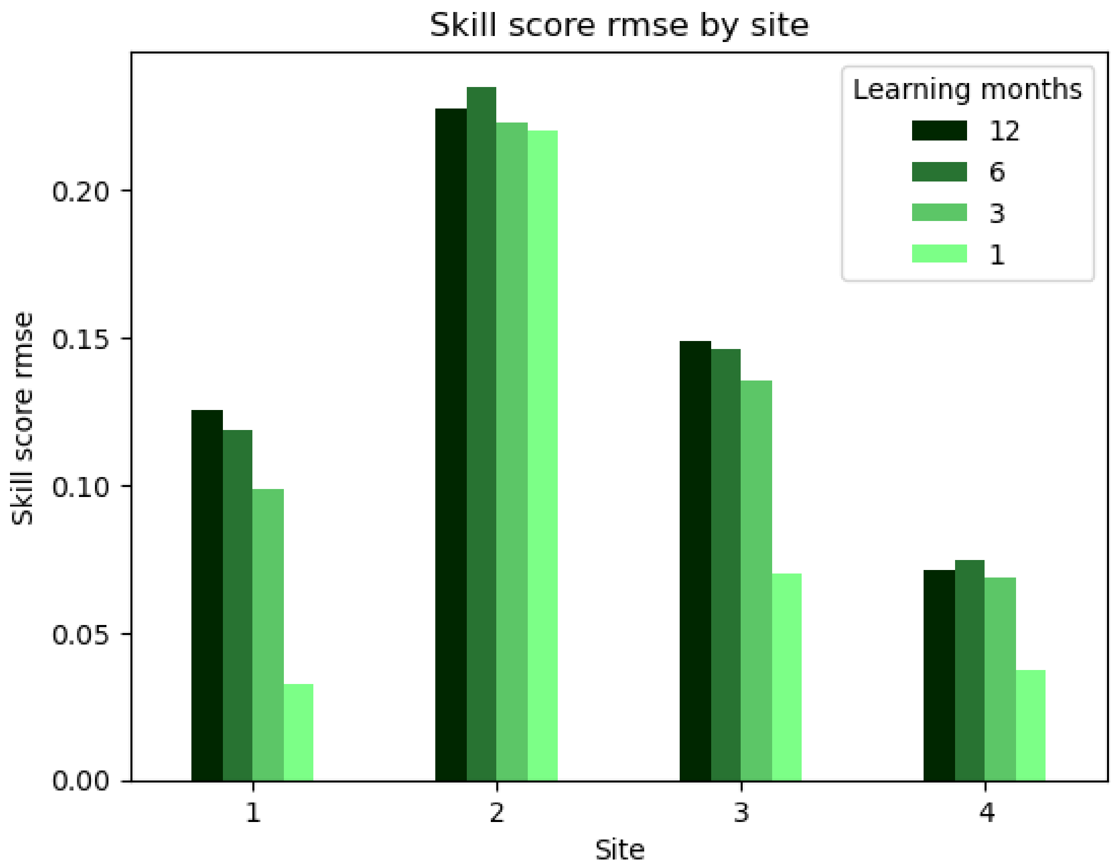


Figure 39: Comparison of the RMSE skill scores of the different learning period durations (in months).

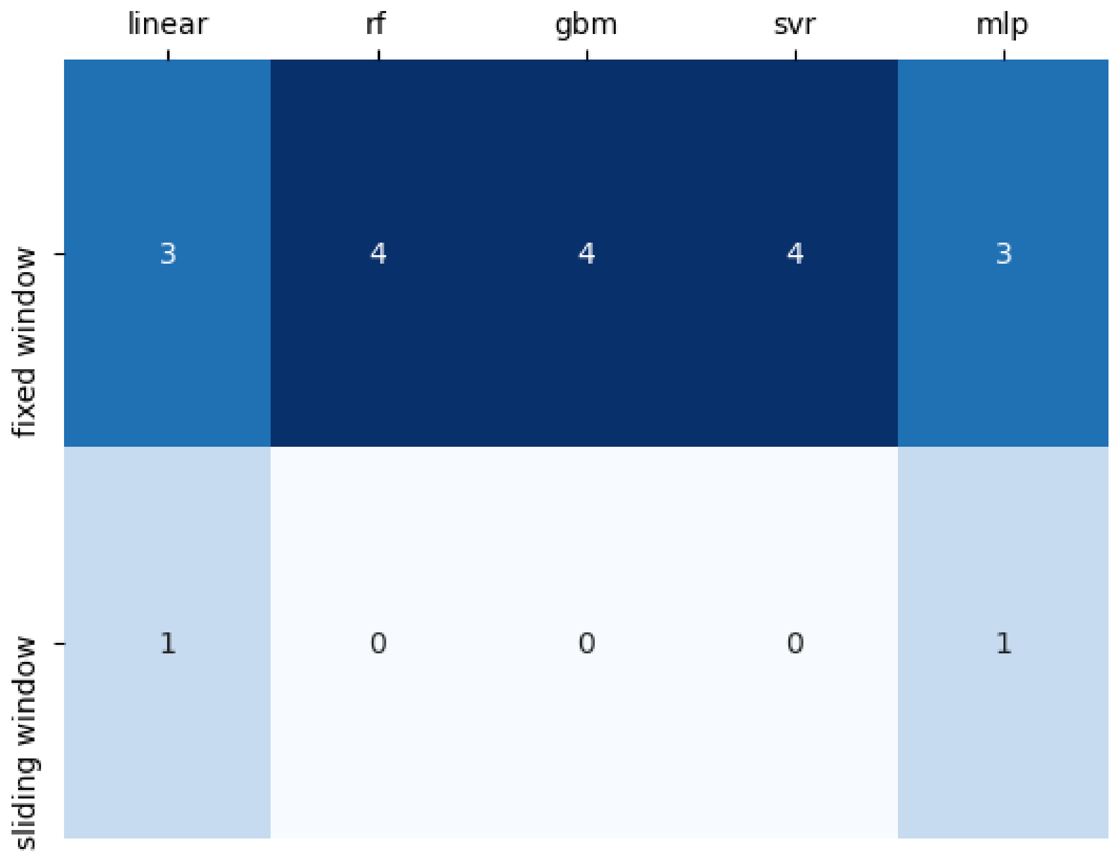


Figure 40: Pairwise systematicity matrix concerning window type for RMSE.

Figure 41: Comparison of the RMSE skill scores for a SVR model.

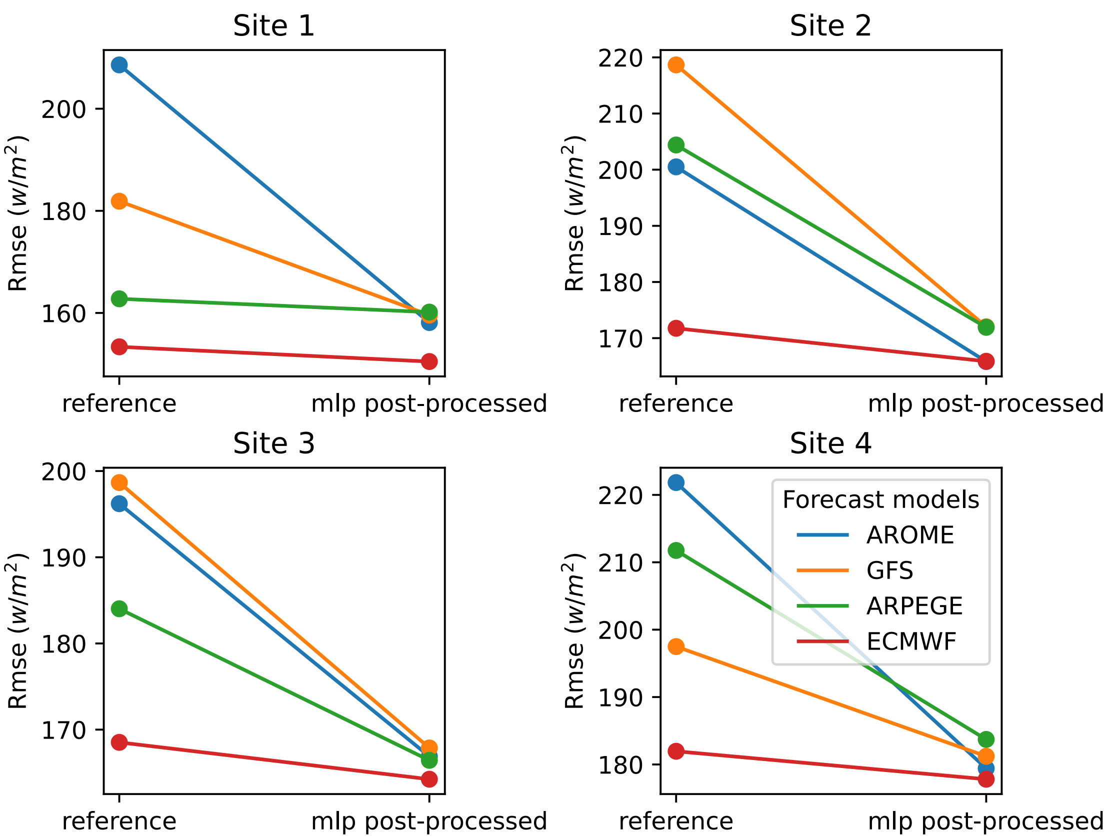


Figure 42: Comparison of the post-processing of four different NWP forecast models on RMSE.