BELGINGUR

Technical report

Ólafur Rögnvaldsson
Heitor Chang
Karolina Stanisławska
Mikołaj Okrzesa

Comparisons of observed and simulated weather

Description of three dynamical downscaling experiments and comparison with observations

Reykjavík
March 2025

Samantekt

Í þessari greinargerð berum við reikniniðurstöður þriggja niðurkvörðunarraða fyrir Ísland saman við veðurmælingar. Reiknirnar nefnast Carra, Icebox og RÁV2 og eru unnar hver með sínu lagi, sem ekki verður lýst nánar hér.

Í örstuttu máli þá eru niðurstöður úr Carra og Icebox nokkuð keimlíkar. Vindhraði og hiti eru t.d. á pari þótt það sjáist vissulega munur milli einstakra mælistöðva. Hermd langbylgjugeislun er að koma betur út í Carra meðan stuttbylgjugeislun lítur betur út í Icebox. Hermanir á hita, vindi og geislun í Carra og Icebox ber mun betur saman við mælingar en hermd gildi úr RÁV2. Úrkoman er sú breyta þar sem munurinn er mestur, en Icebox er að ná úkomuákefðinni betur en Carra og RÁV2. Þarna verður þó að slá ákveðinn varnagla í túlkun þar sem við höfum bara 3 klst. tímaupplausn í Carra (þ.e. "mm/klst" er í raun þriggja klukkustunda meðaltal). Til að flækja samanburðinn enn frekar þá er reiknuð (Carra) úrkoma borin saman við mælda úrkomu á gildistíma útreikninganna (00Z, 03Z, ..., 18Z, 21Z). Þegar úrkoma er uppreiknuð í sólarhrings-, viku- og mánaðarsummur þá er afar lítinn mun að sjá á milli Carra og Icebox, en bæði líkön ofmeta úrkomuna lítillega m.v. mælingar. Ofmatið er að jafnaði 9% í Carra og 14% í Icebox. Samanburður milli mánaða bendir til að Carra vanmeti vetrarúrkomu en ofmeti sumarúrkomu m.v. hefðbundnar úrkomumælingar. Icebox hinsvegar ofmetur úrkomuna óháð árstíð. Þetta ofmat gæti hugsanlega útskýrt villutopp í reiknuðum hita sem kemur fram í maí og júní í Icebox röðinni. Þ.e. of mikil vetrarúrkoma leiðir af sér of mikil snjóalög sem ná of langt inn í sumarið og valda óeðlilegri kælingu. Uppsöfnuð úrkomugildi úr RÁV2 röðinni eru í góðu samræmi við mælingar þrátt fyrir að klukkustundar gildi vanmeti úrkomu þegar ákefð fer yfir 3.5-4 mm/klst.

Samanburður við ákomureikninga af jöklum bendir til að Icebox og Carra ofmeti vetrarúrkomu á jöklum nokkuð meira en gamla RÁV2 röðin. Þarna er vissulega freistandi að benda á að bæði Carra og Icebox eru þvinguð af ERA5 meðan RÁV2 notaðist við gögn úr eldri ERA-Interim endurgreiningunni. Hegðunarmynstri reiknaðrar úrkomu á jöklunum þremur (Hofsjökli, Langjökli og Vatnajökli) ber vel saman við mælda vetrarákomu. Frávikið frá þessu eru þó veturnir 1996-97 og 2018-19 þegar öll þrjú líkönin sýna töluverða aukningu m.v. veturinn á undan meðan mælingar (einkum á Hofsjökli og Vatnajökli) sýna minni ákomu. Ekki er vitað hvað veldur þessum frávikum.

Um mælingarnar er það að segja að úrkoman hefur verið til ákveðinna vandræða. En það var vitað fyrir að úrkomugögn síðustu þriggja til fjögurra ára eru óyfirfarin og bera gagnaraðirnar þess augljós merki. Við höfum eytt töluverðu af augljósum villum út úr samanburðinum en það má gera ráð fyrir að ennþá leynist e-ð af skekkjum inni. Við höfum líka rekist á villur í hita- og vindhraðamælingum sem við höfum hent út. Loks höfum við hreinsað út nokkurn fjölda geislunarmælinga, bæði fyrir stuttbylgju og langbylgju.

Introduction

In this report we compare the results of three re-analysis data sets, named Carra, Icebox, and RÁV2, to observations of temperature, wind, precipitation and radiation fluxes for the twenty-year period from 1 September 1999 to 31 August 2019, and to a less extent to 31 August 2024. To aid with this comparison a graphical tool, built on top of the Verif [1] solution, has been created. This tool can be used to browse and visualize verification results from any atmospheric simulation, as long as the data have been converted to the WOD standardized netCDF file format (for information on the WOD framework see [2]).

In addition, we compare observations of accumulated wintertime precipitation on three large ice caps (Hofsjökull, Langjökull and Vatnajökull) to values from the Carra, Icebox and RÁV2 simulations from the early nineties to the winter of 2023-24, except for RÁV2 which only extends to the winter of 2018-19.

Results of dynamical downscaling

The idea behind dynamical downscaling is relatively simple. Take output from a coarse resolution model, e.g. a Global Circulation Model (GCM), and use it to force a Limited Area Model (LAM) at a higher horizontal and vertical resolution. As resolution is increased, processes governed by the interaction of the large-scale flow and topography become better resolved by the models [3]. Here we present results from three different dynamical downscaling simulations, named Carra, Icebox and RÁV2.

The Belgingur-Carra archive is a sub-set of the C3S Arctic Regional Reanalysis (CARRA) dataset, containing three-hourly short-term forecasts of various surface meteorological variables at 2.5 km horizontal resolution. As of February 2025, the data span from January 1991 to November 2024. For more information on the full Carra dataset see [4].

Within the framework of the Icebox[1] research project, initially led by Statnett in Norway, a one-hourly data set of weather in Iceland has been created. As of February 2025, the data span from September 1990 to October 2024 and are created by dynamically downscaling the ERA5 re-analysis data using V4.1.2 of the WRF-Chem atmospheric model, run at 2 km horizontal resolution and with output written with one hour temporal resolution.

RÁV2 is a one-hourly data set of weather in Iceland ranging from September 1979 to August 2019. It was created by dynamically downscaling the ERA5-Interim re-analysis data using V3.6.1 of the AR-WRF atmospheric model, run at 2 km horizontal resolution. For more information on RÁV2 see [5].

[1] https://www.statnett.no/en/about-statnett/innovation-and-technology-development/our-prioritised-projects/icebox/

A summary of the configuration of the three models is given in Table 1.

Table 1: Summary of model configurations.

MODEL ABBREVIATION / VERSION	RESO-LUTION [KM] / # LEVELS	IC/BC DATA	SHLLOW CONV. SCHEME	PBL SCHEME	MICRO-PHYSICS	LW RAD SCHEME	SW RAD SCHEME	SFC. LAYER	LAND SFC.
CARRA (HARMONIE-AROME CY40)	2.5 / 65	ERA5 + various obs.	EDMFm	HARATU	ICE3	RRTM	Morcrette	SURFEX	SURFEX
RÁV2 (WRF V3.6.1)	2 / 65	ERA-Interim	N/A	MYNN	Morrison	RRTMG	RRTMG	MYNN	NOAH MP
ICEBOX (WRF V4.1.2)	2 / 51	ERA5 + ERA5-Land	N/A	MYNN	Thompson aerosol aware	RRTMG	RRTMG	MYNN	NOAH

The three re-analysis simulations are compared to observations of temperature and wind speed from 55 surface station (cf. Figure 1).

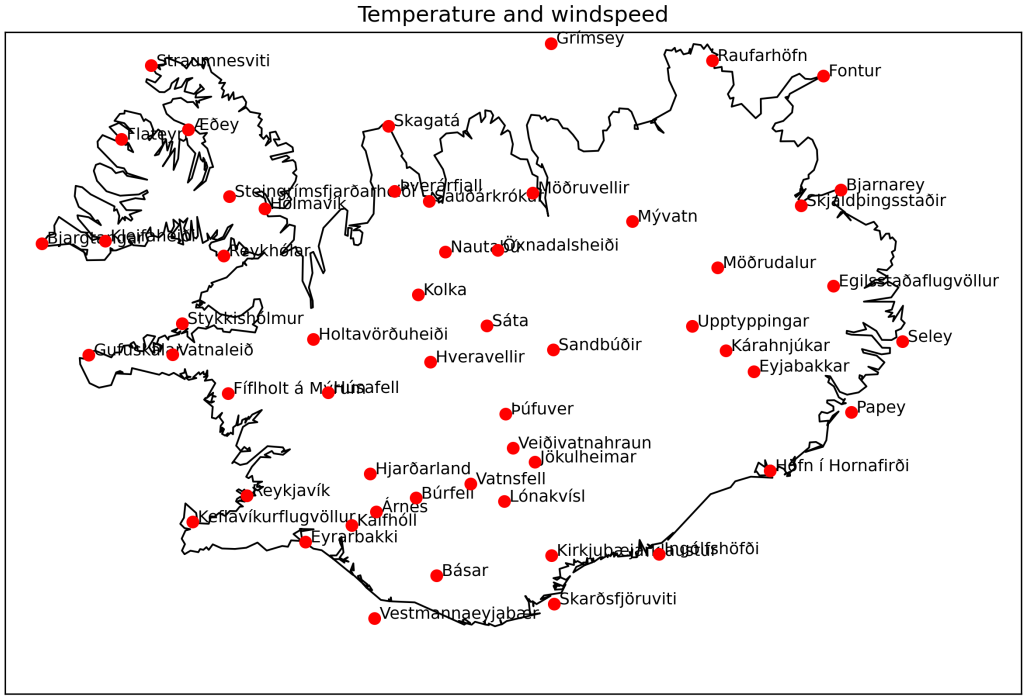


Figure 1: Overview map showing location of the 55 stations used for comparing observed and simulated values of near surface temperature and wind speed.

Simulated precipitation is compared to observations from 52 stations (cf. Figure 2). It is well known that precipitation is notoriously difficult to observe in strong winds and cold weather [6]. To emphasise events where we could expect precipitation observations to be of reasonably good quality, we only investigate cases where simulated (from the Icebox series) wind speed is 8 m/s or less and simulated temperature is above 2°C at the grid cell representing individual observation sites.

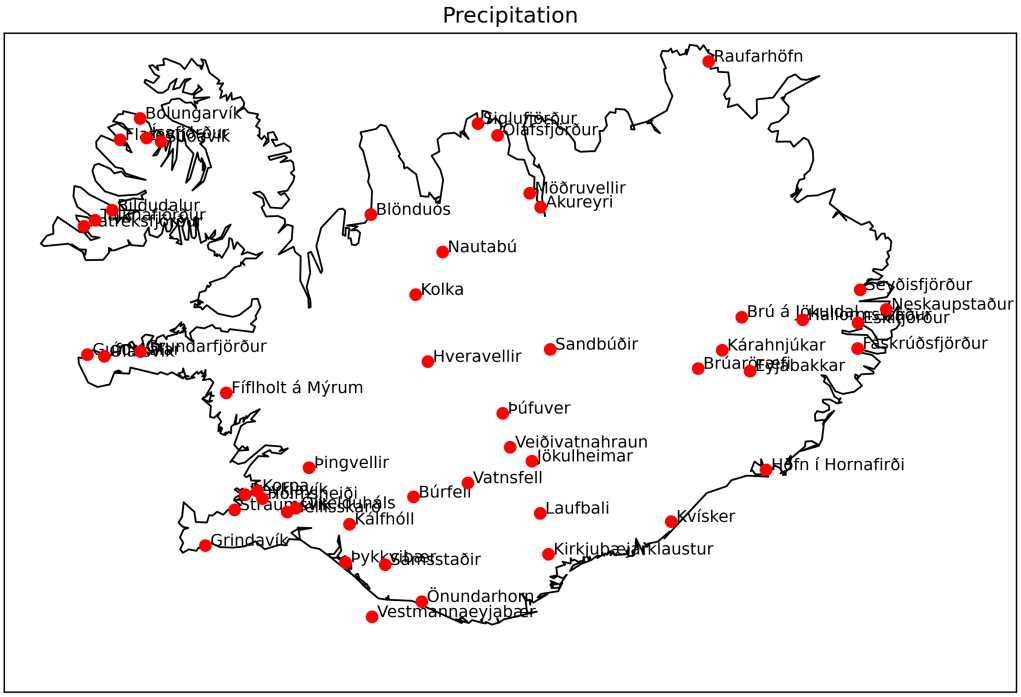


Figure 2: Overview map showing location of the 52 stations used for comparing observed and simulated values of precipitation. We only use observed values at times when simulated wind speed is below 8 m/s and simulated temperature is above 2°C in the grid cell representing the observation site.

To get an estimate of wintertime precipitation over larger areas we compare observed accumulated winter precipitation on chosen icecaps to modelled precipitation for the same regions (cf. Figure 3).

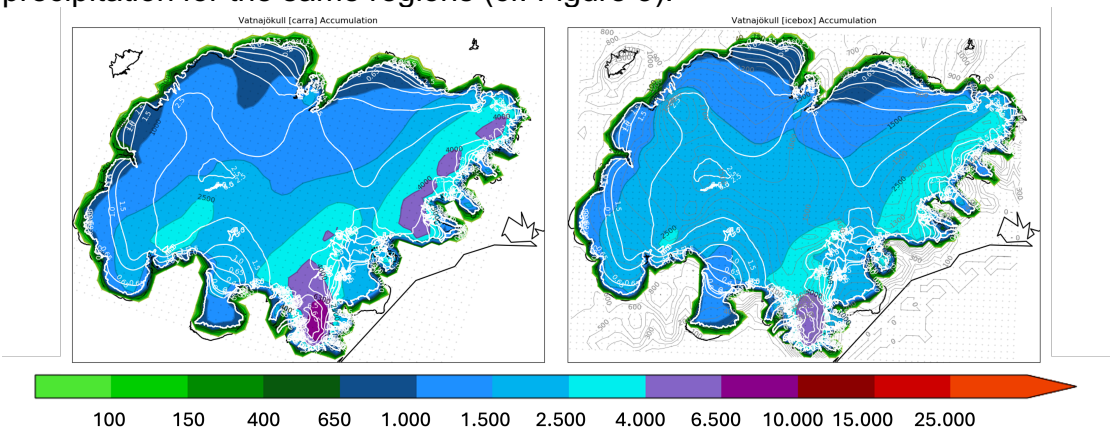


Figure 3: Example of inter-comparison of accumulated winter precipitation (1.464 mm) for Vatnajökull icecap for the winter of 2022-23. White isolines indicate observed accumulation [m] whilst the colour scale [mm] shows the accumulated wintertime precipitation simulated by the Carra (left, 1.984 mm) and Icebox (right, 1.834 mm) models.

Finally, simulations of short- and long wave radiation are compared to observations from 27 and 17 stations, respectively (cf. Figure 4).

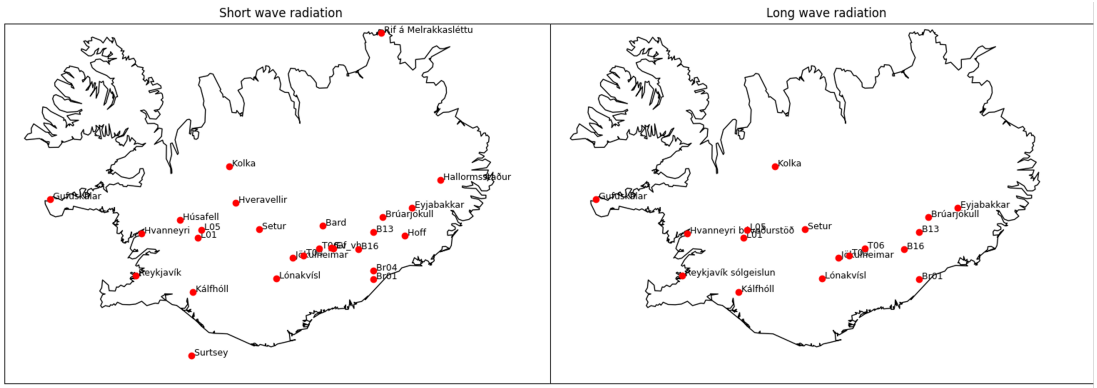


Figure 4: Overview map showing location of the 27 stations used for comparing observed and simulated values of short-wave radiation (left) and 17 stations used for comparing observed and simulated values of long wave radiation (right).

Information from this model vs. observation comparison can be visualised using the Verif web service (https://verif.belgingur.is) that has been developed for this project, and is described in more detail in Appendix A. Currently, users can compare modelled data from a twenty-five year period (1 September 1999 to 31 August 2024), from both Carra and Icebox, to observations. Data from the RÁV2 series is also available for a shorter, twenty-year, period (1 September 1999 to 31 August 2019). The system offers three kinds of plots; scatter diagrams, Taylor diagrams, and Quantile-Quantile plots, and four types of maps showing model Bias, Multiplicative Bias, Root Mean Square Error and Mean Absolute Error. For each location the following statistical parameters are also calculated:

- Root Mean Square Error (RMSE)
- Mean Absolute Error (MAE)
- MBias (Multiplicative Bias)
- Bias
- Spearman's Rank correlation coefficient[2]
- Pearson correlation coefficient[3]
- Standard deviation of observations
- Standard deviation of modelled parameter

These parameters can be viewed as a table on the website or downloaded as a CSV file.

General discussions

Temperature (55 stations)

The Carra and Icebox temperature simulations show a similar pattern. Simulated values from both models are strongly correlated with observations (Pearson correlation is 0.95 and 0.96, respectively) and both have a standard deviation of 5.5°, which is very close to the standard deviation of the observations, which is 5.6°C. The RÁV2 data compare less well with observations. The Pearson correlation is 0.84 and the standard deviation is 5.3°C (cf. Figure 5, top right panel). The Root Mean Square Error (RMSE) of

[2] https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient
[3] https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

both Carra and Icebox is 1.9°C, but 3.4°C for RÁV2. All models capture the distribution well (cf. Figure 5, top left panel) with the Icebox model capturing

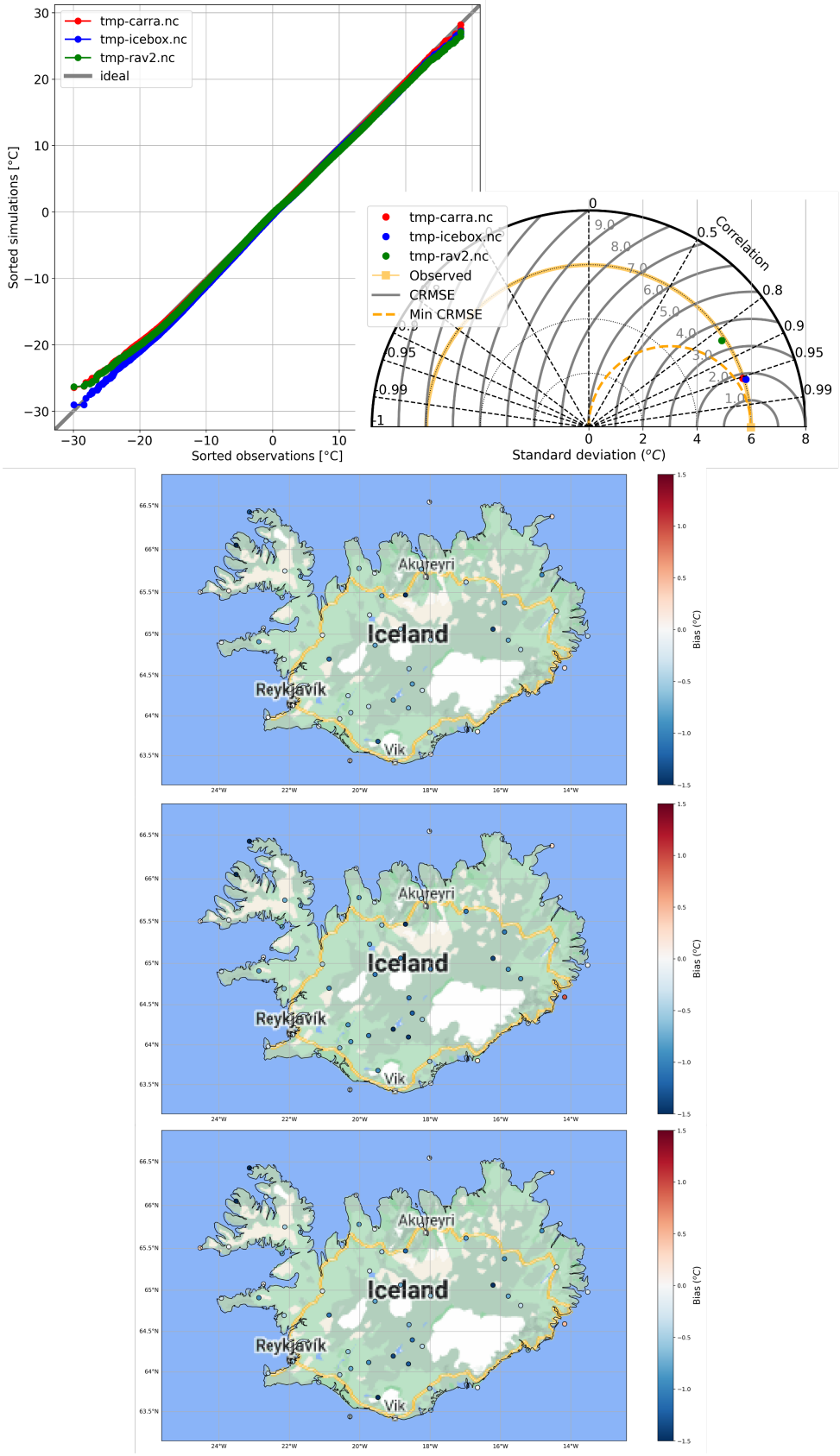


Figure 5: Quantile-Quantile plot (top left), Taylor diagram (top right), and bias of simulated temperature [°C] at 2 metres above ground level. Results from Carra (top), Icebox (middle) and RÁV2 (bottom) simulations between 1 September 1999 and 31 August 2019. See text for details.

the coldest outliers slightly better than Carra and RÁV2. The models' biases show similar pattern (cf. Figure 5, maps), near the coast the bias is close to zero but slightly negative in the interior and highlands.

On average all models have slight negative bias, ranging from -0.5°C for Carra to -0.7°C for Icebox and -0.6°C RÁV2. Mean absolute errors for Carra and Icebox are similar (1.4°C for Carra and Icebox) but RÁV2 shows more than a one-degree greater error (2.5°C). See table 2 for summary of statistical comparison of the models.

Intra-annual variability of RMSE, MAE, Bias, and Pearson correlation for temperature is shown in Figure 6. There are some similarities in the variability between the models as well as noticeable discrepancies. For both Carra and Icebox the correlation is greatest over the winter months, whilst this signal is in general reversed (there is a "dip" in June, July and August) for RÁV2. Icebox and RÁV2 show a large minimum in late spring/early summer in the temperature bias whilst Carra shows much less seasonal variability. The variability of RMSE and MAE is nearly

Figure 6: Monthly values of RMSE (top), MAE (second from top), Bias (second from bottom) and Pearson correlation (bottom) for temperature for the period 1 September 1999 to 31 August 2019.

identical for each model, albeit quite different between models. Again, Carra shows the least variability of the three datasets; winter, spring and early summer having a near constant RMSE and MAE, which then decreases gradually, reaching a minimum in October when it starts increasing again, reaching the near constant value in December. RÁV2 has a maximum in December that then gradually decreases until August when it starts increasing

again. Icebox has near identical values to that of Carra for winter and early spring months but peaks in May and June and then drops below the Carra values, reaching a minimum in September when it starts rising again.

Table 2: Summary of statistical comparison for the 20-year period 1 September 1999 to 31 August 2019, between observed and modelled two-meter temperature. Values in parenthesis are for the 25-year period 1 September 1999 to 31 August 2024. Only the standard deviation is applicable to the observations (last column).

Temperature	Carra	Icebox	RÁV2	Obs
RMSE	1,9 (1,9)	1,9 (1,9)	3,4	N/A
MAE	1,4 (1,4)	1,4 (1,4)	2,5	N/A
Bias	-0,5 (-0,5)	-0,7 (-0,6)	-0,6	N/A
Spearman Corr	0,96 (0,96)	0,96 (0,96)	0,84	N/A
Pearson Corr	0,95 (0,95)	0,96 (0,96)	0,82	N/A
Standard dev	5,5 (5,5)	5,5 (5,6)	5,3	5,6 (5,6)

Wind speed (55 stations)

Comparisons of observed and simulated near-surface winds reveal that the Carra and Icebox models are, as with temperature, showing quantitatively similar results, while the RÁV2 data is of lower quality.

Simulated values from Carra and Icebox show (Pearson) correlation with observations of 0.73 (Carra) and 0.77 (Icebox) and a standard deviation of 4.1 m/s and 3.9 m/s, respectively, which is on par with observed value of 4.1 m/s (cf. Figure 6, top right panel). The Root Mean Square Error (RMSE) is 3.2 m/s for Carra and 3.0 m/s for Icebox. The Pearson correlation of RÁV2 is however only 0.27, standard deviation is 4.4 m/s and the RMSE is considerably greater, or 5.4 m/s. Both the Carra and Icebox models capture the wind speed distribution well (cf. Figure 7, top left panel) up to around 18 m/s (Icebox) and 22 m/s (Carra). At this range the RÁV2 model is overshooting compared to observations. At higher wind speeds both Carra and Icebox tend to underestimate the strength of the wind. This underestimation becomes quite apparent around 36 m/s and above. Interestingly, the RÁV2 model is doing considerably better at these high winds. The models' biases show similar pattern (cf. Figure 7, maps), near the coast the bias tends to be positive, whilst it is slightly negative in the interior and highlands. This, in relation with a general cold temperature bias at higher altitudes, may be an indication of that the models tend to underestimate near-surface mixing in the interior of Iceland. A potential source of this could be too high surface roughness in the models for the area.

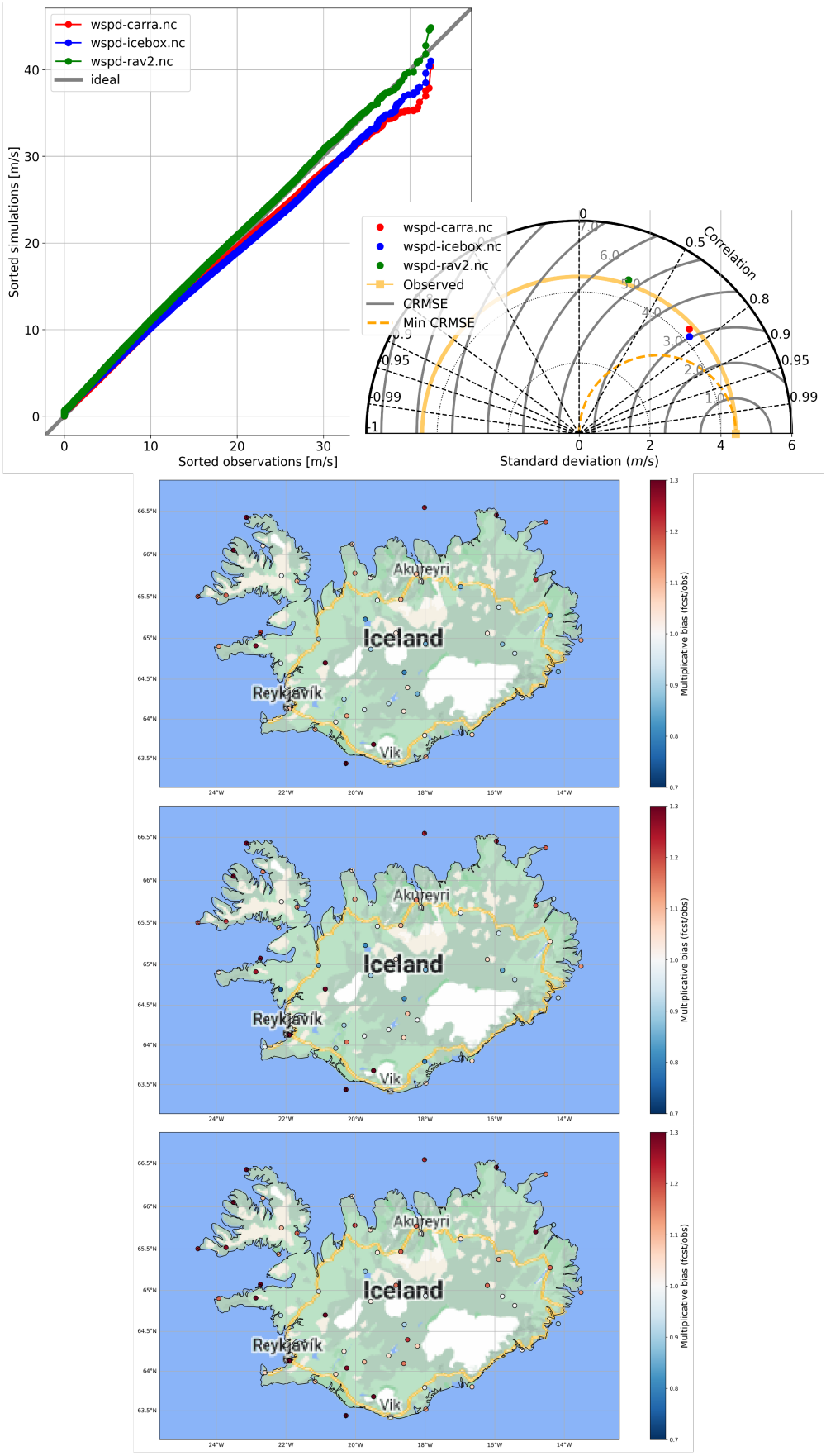


Figure 7: Quantile-Quantile plot (top left), Taylor diagram (top right), and multiplicative bias of simulated wind speed at 10 metres above ground level. Results from Carra (top), Icebox (middle) and RÁV2 (bottom) simulations between 1 September 1999 and 31 August 2019. See text for details.

Out of the three data sets the RÁV2 one matches the ideal Q-Q line the most closely despite having the lowest statistical score in general. Figure 8 shows the frequency distribution of observed and simulated wind speed. The RÁV2 data deviate the most from observations so it remains a mystery as to why the Q-Q fit is so good. A potential reason could be that the number of observations exceeding ~25 m/s is so low, in comparison to slower wind speeds, that the statistical impact of said cases is miniscule with regards to the big picture.

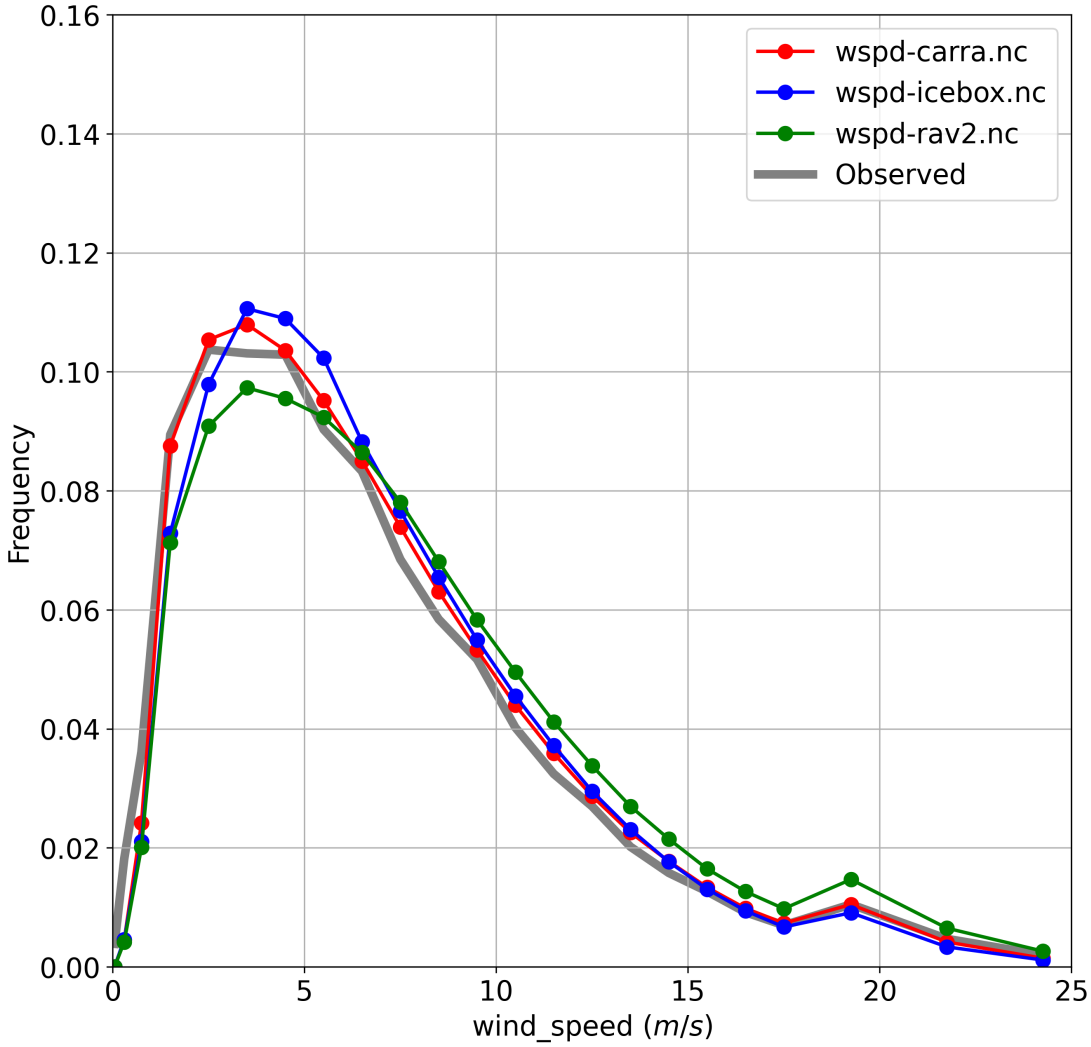
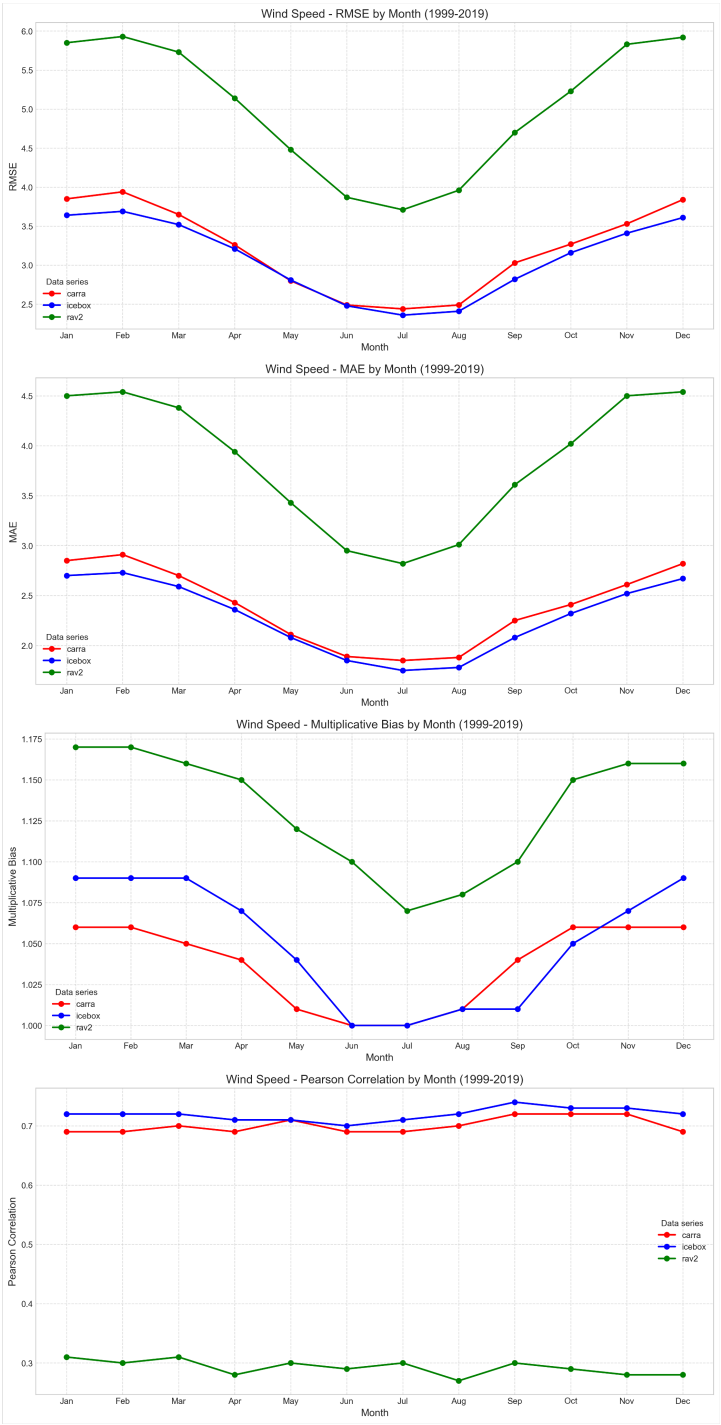


Figure 8: Frequency distribution of observed and simulated wind speed between 1 September 1999 and 31 August 2019. See text for details.

All models have slight positive bias, ranging from 0.3 m/s for Carra to 0.4 m/s for Icebox and 0.9 m/s for RÁV2. Mean absolute errors for Carra and Icebox are similar (2.4 m/s for Carra and 2.3 m/s for Icebox) but RÁV2 shows more than one and a half m/s greater error (4.0 m/s). See Table 3 for a summary of statistical comparison of the three models.



Two locations stand out with respect to positive model bias, high RMSE and a much larger standard deviation than observed. These stations are in Vestmannaeyjabær and Básar á Goðalandi, located in the southern region of Iceland. Why winds at these two locations are so poorly represented in the models can, at least to some extent, be linked to unresolved topography in the vicinity of the stations and, in the case of Básar, underestimation of surface roughness, but the area around the weather station is heavily vegetated and sheltered by local topography. Intra-annual variability of RMSE, MAE, Bias, and Pearson correlation for wind speed is shown in Figure 9. There is a distinct seasonality in RMSE, MAE and the MBias in all three models where the lowest values are during the summer months and maximum values during winter. RMSE and MAE values of Carra and Icebox are

Figure 9: Monthly values of RMSE (top), MAE (second from top), Bias (second from bottom) and Pearson correlation (bottom) for wind speed for the period 1 September 1999 to 31 August 2019.

Table 3: Summary of statistical comparison for the 20-year period 1 September 1999 to 31 August 2019, between observed and modelled near surface wind speed. Values in parenthesis are for the 25-year period 1 September 1999 to 31 August 2024. Only the standard deviation is applicable to the observations (last column).

Wind speed	Carra	Icebox	RÁV2	Obs
RMSE	3,2 (3,2)	3,1 (3,1)	5,4	N/A
MAE	2,4 (2,4)	2,3 (2,3)	4,0	N/A
Bias	0,3 (0,3)	0,4 (0,4)	0,9	N/A
Spearman Corr	0,71 (0,71)	0,74 (0,73)	0,26	N/A
Pearson Corr	0,73 (0,74)	0,77 (0,77)	0,27	N/A
Standard dev	4,1 (4,2)	4,0 (4,0)	4,4	4,1 (4,2)

very similar but RÁV2 shows 2 m/s (summer) to 3 m/s (winter) larger values. The same pattern is seen for the multiplicative bias. There is not much seasonality seen in the Pearson correlation, Icebox and Carra have a correlation slightly exceeding 0.7 whilst RÁV2 has a much lower correlation of, or just below, 0.3.

Precipitation (52 stations)

To minimize the effects of strong winds and low temperature on the quality of the precipitation observations we only investigate cases where simulated (from the Icebox series) wind speed is 8 m/s or less and simulated temperature is above 2°C at the grid cell representing individual observation sites. By this we limit greatly the cases when precipitation is under-observed due to strong winds and/or snowy conditions.

Carra and Icebox have Pearson correlation values of 0.45 and 0.41, respectively and RMSE of 0.39 and 0.43 mm/hr. The standard deviation of Carra is 0.32 mm/hr but that of Icebox is 0.42 mm/hr, observed value is 0.41 mm/hr (cf. Figure 10, bottom right panel). For RÁV2 these values are 0.06 (Pearson correlation), 0.56 mm/hr (RMSE) and 0.34 mm/hr (standard deviation). Carra and Icebox show a positive multiplicative bias (MBias) of 1.14 and 1.22, respectively, whilst RÁV2 has a MBias of 0.9. The mean absolute error (MAE) of Carra and Icebox is the same, 0.13 mm/hr, but RÁV2 has MAE of 0.18 mm/hr. See Table 4 for a statistical summary.

Both Carra and Icebox show in general a slight positive bias in simulated precipitation. The exception being two stations in SW-Iceland, Hellisskarð and Ölkelduháls and one station in SE-Iceland, Kvísker. The distance between Hellisskarð and Ölkelduháls is only about six kilometres, and both stations are located in relatively complex topography upstream of southeasterly winds coming in from the ocean. The Kvísker station is located at the foot of Mt. Öræfajökull, also upstream of southeasterly winds coming in from the ocean. Hence it should be expected that orographic lifting in combination with local topography (unresolved at 2-3 km horizontal resolution) could lead to general underestimation of simulated precipitation.

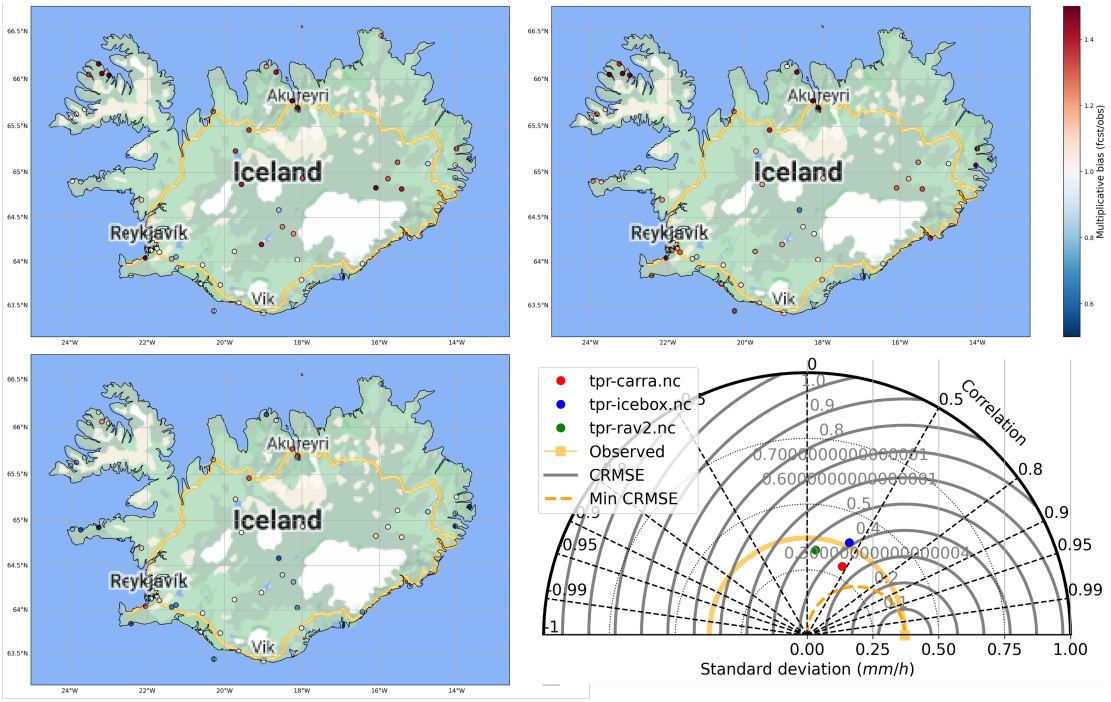


Figure 10: Multiplicative bias maps and Taylor diagram (bottom right) of simulated precipitation rate. Results from Carra (top left), Icebox (top right) and RÁV2 (bottom left) simulations between 1 September 1999 and 31 August 2019. See text for details.

Looking at quantile-quantile plots of hourly to monthly aggregated values of precipitation, it is seen that the Carra simulations start to underestimate the precipitation around a rate of 1.5-2 mm/hr and RÁV2 around 3-3.5 mm/hr whilst Icebox shows no signs of underestimation (cf. Figure 11, top left). However, when the data is aggregated in to daily, weekly and monthly values, this underestimation is no longer visible (cf. Figure 11, top right and bottom panels). On the contrary, both Carra and Icebox are seen to overestimate the precipitation, which agrees with the positive bias found in these (cf. Figure 10). Summary of statistical comparison of the three models is given in Table 4.

Table 4: Summary of statistical comparison for the 20-year period 1 September 1999 to 31 August 2019, between observed and modelled hourly precipitation rate. Values in parenthesis are for the 25-year period 1 September 1999 to 31 August 2024. Only the standard deviation is applicable to the observations (last column).

Precip rate	Carra	Icebox	RÁV2	Obs
RMSE	0,39 (0,40)	0,43 (0,45)	0,56	N/A
MAE	0,13 (0,13)	0,13 (0,14)	0,18	N/A
MBias	1,14 (1,09)	1,22 (1,14)	0,9	N/A
Spearman Corr	0,44 (0,42)	0,44 (0,41)	0,1	N/A
Pearson Corr	0,45 (0,43)	0,41 (0,39)	0,06	N/A
Standard dev	0,32 (0,32)	0,42 (0,41)	0,34	0,40 (0,41)

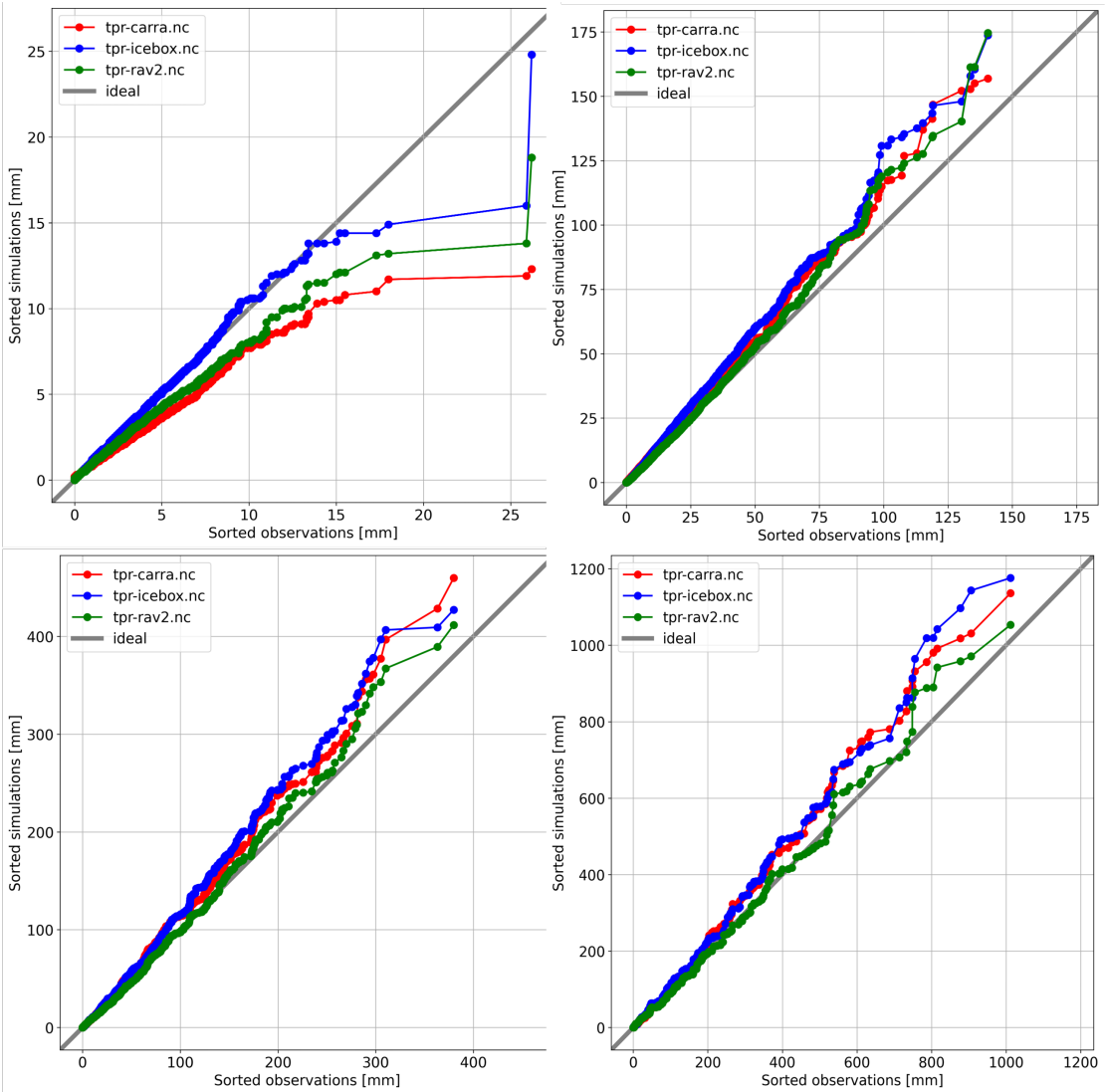


Figure 11: Quantile-Quantile plots of hourly (top left), daily aggregated (top right), weekly aggregated (bottom left) and monthly aggregated (bottom right) simulated precipitation [mm]. Results from Carra (red line), Icebox (blue line) and RÁV2 (green line) simulations between 1 September 1999 and 31 August 2019. See text for details.

Intra-annual variability of RMSE, MAE, Bias, and Spearman rank correlation for precipitation rate is shown in Figure 12. All three model show strong seasonality in RMSE and MAE, with a clear minimum in the summer and maximum during winter. The Carra data set has the lowest RMSE in general and RÁV2 the highest. Carra and Icebox have near identical MAE that are lower than the RÁV2 values. The multiplicative bias of Carra indicates the model is underestimating wintertime precipitation but overestimating the summertime precipitation. Icebox shows overestimation all year around with much less seasonality, although there are signs of spring and autumn maxima. The RÁV2 data set is underestimating the precipitation rate with the exception in March and December. The Spearman rank correlation of RÁV2 is quite low, or less than 0.15, and shows little seasonality. The correlation of Carra and Icebox is similar with highest values in the winter (between 0.4 and 0.5) and minima during summer (between 0.35 and 0.4).

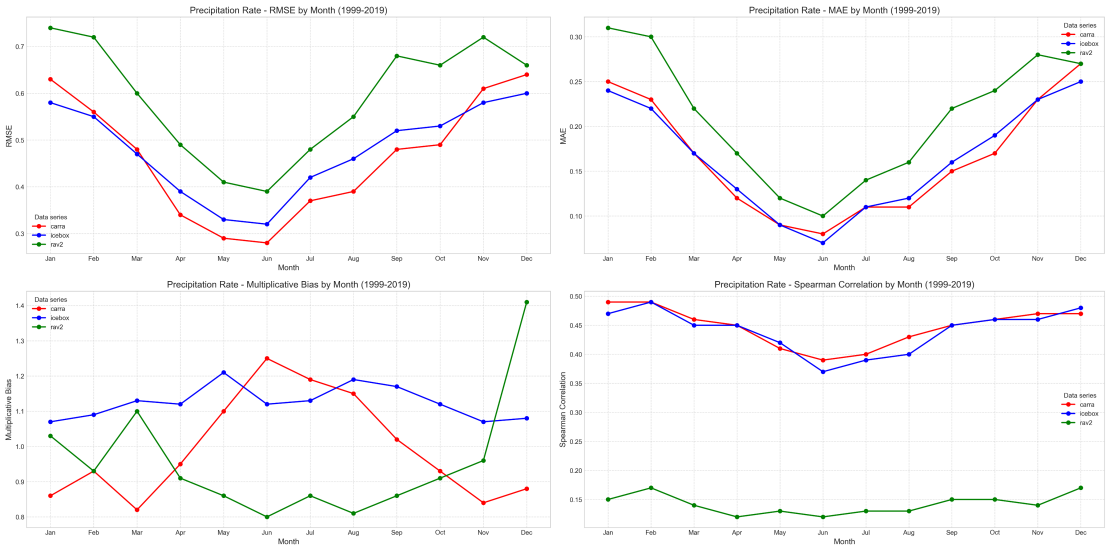


Figure 12: Monthly values of RMSE (top left), MAE (top right), MBias (bottom left) and Spearman rank correlation (bottom right) for precipitation rate for the period 1 September 1999 to 31 August 2019.

The number of observations varies considerably within the year, cf. Figure 13.

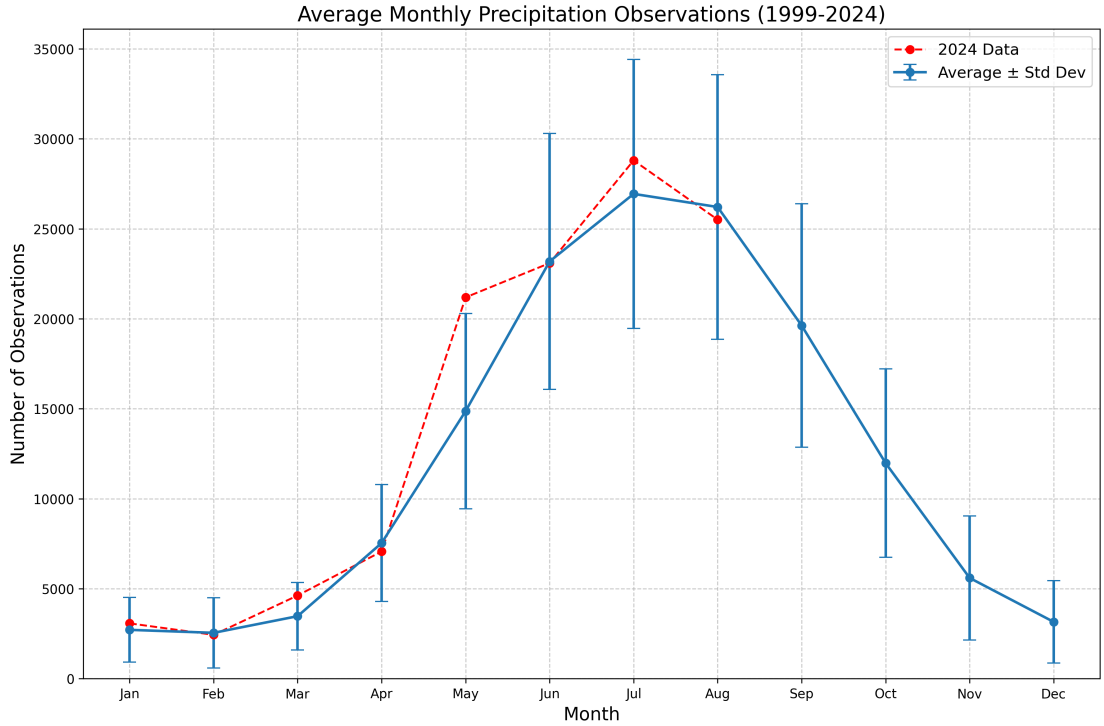


Figure 13: Average (blue line) number of monthly precipitation observations between September 1999 and August 2024, as used in this study. Red line shows data for January – August 2024 and vertical bars show the standard deviation.

On average there is a factor of ten difference between the number of available observations during winter and summer months. This should be kept in mind when interpreting the intra-annual variability shown in Figure 12.

Accumulated wintertime precipitation on icecaps

To get an estimate of wintertime precipitation over larger areas we compare observed accumulated winter precipitation on chosen icecaps (Hofsjökull, Langjökull, and Vatnajökull) to modelled precipitation for the same regions.

For Hofsjökull, the comparison period is the winters of 1991-92 to 2023-24 (Carra, 33 years), 1990-91 to 2023-24 (Icebox, 34 years), and 1988-89 to 2018-19 (RÁV2, 31 year). For Langjökull, the comparison period is the winters of 1996-97 to 2023-24 (Carra and Icebox, 28 years), and 1996-97 to 2018-19 (RÁV2, 23 years). For Vatnajökull, the comparison period is the winters of 1991-92 to 2023-24 (Carra and Icebox, 33 years), and 1991-92 to 2018-19 (RÁV2, 28 years). This comparison is shown in Figures 14 to 16 with statistical summary given in Tables 5 to 7.

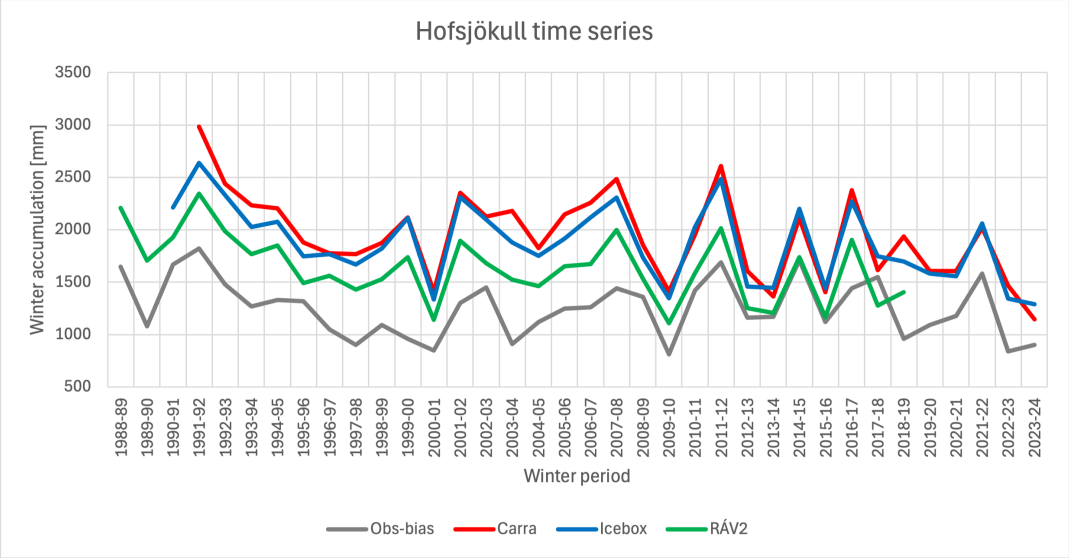


Figure 14: Modelled and observed wintertime accumulation of snow on Hofsjökull icecap in central Iceland. Observed values are taken from https://islenskirjoklar.is, accessed on 15 February 2025. Note that observed values have been bias-corrected by 300mm (Tómas Jóhannesson, personal communications).

Table 5: Summary of statistical comparison for the 28-year period 1991-92 winter to the 2018-19 winter between observed and modelled accumulated water equivalent of snow. Values in parenthesis are for 33-year (winter of 1991-92 to 2023-24 for Carra), 34-year (winter of 1990-91 to 2023-24 for Icebox) and 31-year (winter of 1988-89 to 2018-19 for RÁV2) periods. Only the standard deviation is applicable to the observations (last column).

Hofsjökull	Carra	Icebox	RÁV2	Obs
RMSE	812 (770)	703 (668)	414 (424)	N/A
MAE	754 (708)	663 (628)	367 (378)	N/A
Bias	754 (708)	663 (628)	348 (361)	N/A
Spearman Corr	0,59 (0,64)	0,70 (0,74)	0,68 (0,70)	N/A
Pearson Corr	0,64 (0,67)	0,73 (0,77)	0,68 (0,71)	N/A
Standard dev	389 (407)	342 (355)	298 (308)	261 (274)

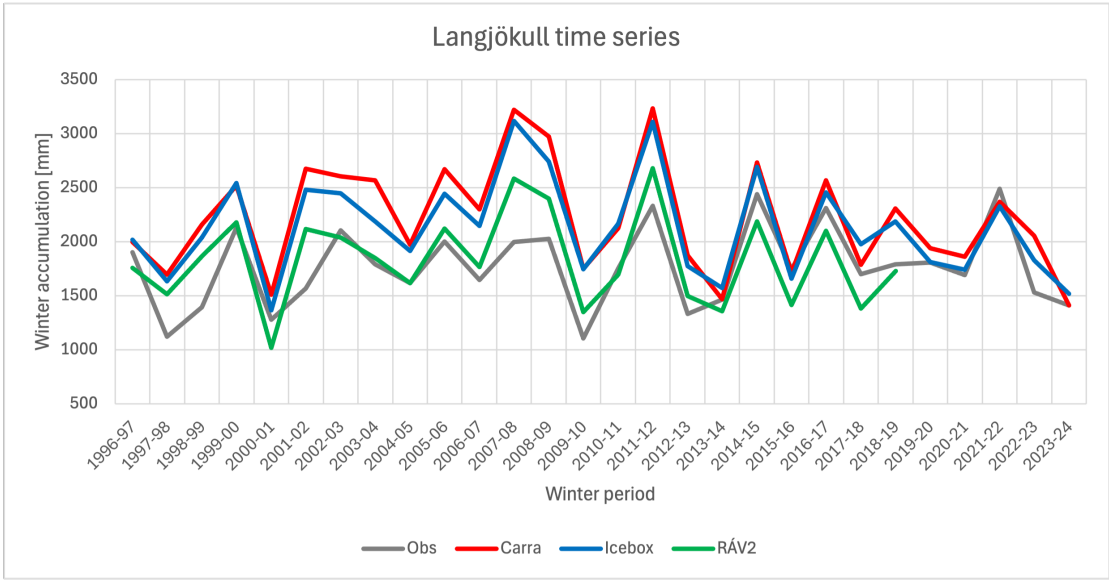


Figure 15: Modelled and observed wintertime accumulation of snow on Langjökull icecap in central Iceland. Observed values are taken from https://islenskirjoklar.is, accessed on 15 February 2025.

Table 6: Summary of statistical comparison for the 23-year period 1996-97 winter to the 2018-19 winter between observed and modelled accumulated water equivalent of snow. Values in parenthesis are for 28-year (winter of 1996-97 to 2023-24 for Carra and Icebox) period. Only the standard deviation is applicable to the observations (last column).

Langjökull	Carra	Icebox	RÁV2	Obs
RMSE	614 (567)	511 (468)	282	N/A
MAE	518 (460)	437 (381)	230	N/A
Bias	515 (449)	429 (363)	73	N/A
Spearman Corr	0,76 (0,73)	0,81 (0,78)	0,77	N/A
Pearson Corr	0,75 (0,72)	0,80 (0,76)	0,76	N/A
Standard dev	508 (497)	460 (452)	413	368 (369)

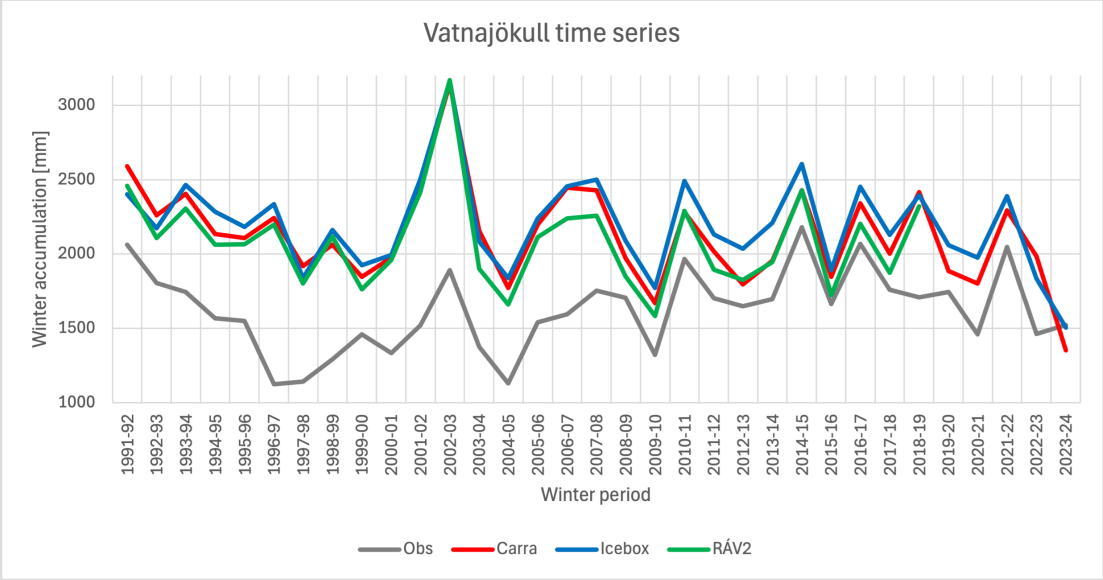


Figure 16: Modelled and observed wintertime accumulation of snow on Vatnajökull icecap in SE-Iceland. Observed values are taken from https://islenskirjoklar.is, accessed on 15 February 2025.

Table 7: Summary of statistical comparison for the 28-year period 1991-92 winter to the 2018-19 winter between observed and modelled accumulated water equivalent of snow. Values in parenthesis are for 33-year (winter of 1991-92 to 2023-24 for Carra and Icebox) period. Only the standard deviation is applicable to the observations (last column).

Vatnajökull	Carra	Icebox	RÁV2	Obs
RMSE	622 (586)	671 (633)	554	N/A
MAE	555 (514)	622 (575)	472	N/A
Bias	555 (504)	622 (574)	472	N/A
Spearman Corr	0,56 (0,55)	0,62 (0,61)	0,54	N/A
Pearson Corr	0,54 (0,50)	0,60 (0,58)	0,52	N/A
Standard dev	305 (325)	289 (307)	316	266 (269)

Comparison of observed wintertime accumulation on the three icecaps show that both Carra and Icebox overestimate wintertime precipitation more than the old RÁV2 series. The RÁV2 data also show lower RMSE and MAE than both Carra and Icebox, whilst Icebox shows the highest correlation of the three models. It is tempting to point out that both Carra and Icebox are forced by ERA5 re-analyses, whilst RÁV2 used data from the older ERA-Interim reanalysis. It is also worth pointing out the observed wintertime accumulation is most likely underestimating the total wintertime precipitation. As pointed out in [7] “Due to factors such as occasional winter thaw events, winter precipitation that falls as rain, and sublimation of snow, this should be a slight underestimation of the actual precipitation on the glacier“. The interannual changes of the simulated precipitation on the three icecaps (Hofsjökull, Langjökull and Vatnajökull) compare well with the observed changes. The exception from this are the winters of 1996-97 and 2018-19 when all three models show a considerable increase, compared to the previous winter, while observations (especially on Hofsjökull and Vatnajökull) show a decrease in precipitation. It is not known what causes this deviation.

Short- and long wave radiation (27 and 17 stations, respectively)
Landsvirkjun operates several weather stations that observe incoming short- and long wave radiation fluxes during summer and autumn months. This data, in combination with data from year-around stations, has been compared to simulated values.

The Pearson correlation of simulated short-wave radiation by Carra is 0.79 whilst the correlation for Icebox is 0.85. Standard deviation of the simulations is 182 W/m^2 (Carra) and 200 W/m^2 (Icebox) whilst observed value is 192 W/m^2. The RMSE of Carra is 123.1 W/m^2 and 106.3 W/m^2 for Icebox. For RÁV2 these values are 0.74 (correlation), 207 W/m^2 (standard deviation) and 142.5 W/m^2 (cf. Figure 17, top right). All models capture the distribution reasonably well but start to have issues with observed values greater than 800 W/m^2 (cf. Figure 17, top left). This may be linked to downward reflection from low level clouds and/or fog over snow covered or glaciated surface. Note that the influence of the horizontal diffuse transport is not considered in the 1D radiative transfer models (RTMs) that are used in these simulations. Low clouds, such as stratocumulus, facilitate the horizontal escape of diffuse irradiance beyond cloud boundaries while simultaneously increasing interactions with the surface due to high albedo. One dimensional RTM models treat clouds as plane-parallel layers, neglecting these 3D effects. This phenomenon becomes especially significant in very high-resolution data on short temporal scales.

The Carra model has a negative bias of short-wave radiation at most of the highland stations and most of the stations on Vatnajökull icecap (cf. Figure 16, maps), the exception being Hveravellir and stations GF and GV_VH on Vatnajökull icecap. It is worth noting that there is considerable difference in observed bias in all model simulations between stations GF and GV_VH, even though they are located very close to each other near the centre of Vatnajökull icecap. Icebox and RÁV2 have more neutral and positive biases, with Icebox outperforming RÁV2 at most locations. Table 8 summarises the statistical comparison of the downward short-wave radiative fluxes for all three models.

Table 8: Summary of statistical comparison for the 20-year period 1 September 1999 to 31 August 2019, between observed and modelled downward short-wave radiation fluxes. Values in parenthesis are for the 25-year period 1 September 1999 to 31 August 2024. Only the standard deviation is applicable to the observations (last column).

Short-wave	Carra	Icebox	RÁV2	Obs
RMSE	123,1 (119,8)	106,3 (104,3)	142,5	N/A
MAE	71,1 (69,0)	52,5 (51,4)	76,3	N/A
MBias	0,97 (0,97)	1,10 (1,09)	1,14	N/A
Spearman Corr	0,81 (0,81)	0,90 (0,90)	0,84	N/A
Pearson Corr	0,79 (0,80)	0,85 (0,86)	0,74	N/A
Standard dev	182 (180)	200 (198)	207	192 (190)

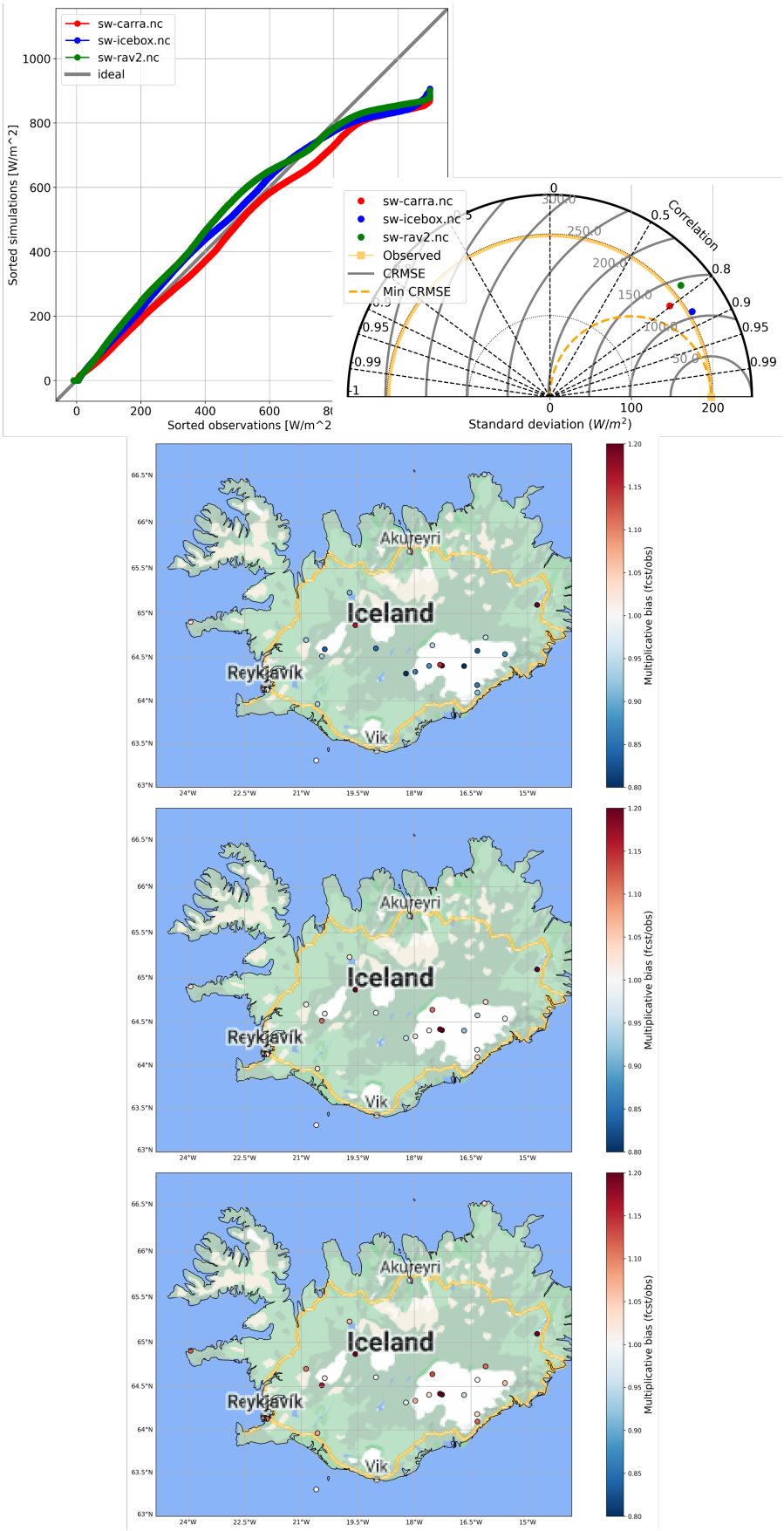


Figure 17: Quantile-Quantile plot (top left), Taylor diagram (top right), and multiplicative bias of simulated downward short-wave flux. Results from Carra (top), Icebox (middle) and RÁV2 (bottom) simulations between 1 September 1999 and 31 August 2019. See text for details.

Intra-annual variability of RMSE, MAE, Bias, and Pearson correlation for short-wave radiation fluxes is shown in Figure 18.

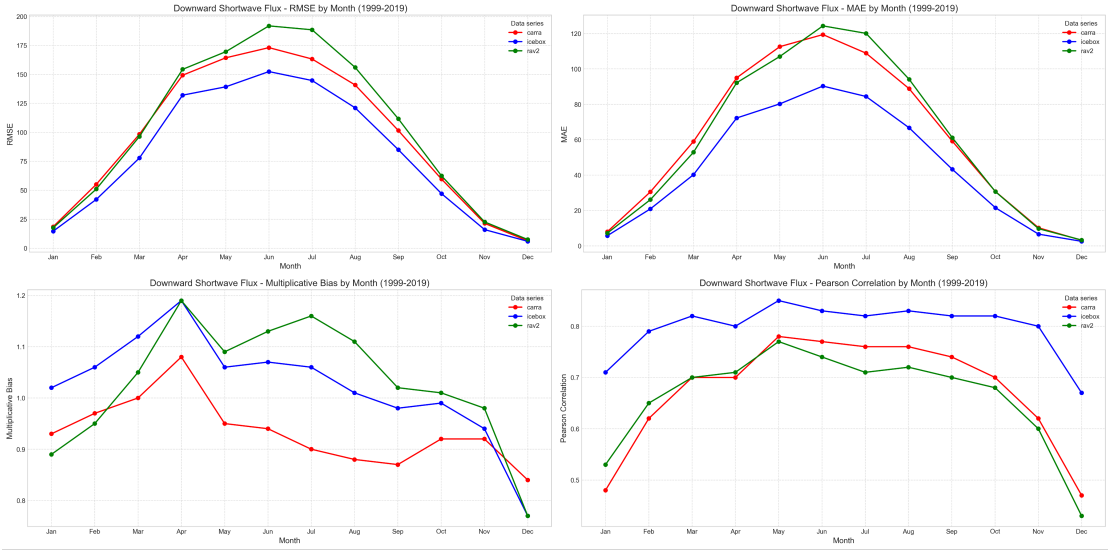


Figure 18: Monthly values of RMSE (top left), MAE (top right), MBias (bottom left) and Pearson correlation (bottom right) for short-wave fluxes for the period 1 September 1999 to 31 August 2019.

All three models show strong seasonality for RMSE, MAE with maximum values found during the summer months. The RÁV2 dataset has the largest errors and Icebox the lowest. This seasonal signal is not as strong for the multiplicative bias as it is for the RMSE and MAE. There is still a distinct maximum found in April in all datasets. In general, Carra underestimates the fluxes (except in April) whilst Icebox overestimates it between January and July (peaking in April). RÁV2 underestimates the radiation during winter and overestimates it over the summer months. All three models show the largest relative underestimation in December. The Pearson correlation has a strong seasonality in all three datasets. Minimum values are found during the winter months and greatest correlation is during March and October. The Icebox dataset is the exception where only January and December have notably lower correlation compared to the rest of the months.

The observed Pearson correlation of long wave radiation is 0.71 for Carra and 0.67 for Icebox. The RMSE is 33.6 W/m^2 for Carra and 36.4 W/m^2 for Icebox. The standard deviation of both models is greater than that of observations (44 W/m^2 and 45 W/m^2 vs. observed value of 38 W/m^2, cf. Figure 19, top right). The correlation and RMSE of the RÁV2 data are of notably less quality, with correlation equal to 0.3 and RMSE equal to 51.4 W/m^2. The standard deviation of RÁV2 is similar to that of both Carra and Icebox, or 46 W/m^2. The multiplicative bias (MBias) of Carra and Icebox is equal to 0.95 and 0.96 for RÁV2. Carra has the lowest mean absolute error (MAE), or 24.3 W/m^2, followed by Icebox (MAE equal to 26.9 W/m^2) and RÁV2 showing the largest MAE of 39.7 W/m^2. All models tend to underestimate long wave radiation for values less than 300 W/m^2, but capture higher values quite well. This is seen clearly in Figure 19, top left panel. Table 9 summarises this statistical comparison.

Table 9: Summary of statistical comparison for the 20-year period 1 September 1999 to 31 August 2019, between observed and modelled downward long wave radiation fluxes. Values in parenthesis are for the 25-year period 1 September 1999 to 31 August 2024. Only the standard deviation is applicable to the observations (last column).

Long wave	Carra	Icebox	RÁV2	Obs
RMSE	33,6 (31,8)	36,4 (34,9)	51,4	N/A
MAE	24,3 (22,8)	26,9 (25,4)	39,7	N/A
MBias	0,95 (0,96)	0,95 (0,96)	0,96	N/A
Spearman Corr	0,72 (0,76)	0,68 (0,72)	0,31	N/A
Pearson Corr	0,71 (0,76)	0,67 (0,71)	0,30	N/A
Standard dev	44 (45)	45 (46)	46	38 (39)

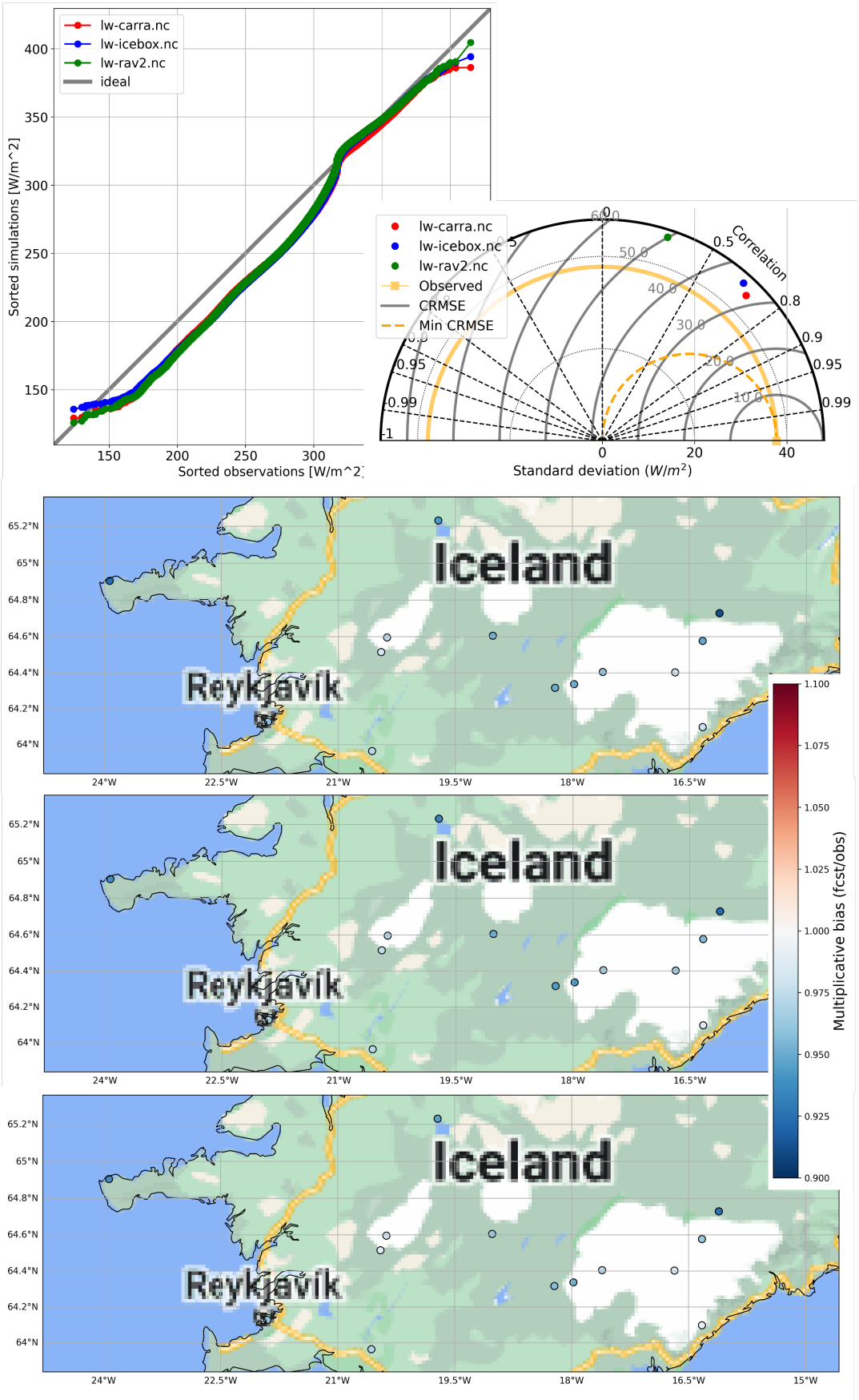


Figure 19: Quantile-Quantile plot (top left), Taylor diagram (top right), and multiplicative bias of simulated downward long wave flux. Results from Carra (top), Icebox (middle) and RÁV2 (bottom) simulations between 1 September 1999 and 31 August 2019. See text for details.

Intra-annual variability of RMSE, MAE, Bias, and Pearson correlation for longwave radiation fluxes is shown in Figure 20.

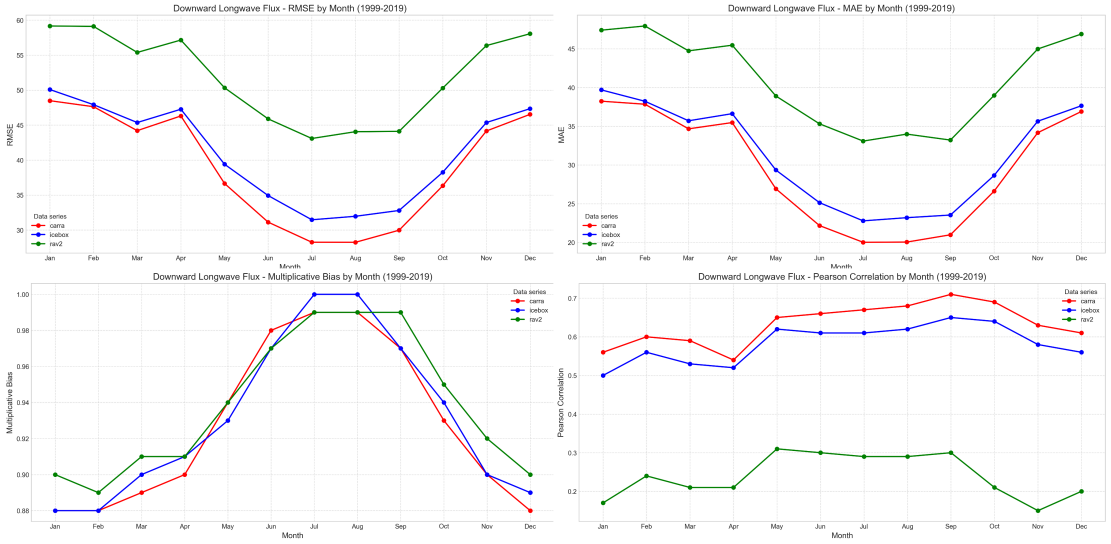


Figure 20: Monthly values of RMSE (top left), MAE (top right), MBias (bottom left) and Pearson correlation (bottom right) for longwave fluxes for the period 1 September 1999 to 31 August 2019.

All datasets have a strong seasonal signal in RMSE, MAE, and MBias where the winter months have the largest error and summer months the smallest. April stands out as it has a slightly greater RMSE and MAE errors than the previous month. Carra has the lowest error, followed closely by Icebox but RÁV2 has distinctly greater RMSE and MAE errors than the other two datasets. All datasets have the highest Pearson correlation during summer and autumn months and lowest during winter and spring time.

Summary

Results from simulated weather, using three different atmospheric re-analyses, Carra, Icebox, and RÁV2 have been compared to observations from over 50 weather stations in Iceland for the twenty-year period between 1 September 1999 and 31 August 2019. In addition, simulations have been compared to observations of accumulated wintertime precipitation on the three icecaps Hofsjökull, Langjökull, and Vatnajökull. For Hofsjökull, the comparison period is the winters of 1991-92 to 2023-24 (Carra, 33 years), 1990-91 to 2023-24 (Icebox, 34 years), and 1988-89 to 2018-19 (RÁV2, 31 year). For Langjökull, the comparison period is the winters of 1996-97 to 2023-24 (Carra and Icebox, 28 years), and 1996-97 to 2018-19 (RÁV2, 23 years). For Vatnajökull, the comparison period is the winters of 1991-92 to 2023-24 (Carra and Icebox, 33 years), and 1991-92 to 2018-19 (RÁV2, 28 years).

The general conclusion is that data from the Carra and Icebox re-analyses' series are of similar quality, outperforming the older RÁV2 model on most fronts. Simulated near surface temperature and wind speed are on par, whilst Icebox captures short range radiation better than Carra. The opposite holds true for long wave radiation where Carra outperforms Icebox. Simulated hourly precipitation rate is in better agreement with observations in Icebox than data from Carra and RÁV2. When precipitation data is aggregated over longer periods of days, weeks and months, this difference becomes less prominent. Indeed, Carra and Icebox tend to overestimate the precipitation

whilst RÁV2 is in better agreement with both daily, weekly, and monthly precipitation. Intra-annual comparison indicates that Carra underestimates the wintertime precipitation and overestimates the summertime precipitation compared to observations. Icebox on the other hand overestimates the precipitation independent of time of year. This might explain the peak in RMSE and MAE temperature errors seen in May and June in the Icebox dataset (cf. Figure 6). I.e. too extensive snow cover, caused by excessive wintertime precipitation, could lead to abnormally cold springs and early summer. Simulated wintertime precipitation compares more favourably to observations from Hofsjökull and Langjökull icecaps in Icebox than Carra, results for Vatnajökull are mixed where Carra has lower RMSE, MAE and Bias but Icebox correlates better with observations, and standard deviation is closer to that of the observations. However, the wintertime precipitation from RÁV2 is in better agreement with observed accumulated precipitation on all three icecaps, except for correlation where the Icebox simulation is better.

To aid with this comparison an on-line verification solution has been created (https://verif.belgingur.is – see description in Appendix A). This tool can be used to browse and visualize verification results from the Carra and Icebox atmospheric simulation. Data, both observations and simulations alike, from individual observational locations can be downloaded as simple text files. In addition, users can view a set of statistical properties in a table format and/or download results as a CSV file.

Acknowledgements

The authors acknowledge comments and suggestions made by Hálfdán Ágústsson and Örnólfur E. Rögnvaldsson that improved the report considerably. Enlightning discussions with Tómas Jóhannesson on observations made on the three icecaps are acknowledged and appreciated.

Appendix A

In this appendix, a graphical tool, built on top of the Verif [1] solution, is described. This tool can be used to browse and visualize verification results from any atmospheric re-analysis simulation, as long as the data have been converted to the WOD standardized netCDF file format. General guidelines on how to prepare data for the Verif solution can be found on Verif's Wiki page (https://github.com/WFRT/verif/wiki/Arranging-my-own-data). There, the user can find information on how the Verif package specifies the data format and how to load the data into the NetCDF files to be read by the Verif system. The WOD API system (URL of WOD RESTful API: https://wod.belgingur.is/api/v2/ui/, and for further information see https://github.com/Belgingur/WOD-Documentation/wiki/Getting-Started-With-WOD-APIs) can be used to download data, observations, and model data alike, which then are fed into Verif's pre-processing tools (again, we refer to the Verif Wiki page for further instructions) to create the files that are eventually interpreted by Verif. This simplifies considerably the process of comparing results from different modelling systems as the task of converting model data into a unified format has already been conducted within the WOD framework. That is, the user can use the same API to access results from a plethora of atmospheric models.

The landing page for the Verif web service is shown in Figure A1 (top panel). Once the user has logged on, he/she needs to select a file containing observations and model simulations of one variable. This is how Verif works, i.e., it operates on one file at a time, where the said file contains the observed and simulated data of a single variable.

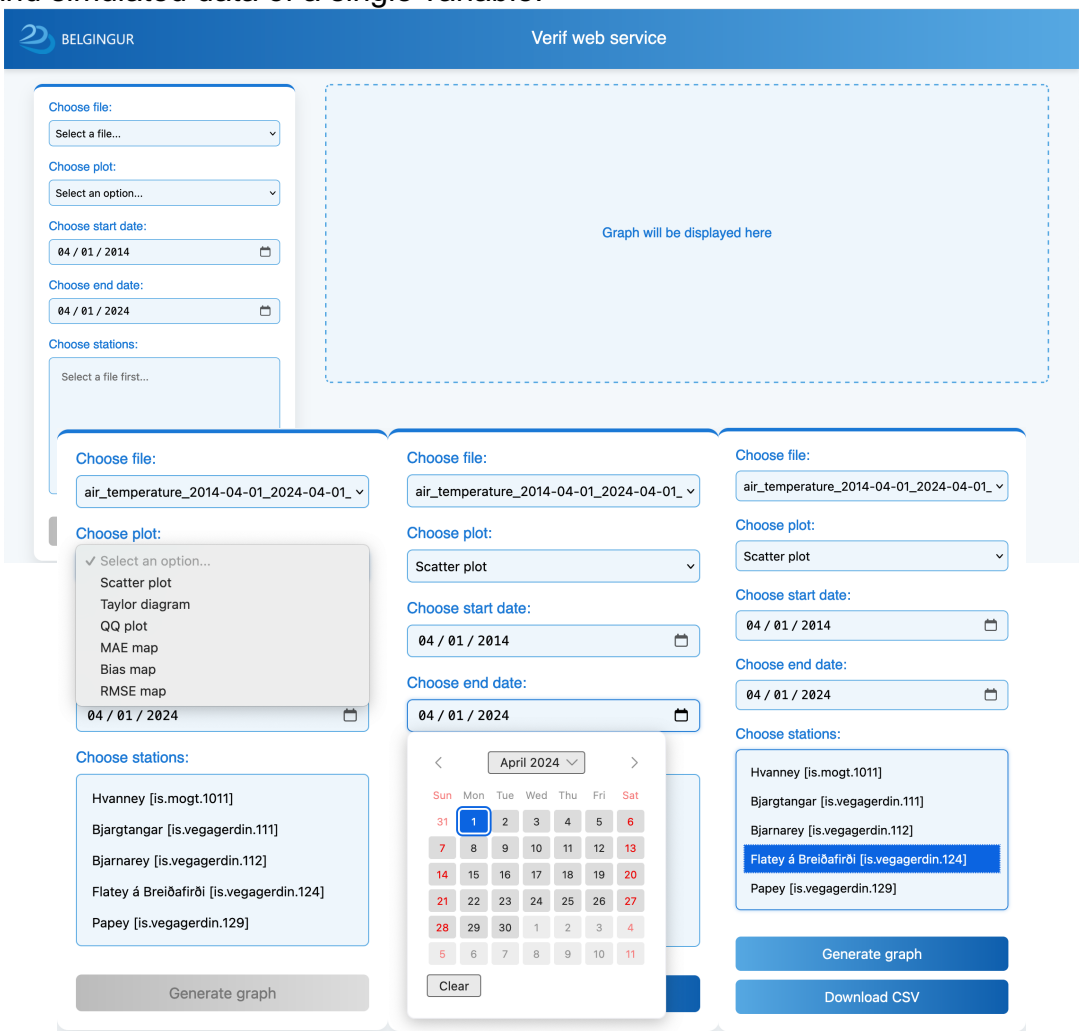


Figure A1. The landing page (top panel) of the Verif web service offers the user the choice of a set of observed and modelled variables as well as plot options (lower panel, left); data range options (lower panel, middle); and the option of customizing which observation locations are to be investigated (lower panel, right). In addition, users can view a set of statistical properties in a table format and/or download results as a CSV file (not shown).

If a single observation station is chosen, the user can also download observed and simulated data in a text format by clicking the “Download CSV” button. In addition to creating scatter plots, Taylor diagrams, and quantile–quantile plots (cf. Figure A2, top panels), the user can also plot three different types of maps (cf. Figure A2, bottom panels).

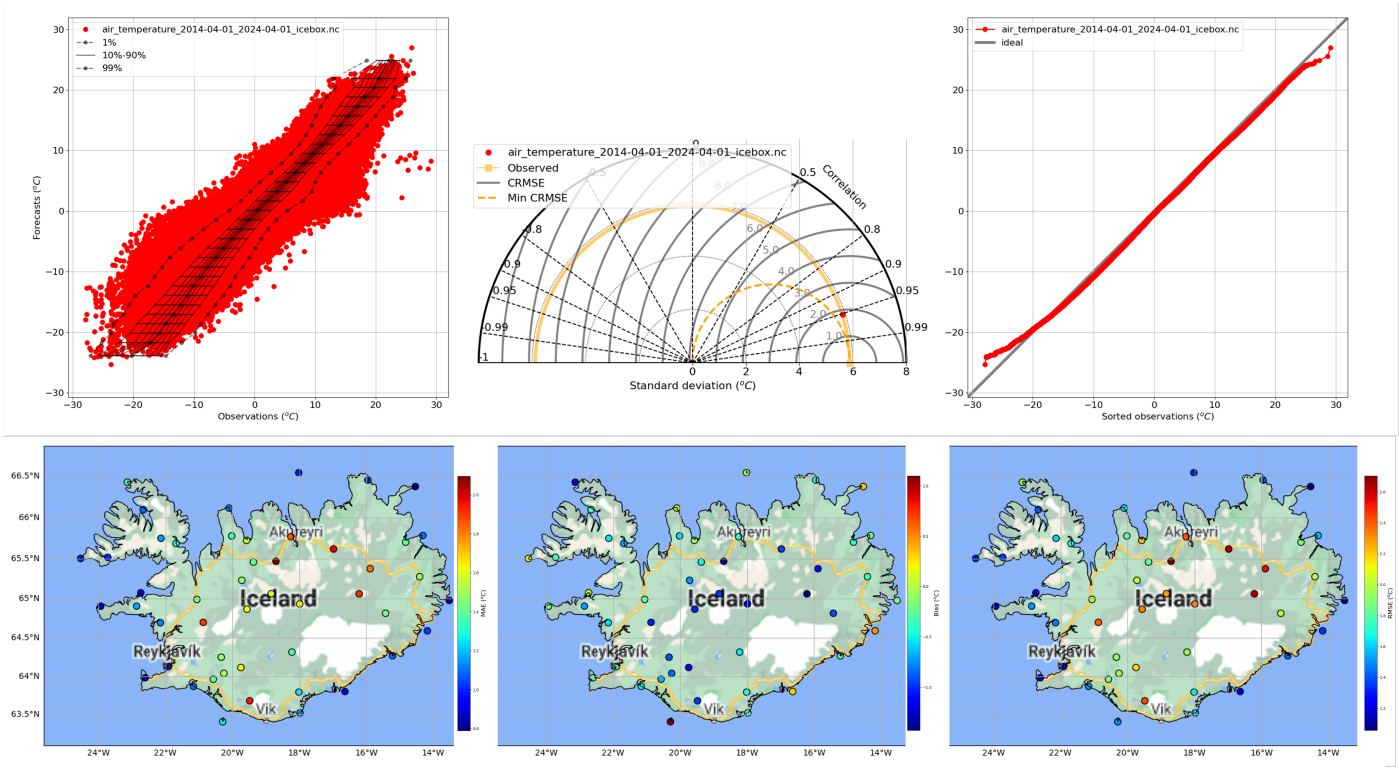


Figure A2. The Verif web service offers seven types of graphs. These are scatter plots (top left), Taylor diagrams (top centre), quantile–quantile plots (top right), and maps showing mean absolute error (bottom left), bias (bottom centre), root-mean-square error (bottom right), and multiplicative bias (not shown).

References

[1] Nipen, T.N., R. B. Stull, C. Lussana, and I. A. Seierstad, 2023. Verif: A Weather-Prediction Verification Tool for Effective Product Development. *Bull. Am. Meteorol. Soc.*, 104, E1610-E1618. Available on-line: https://doi.org/10.1175/BAMS-D-22-0253.1 (last access: 21 February 2025).

[2] Rögnvaldsson, Ó., K. Stanislawska, and J. A. Hackerott, 2025. The Weather On-Demand Framework. *Atmosphere*, *16*(1), 91. https://doi.org/10.3390/atmos16010091 (last access: 21 February 2025).

[3] Rögnvaldsson, Ó., 2013. Numerical simulations of winds and precipitation in Iceland - Die zweite Aufgabe der theoretischen Meteorologie. Dissertation for the degree of philosophiae doctor. University of Bergen, Norway. ISBN: 978-82-308-2225-8. Available on-line: https://bora.uib.no/handle/1956/6700 (last access: 21 February 2025).

[4] Schyberg H., and co-authors, 2020. Arctic regional reanalysis on single levels from 1991 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available on-line: https://doi.org/10.24381/cds.713858f6 (last access: 21 February 2025).

[5] Rögnvaldsson, Ó., 2020. Observed and simulated weather: Description of dynamical downscaling experiments for the water year 2014-2015 and comparison with observations, Technical report. Available on-line: http://ftp.belgingur.is/publications/ObsSim-comparison-TechReport.pdf (last access: 21 February 2025).

[6] Friðriksson, R. F., and H. Ólafsson, 2005. Estimating errors in wintertime precipitation observations by comparison with snow observations. *Croatian Meteorological Journal*, **Vol. 40**, No. 40, 2005, pp. 695-697. Available on-line: https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=97255 (last access: 21 February 2025).

[7] Helgason, H. B., and B. Nijssen, 2024. LamaH-Ice: LArge-SaMple DAta for Hydrology and Environmental Sciences for Iceland, *Earth Syst. Sci. Data*, 16, 2741–2771. Available on-line: https://doi.org/10.5194/essd-16-2741-2024 (last access: 21 February 2025).