Project 3: Ensembles
On the use of Ensembles for Weather Forecasting in Lugano

Matea Leahy, University College Dublin
Advanced Computational Science

Accepted: 20/04/2021

Abstract
The following report consists of an analysis of the Weather in Lugano, Switzerland by use of ensemble forecasting. The ensemble represents the mean of results taken from 9 different forecast models. This report is split into two sections. In section one I perform an analysis on the uncertainties and skill scores of the Ensemble and single model forecasts by using deterministic and probabilistic methods to show the improvements achieved by using a multi-model instead of a single-model ensemble. In order to understand the success of the multi-model approach a number of skill scores are employed to compare the relative performance of the models when separate and when combined. A comparison has also been made to how Climatology succeeded in predicting the weather by comparison to the observed data from the weather in Lugano in December and January of 2020/2021. These skill scores are computed for each of the 9 models used within my ensemble and are compared to the performance of the Ensemble itself. Some improvements can be made and a discussion of the systematic errors in place is provided. Section two consists of an investigation of climate projections for the city Lugano where a number of models are analysed to deduce how well they represent the climate in Lugano. The projected change in temperature for each model is then computed and the results are analysed.

Key words: Weather Forecasting – Ensemble – Evaluating Forecasts

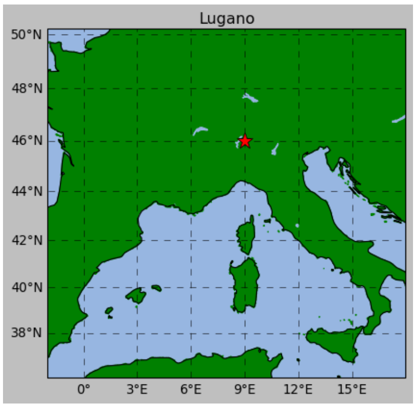

Figure 1. Lugano on the map

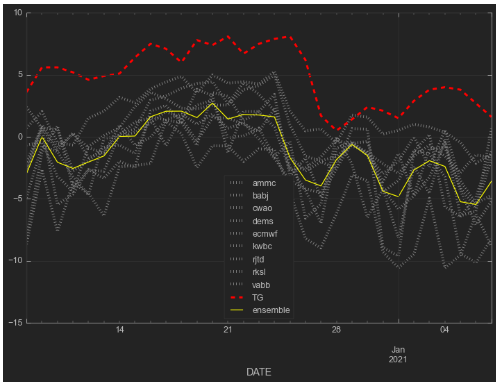

Figure 2. A plot of the individual model's Bias and forecast errors

1 Skill Scores

The calculation of skill scores to analyse the uncertainty of models is important in weather forecasting. Both probabilistic and deterministic methods are used to discuss how these models may perform. Something that must first be mentioned is that all analysis for this report has been performed in Python where the nearest integer longitude and latitude are taken. Hence where Lugano is represented by 46.0° Latitude and 8.97° longitude, the points used in the report are 46.0° latitude and 9.0° longitude (unfortunately not actually on the lake like I had originally suspected after a zoomed in view of my Cartopy plot). This preemptively leads to a small amount of some systematic Bias. Another point worth mentioning before the computation of the skill scores is that the date range used in the models for the multi model ensemble is from the 8th of December 2020 to the 7th of January 2021. This short window of time over which the skill scores are calculated does not necessarily give enough information to make definitive changes to the model based on the results. Hence trying to increase the specific match of this one process can leave the model more exposed to errors for other processes. For example, my models being well behaved over the course of these winter months does not necessarily mean corrections will apply for summer months which can follow different trends. However a discussion will still be presented for the behaviour of the model with the systematic errors removed simply in reference to how one might make improvements in general. All discussions of the respective skill scores is included in the error exploration section below. The trends in the T2M temperatures for each of the forecast models is compared to the multi-model ensemble and the observed values in figure 2.

Table 1. Forecast Bias (°C)

	ammc	babj	cwao	dems	ecmwf	kwbc	rjtd	rksl	vabb	Ensemble Bias
Bias	-6.215	-6.652	-6.999	-4.989	-3.022	-4.759	-9.701	-5.596	-6.603	-6.060

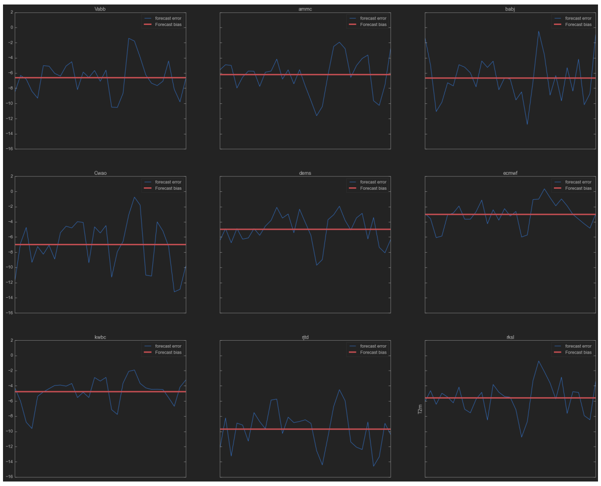

Figure 3. A plot of the individual model's Bias and forecast errors

1.1 Forecast Bias

The Forecast Bias of a model is a deterministic skill measure of the tendency of the model to over-forecast or under forecast which leads to error. The method for measuring the bias I have taken is to compute the mean of the forecast errors when compared to the observed temperatures on that day. The formula used to compute the forecast errors is

$$Forecast - Observed$$

which is calculated for each day, for each model and an arithmetic average is taken of the forecast errors over the 31 days to provide an overall Bias for the corresponding model. The values computed for the error of each day are displayed in Table 1.1 in the appendix and the mean errors (bias) are shown in table 1.

The Biases for all of my forecasts are highly negative with values ranging from -3.022°C corresponding to the ecmwf model to -9.7°C for the rjtd model and an increase in uncertainty can be seen as time increases. The strong Bias in the negative direction suggests that the models used highly underestimated the temperature Lugano and were consistently below the true values of 2 metre temperature(°C). The Bias of the multi-model ensemble is approximately -6.6°C which performs a good deal better than the rjtd whose error plot stands out as an outlier from the others. However this strong negative bias gives good information for where there may exist some systematic errors in all of the models. As well as this, when looking at the plot of the ensemble errors it seems to suggest that the variation in the error increases over time.

There are a number of reasons why the forecast may have been so low one of which is that when assuming the use of 'nearest' longitude and latitude for Lugano. Other reasons include the fact that the weather in January and December was simply especially sunny to what one would normally predict, even by eyeing the data, with no days dropping below zero degrees Celcius. As well as this, when calculating the T2m temperature to be used an average was actually taken over four different time

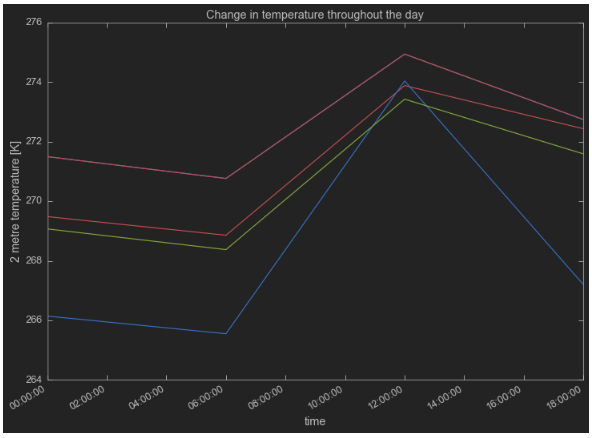

Figure 4. Change in temperature over hourly periods throughout the day for a small sample of forecast models.

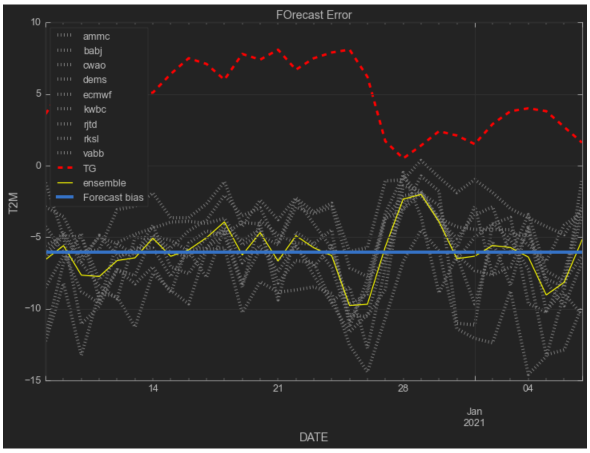

Figure 5. A plot of the Ensemble Errors with Ensemble Bias indicated as a blue line, compared to observed data.

periods including midnight. There is a systematic negative bias in place for many models for the evenings in winter and the observed data does not account for the colder evening temperatures. The temperature variance over one day for a sample of the models in my ensemble is shown in figure 4.

Other more complicated factors such as expected cloudiness which was possibly overestimated in this case however this data has been neglected for this project. Irregardless, this strong Bias gives the first insight as to which systematic errors should be considered. Figure 5 displays the errors in the Midday 2 metre temperatures of my ensemble (yellow) and the corresponding single-models with the ensemble bias marked in blue and the true values of the observed data (red), a similar trend of the weather over the 31 days can be seen. Editing the model to reduce systematic biases will thus be a good idea however it may increase the root-mean-square error (RMSE) because different kinds of errors may no longer partially cancel each other and thus this will be investigated next.

1.2 RMSE

The RMSE of an ensemble is a very commonly used statistical technique to quantify the uncertainty of forecast models.

Table 2. RMSE

ammc	babj	cwao	dems	ecmwf	kwbc	rjtd	rksl	vabb	RMSE
6.656429	7.246194	7.696618	5.351033	3.41782	5.071943	10.033387	5.981851	6.943227	6.303707

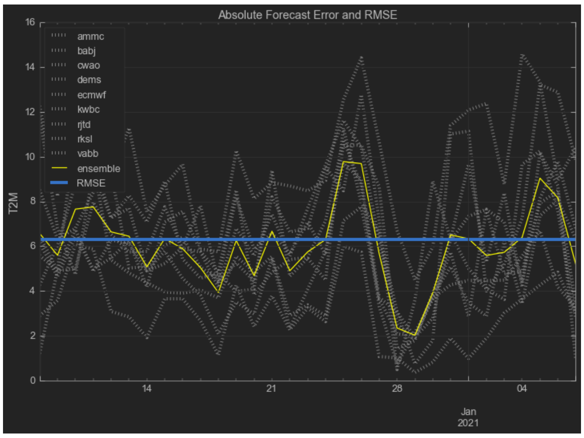

Figure 6. A plot of the Absolute value of the Ensemble Errors with Ensemble RMSE indicated as a blue line

$$RMSE = \sqrt{mean((Forecast - observed)^2)}$$

The RMSE gives us information on how well a model can fit out observed data. The forecast errors are squared to ensure that errors that have opposite signs will not cancel. The mean of these errors are then computed over the 31 days and finally the square root is taken to give the RMSE skill of each model. As can be seen in figure 5 below, computing the absolute errors has shown an increase in variance as time increases.The large spread for the later days indicates a less predictable event. The plot of the absolute errors also verifies that the size of our error is larger where the temperature of the observed data is lower which occurs during the end dates.

1.3 Rank Histogram

The Rank Histogram answers the question of how well our ensemble spread represents the uncertainty of the observations. The 9 forecast models are put in order from smallest to largest. This results in 10 'bins' where the endpoints are 'less than the lowest' and 'greater than the largest'. The observed values for each day are then placed into one of these bins and the counts are displayed in the histogram. For a well behaved system, the observed values should act as one of the models and thus should have somewhat equal chance of being anywhere between any two models. Prior to bias removal this is clearly not the case as can be seen in figure 7.

This Rank Histogram displays clear negative bias as the majority of the observed t2m temperatures lie after the position of the largest value for any of our models. This has much room for improvement and will be brought back to use post systematic error removal.

1.4 ACC

The daily scores for the anomaly correlation consisted of +/- 1 values. This can be easily understood as: for an AC of 1 the forecast and the observed data's relationship to the climate are positively correlated and likewise, if the AC is -1 then they have a negative correlation, ie. the two errors have moved in

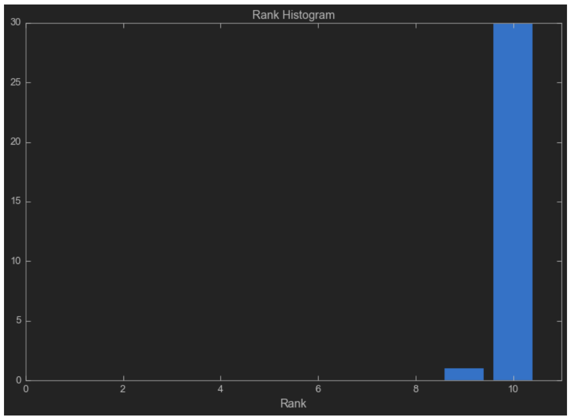

Figure 7. Rank Histogram for the Observed forecast

Table 3. Summary of results

	Bias	RMSE	ACC
ammc	-6.2147	6.65643	-0.229458
babj	-6.65202	7.24619	-0.16915
cwao	-6.99926	7.69662	-0.014912
dems	-4.98942	5.35103	0.023069
ecmwf	-3.02217	3.41782	0.152552
kwbc	-4.75929	5.07194	0.0233562
rjtd	-9.70075	10.0334	-0.150744
rksl	-5.59659	5.98185	-0.185568
vabb	-6.60325	6.94323	-0.106294
ensemble	-6.05972	6.30371	-0.0963579

'opposite directions' and where my forecasts have error w.r.t. climate values in the negative direction the observed data w.r.t. climate has errors in the positive direction. A ACC value of 1 gives indication towards a good forecast so the positive values are preferred. The average of these scores over the 31 days is then computed for each model giving the Anomaly correlation coefficient.

$$ACC = \frac{mean((f-c)(o-c))}{\sqrt{(mean((f-c)^2))(mean(o-c)^2))}}$$

This essentially works as a 'pattern' detection skill score between my models and the results from climatology. Larger ACC values indicate more accurate forecasts compared to climatology without taking into account the bias. Thus looking at the results below it seems to show that my models that had the largest bias still seem to be under-performing with respect to climatology. An ACC of less than 0.5 for an ensemble indicates a model that has an impractically useful skill.

1.5 Error Exploration

Throughout the calculation of the skill scores it has become clear that the system displays large systematic errors, predominantly systematic bias. Thus to reach the full potential of the model, systematic bias must be reduced. I have elected remove bias on a one-to-one basis as it seems to me that each model ammc, babj etc will have specific systematic errors associated to it. An alternative method of subtracting the average of the bias across all of the models was investigated however further analysis showed that this fared badly when compared to the removing on an individual basis.

The logic behind the existence of systematic bias has been

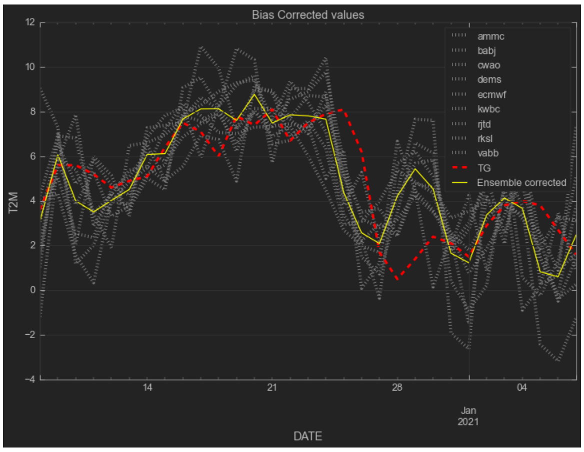

Figure 8. t2m temperatures of the bias corrected models

touched on in section 1.1 and my method is to investigate my errors and then adjust my model and repeat this investigation in the hopes that improvement can be seen. To begin, I subtracted the individual biases from their respective models. The result is seen in figure 8. As before the observed temperature values are shown in red, my ensemble in yellow and single models in the grey background. Nice! A large improvement is seen and now the Ensemble seems to follow the shape and trend of the observed data. The rank histogram however shows that this result might possibly be over-dispersive as there is that large peak in the middle meaning I might have modelled to closely to my observed data. when checking the new skill scores however, looking at the new summary results for the model without bias, there is now a much higher anomaly coefficient, now positive and closer to 1 which represents a good model, the RMSE scores have strictly decreased thus I can conclude a significant improvement has been made. Furthermore, when considering this improved forecast model over the December months the ACC score is approximately 0.85 which indicates an extremely skill full 'near perfect' forecast model and when compared to the January values where the ACC score is approximately 0.3 it is clear that the later month a model that doesn't have practically useful skill. From this my suspicions when eyeing the plots have been verified that as time goes on, my model has been less successful, irregardless of bias removal. This can also be clearly seen in figure 8 where the t2m temperatures as we enter the month of January have significant variance. It is possible that a weighted bias removal approach could be preferred specifically for the month of January, however this could lead to overfitting which is not ideal.

1.5.0.1

2 Climate Analysis

2.1 Historical

To investigate how well each climate model represents the climate for a city, it is important to take a look at how the historical predictions fared as that allows for an understanding of the uncertainty. However with such amounts of data it is difficult to analyse whether the models perform well by simply looking at day to day values for the temperatures or errors of the predictions vs observed as was done in section one. As can be seen in figure 10 there is a lot of information but this method

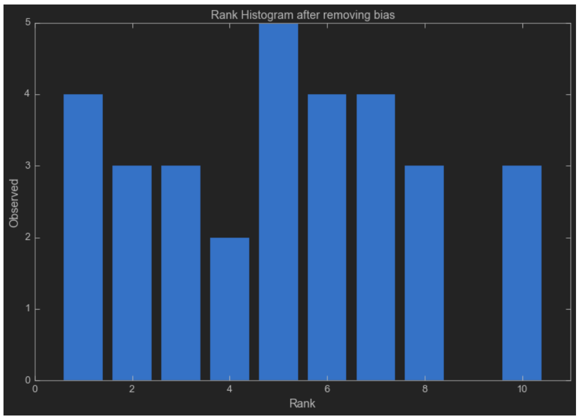

Figure 9. t2m temperatures of the bias corrected models

Table 4. Summary of results after bias removal

	Bias	RMSE	ACC
ammc	0.0	2.38444	-0.229458
babj	0.0	2.87366	-0.16915
cwao	0.0	3.2013	-0.014912
dems	0.0	1.93372	0.023069
ecmwf	0.0	1.59624	0.152552
kwbc	0.0	1.75321	0.0233562
rjtd	0.0	2.5621	-0.150744
rksl	0.0	2.11205	-0.185568
vabb	0.0	2.14606	-0.106294
ensemble	0.0	1.73683	0.727602

Table 5. Summary of results

	Histotical_CNRM	Historical_ICHEC	observed
count	9862.000000	9862.000000	9862.000000
mean	7.762021	7.598235	12.363648
std	8.631238	8.120338	7.016912
min	-22.052325	-17.176936	-6.800000
25%	1.050086	1.404845	6.300000
50%	7.606908	7.545782	12.100000
75%	14.816835	13.867827	18.400000
max	29.515009	27.376093	28.300000

of plot doesn't necessarily reveal much. Here, a much more useful tool is to consider a plot of the distribution densities as this can tell a lot about the behaviour of a model. From Figure 11 it can be seen that both of the historical models have predicted a lower temperature in general, with mean temperatures of 7.76°C (CNRM) and 7.59°C (ICHEC) compared to the true mean of 12.36°C. This can also be seen in the fact that the minimum temperatures for the historical CNRM and ICHEC are -22°C and -17°C respectively, whereas the true min temperature was only -6.8°C. As well as this the both historical models predicted that the temperatures will stray from the mean value more often which is evident in the fatter curve and larger standard deviation, which means they were ready for more extreme temperatures than actually occurred.

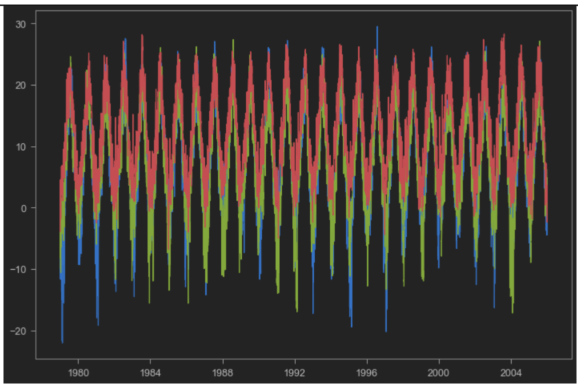

Figure 10. Historical data,

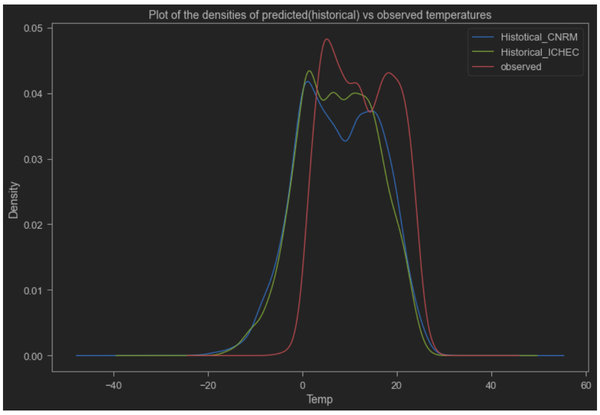

Figure 11. Historical densities vs. observed

2.2 Projections

The same method is employed for analysing the projected change in temperature for each climate model. These are two different possible futures for our climate with the 4.5 projections assuming we reduce our carbon emissions and the 8.5 assuming we do not. The projected change in temperature is expecting relatively large temperature deviation from the mean values predicted to be 9.88°C, 9.90°C, 11.0°C, 11.46°C. Thus in the absence of a reduction of our carbon emissions the expected mean temperature is around 1°C higher than if we were to reduce emissions.

This translates also to the minimum temperature values where the minimum temperature for reduced carbon emissions are near 1°C lower than those where we haven't.The ICHEC model predicts higher variation in temperature and a larger standard deviation for the 8.5 case but a lower variation in the case with reduced carbon emissions than the CNRM model. Thus the ICHEC model is expecting a more dramatic change for our climate if we do not reduce our carbon emissions.

Table 6. Description of Model Predicion data

	CNRM_4.5	ICHEC_4.5	CNRM_8.5	ICHEC_8.5
count	22280.000000	22280.000000	22280.000000	22280.000000
mean	9.881429	9.905543	11.002961	11.465478
std	8.013612	8.050062	7.815892	8.193635
min	-21.278004	-17.843136	-14.725018	-13.485579
25%	3.550122	3.382454	4.774521	4.896466
50%	9.235410	9.316551	10.352199	10.487144
75%	16.337056	16.163395	17.096353	17.893816
max	30.701471	30.033594	33.250635	33.514978

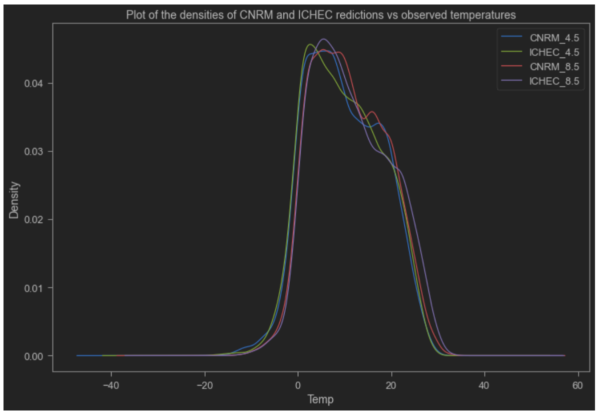

Figure 12. Plot of the densities of CNRM and ICHEC redictions vs observed temperatures

[]

Table 7. Forecast Errors

	ammc	babj	cwao	dems	ecmwf	kwbc	rjtd	rksl	vabb	Ensemble
2020-12-08	-5.584467	-1.205988	-11.795099	-6.472406	-2.890100	-4.262964	-12.267725	-6.134180	-8.516412	-6.569927
2020-12-09	-4.900201	-4.944940	-6.809991	-4.894220	-3.570703	-6.033655	-8.232751	-4.640497	-6.316248	-5.593690
2020-12-10	-4.994897	-11.069116	-4.735498	-6.750940	-6.079401	-8.797113	-13.251031	-6.430933	-6.802820	-7.656861
2020-12-11	-7.966327	-9.836932	-9.331317	-4.871967	-5.870135	-9.621539	-8.910663	-4.974139	-8.396533	-7.753284
2020-12-12	-6.530939	-7.275629	-7.250909	-6.274530	-3.082635	-5.381158	-9.166071	-5.475458	-9.306940	-6.638252
2020-12-13	-5.742804	-7.690375	-8.259039	-6.121985	-2.842963	-4.802374	-11.282843	-6.239966	-5.023260	-6.445068
2020-12-14	-5.766779	-4.899683	-7.082239	-4.881616	-1.916254	-4.372064	-7.528101	-4.171503	-5.086237	-5.078275
2020-12-15	-7.753607	-5.222296	-8.901373	-5.770819	-3.647070	-3.951788	-8.714514	-7.080908	-6.044287	-6.342963
2020-12-16	-5.885956	-5.968689	-5.428101	-4.560028	-3.649750	-3.899750	-9.659760	-7.553131	-6.400085	-5.889472
2020-12-17	-5.743677	-7.808374	-4.571619	-3.805078	-2.672540	-4.040613	-5.874750	-5.770380	-5.066705	-5.039304
2020-12-18	-4.159149	-4.410797	-4.796051	-2.085754	-1.130585	-3.691376	-5.749115	-4.855621	-4.511536	-3.932220
2020-12-19	-6.792645	-5.254498	-3.972028	-3.502087	-4.249127	-5.545972	-10.275525	-8.504376	-8.177075	-6.252593
2020-12-20	-5.575262	-4.435980	-4.081213	-2.967444	-2.410254	-4.852484	-8.124884	-3.825079	-5.868231	-4.682315
2020-12-21	-7.417841	-8.197534	-9.385187	-5.434900	-3.777307	-5.537286	-8.847406	-4.838403	-6.586237	-6.669122
2020-12-22	-5.556537	-6.639270	-4.660297	-2.331989	-2.261096	-2.904071	-8.693347	-5.400775	-5.666461	-4.901538
2020-12-23	-7.692810	-6.799744	-5.466736	-4.074097	-3.238007	-3.372101	-8.476135	-5.508942	-7.082245	-5.745646
2020-12-24	-9.625830	-9.539801	-4.488470	-5.789374	-2.630499	-2.888678	-8.939825	-7.141364	-5.613409	-6.295250
2020-12-25	-11.624963	-8.481317	-11.278223	-9.713373	-5.997247	-7.140436	-12.550592	-10.758630	-10.491052	-9.781759
2020-12-26	-10.409320	-12.768237	-7.929401	-8.972339	-5.729053	-7.763202	-14.414020	-8.706500	-10.527728	-9.691089
2020-12-27	-6.226123	-7.138507	-6.587482	-3.712146	-1.068500	-3.644275	-10.668292	-3.312335	-8.660510	-5.668686
2020-12-28	-2.491608	-0.487152	-3.077728	-3.058563	-1.004700	-2.095001	-6.705597	-0.726044	-1.420166	-2.340729
2020-12-29	-1.930334	-3.589240	-0.740363	-1.939337	0.353052	-1.911749	-4.502997	-2.142554	-1.780890	-2.020490
2020-12-30	-2.781226	-8.904425	-1.786597	-3.804236	-0.815405	-3.677527	-5.956732	-3.633826	-3.912115	-3.919121
2020-12-31	-6.557733	-6.329279	-11.017023	-5.081140	-1.886774	-4.274866	-11.395441	-5.725824	-6.250696	-6.502086
2021-01-01	-5.009155	-9.658417	-11.138947	-3.474762	-0.983032	-4.466461	-12.082367	-2.858276	-7.332581	-6.333777
2021-01-02	-4.076971	-5.293768	-4.000830	-2.878058	-1.892340	-4.457495	-12.386816	-7.628363	-7.649725	-5.584930
2021-01-03	-3.623395	-8.372327	-5.202527	-6.249188	-3.014447	-4.483929	-8.768018	-4.768018	-7.072003	-5.728206
2021-01-04	-9.646759	-4.173004	-7.241425	-3.430725	-3.641724	-5.556183	-14.591309	-4.870026	-4.411011	-6.395796
2021-01-05	-10.254865	-10.169659	-13.220654	-7.360608	-4.263043	-6.690503	-13.338574	-7.924939	-8.147260	-9.041123
2021-01-06	-7.580676	-8.706104	-12.881763	-8.076892	-4.828601	-4.169574	-8.945422	-8.472552	-9.791705	-8.161477
2021-01-07	-2.752832	-0.941675	-9.858881	-6.331293	-2.997064	-3.251855	-10.422510	-3.420709	-6.788446	-5.196141

Table 8. Forecast Bias

	ammc	babj	cwao	dems	ecmwf	kwbc	rjtd	rksl	vabb
Bias = average(Forecast - Observed	-6.2147	-6.652024	-6.999258	-4.989416	-3.022171	-4.759292	-9.700746	-5.596589	-6.603245

Table 9. RMSE

	ammc	babj	cwao	dems	ecmwf	kwbc	rjtd	rksl	vabb	RMSE ensemble
Mean Squared Error	6.656429	7.246194	7.696618	5.351033	3.41782	5.071943	10.033387	5.981851	6.943227	6.303707