UiT

THE ARCTIC
UNIVERSITY
OF NORWAY

CLARA-A2 dataset Improvements and shortcomings in high latitude regions

Bilal Babar
Researcher
Energy and Climate group
University of Tromsø – The Arctic University of Norway

Outline

- Part 1: Improvements of CLARA-A2 dataset

Babar, B., Graversen, R., & Boström, T. (2018). Evaluating CM-SAF solar radiation CLARA-A1 and CLARA-A2 datasets in Scandinavia. *Solar Energy*, *170*, 76-85.

- Part 2: Comparing CLARA-A2, SARAH-2 and ERA5 datasets

Babar, B., Graversen, R., & Boström, T. (2019). Solar radiation estimation at high latitudes: Assessment of the CMSAF databases, ASR and ERA5. *Solar Energy*, *182*, 397-411.

- Part 3: Improving CLARA-A2 dataset with machine learningn

Babar, B., Luppino, T,L., Boström, T & Anfinsen, S. (under review). Random forest regression for improved mapping of solar radiation at high latitudes.
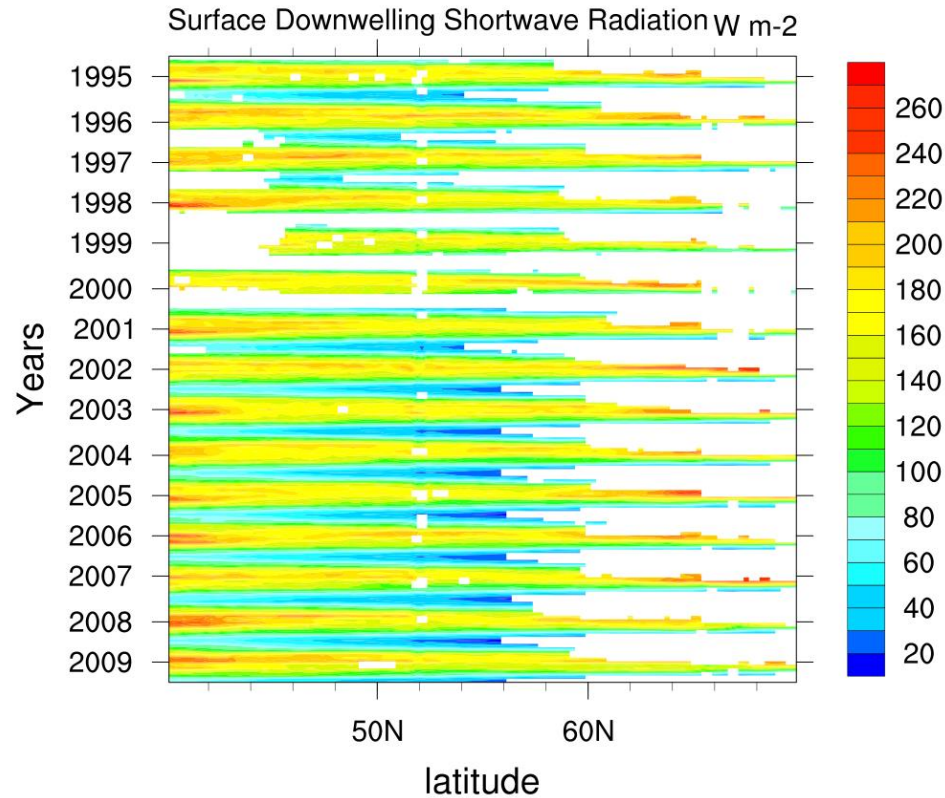
Area of investigation
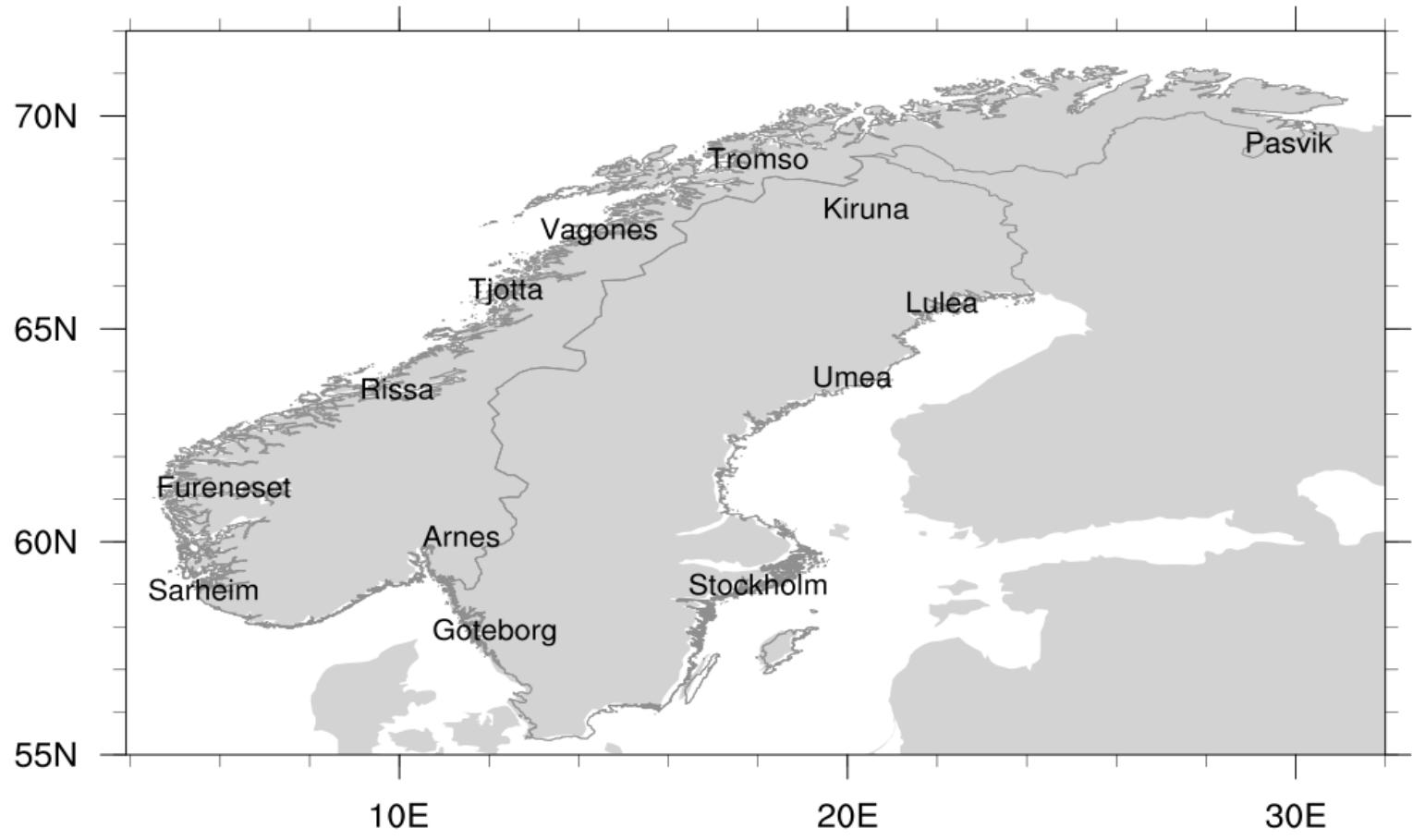
Part 1
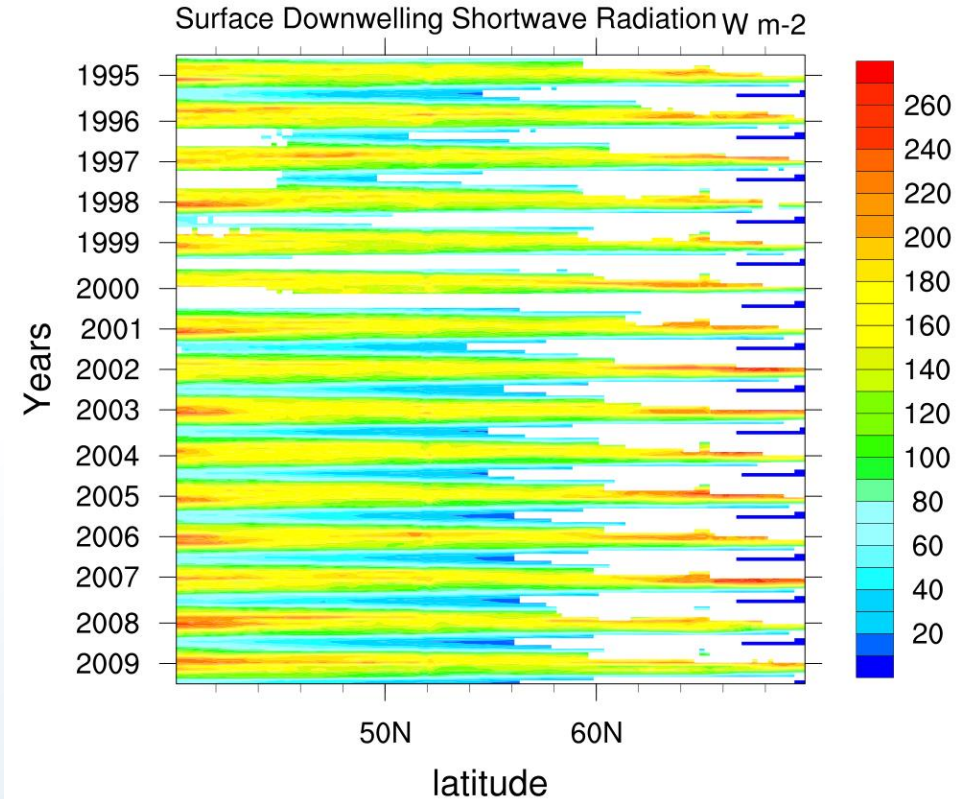## From CLARA-A1 to CLARA-A2

- CLARA-A2 has smaller amount of missing values
- The data points that were available in CLARA-A1 have improved accuracy
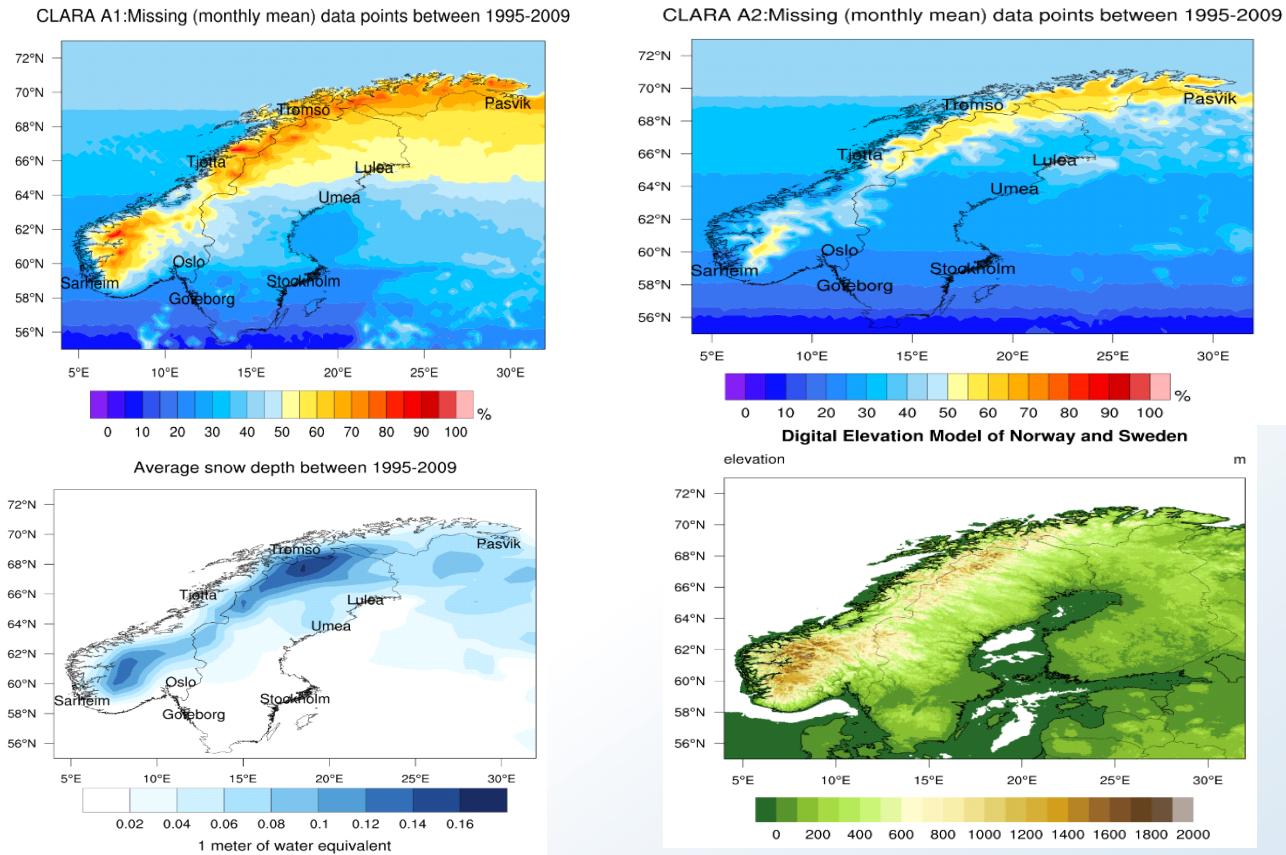- New data points have large errors

Missing data in CLARA-A1 and CLARA-A2

CLARA-A1, longitude=15, latitude=40:70, 1995-2009

CLARA-A2, longitude=15, latitude=40:70, 1995-2009

Missing data in CLARA-A1 and CLARA-A2

INCREASE IN AVAILABILITY OF CLARA A2 DATA
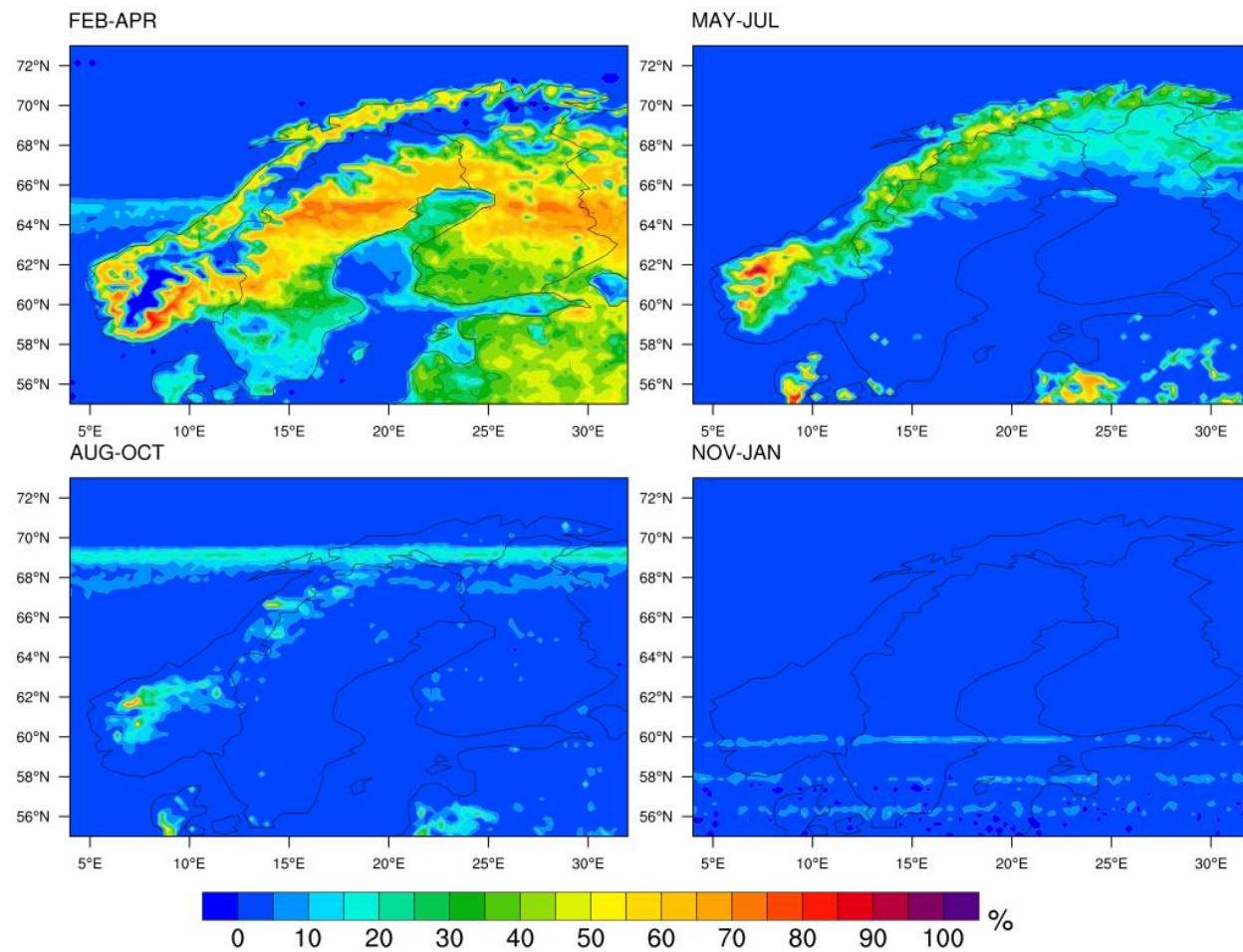
Part 2
## Comparing CLARA-A2 with SARAH-2 and ERA5

- 31 locations and 16 years of data were analysed.
- The analysis was performed by dividing locations into 4 distinct groups, i.e. above 65N, below 65N, coastal and inland.
- A sky-stratification and a seasonal error analysis was performed to assess the datasets.

Accuracy of the datasets

	RMSD(Wm^{-2})			MAD(Wm^{-2})			MBD(Wm^{-2})		
	CLARA	SARAH	ERA5	CLARA	SARAH	ERA5	CLARA	SARAH	ERA5
All Sites	8.6 (18.6)	**7.7** (**17.8**)	10.2 (27.0)	5.6 (13.1)	**5.0** (**11.8**)	6.8 (17.4)	**-1.1** (**-1.6**)	-1.9 (-2.4)	4.1 (4.5)
Above 65°N	**10.0** (**18.8**)	-	12.9 (29.6)	5.7 (11.7)	-	8.5 (17.5)	**-2.4** (**-2.9**)	-	6.3 (6.7)
Below 65°N	8.4 (18.6)	**7.7** (**17.8**)	9.8 (26.7)	5.6 (13.3)	**5.0** (**11.8**)	6.6 (17.4)	**-1.0** (**-1.4**)	-1.9 (-2.4)	3.9 (4.2)
Coastal	7.8 (17.6)	**7.1** (**16.5**)	10.4 (27.1)	4.9 (12.2)	**4.5** (**11.0**)	6.8 (17.0)	**-0.7** (**-1.1**)	-1.6 (-1.9)	4.3 (4.7)
Inland	9.1 (19.4)	**8.0** (**18.5**)	10.1 (27.0)	6.1 (13.8)	**5.2** (**12.2**)	6.8 (17.7)	**-1.5** (**-1.9**)	-2.1 (-2.7)	4.0 (4.2)

Sky Stratification analysis

CLARA
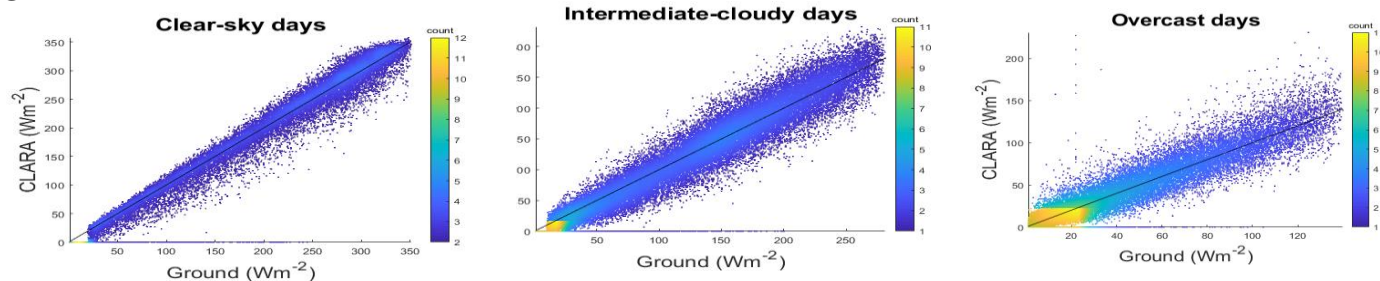
CLARA	RMSD (Wm^{-2})	MBD (Wm^{-2})
Clear-sky	21.5	-3.9
Intermediate-cloudiness	22.1	-3.2
Overcast	12.9	-0.1

SARAH
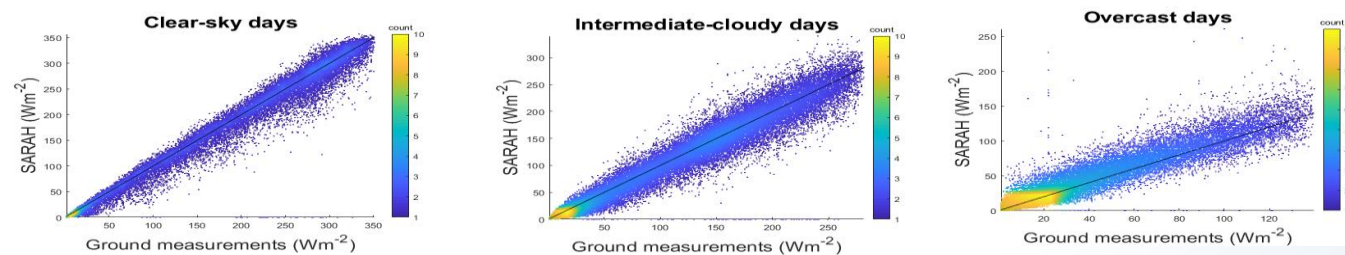
SARAH	RMSD (Wm^{-2})	MBD (Wm^{-2})
Clear-sky	20.4	-5.5
Intermediate-cloudiness	20.1	-2.9
Overcast	13.2	4.5

ERA5
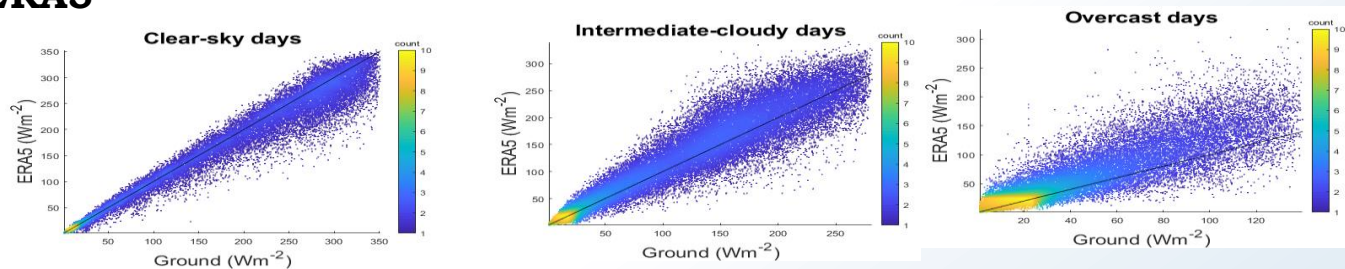
ERA5	RMSD (Wm^{-2})	MBD (Wm^{-2})
Clear-sky	25.4	-9.9
Intermediate-cloudiness	28.4	8.7
Overcast	29.6	15.3

Part 3
Random forest regression on CLARA-A2 and ERA5

- CLARA-A2, ERA5, latitude, altitude, clear-sky indices and solar zenith angle as input
- 20% data for training and 80% for testing
- The model was trained on Norwegian locations and tested on Swedish locations

Error Analysis of Norwegian Locations

	RMSD (Wm^{-2})			MAD (Wm^{-2})			Bias (Wm^{-2})		
	CLARA	ERA5	RFR	CLARA	ERA5	RFR	CLARA	ERA5	RFR
NIBIO sites	8.6 (18.6)	10.2 (27.0)	**6.6 (15.5)**	5.6 (13.1)	6.8 (17.4)	**4.3 (10.1)**	-1.1 (-1.6)	4.1 (4.5)	**-0.2 (0.0)**
Above 65°N	9.6 (16.0)	10.1 (26.3)	**6.5 (13.5)**	5.7 (11.7)	8.5 (17.5)	**4.2 (8.1)**	-2.4 (-2.9)	6.3 (6.7)	**-0.2 (0.4)**
Below 65°N	9.7 (19.5)	12.7 (26.8)	**8.0 (15.7)**	5.6 (13.3)	6.6 (17.4)	**5.4 (10.4)**	-1.0 (-1.4)	3.9 (4.2)	**0.1 (-0.1)**
Coastal	9.7 (16.7)	10.1 (26.7)	**6.6 (16.2)**	4.9 (12.2)	6.8 (17.0)	**4.3 (10.7)**	-0.7 (-1.1)	4.3 (4.7)	**-0.2 (-0.3)**
Inland	8.2 (20.8)	11.2 (26.7)	**6.6 (14.5)**	6.1 (13.8)	6.8 (17.7)	**4.6 (9.4)**	-1.5 (-1.9)	4.0 (4.2)	**0.1 (0.3)**

Error Analysis of Swedish locations

	RMSD (Wm^{-2})			MAD (Wm^{-2})			Bias (Wm^{-2})		
	CLARA	ERA5	RFR	CLARA	ERA5	RFR	CLARA	ERA5	RFR
Kiruna	17.2 (26.6)	**7.6** (24.0)	11.0 (**18.7**)	10.1 (16.6)	**4.9** (14.4)	6.7 (**11.7**)	-7.0 (-8.2)	**-2.3** (**-2.5**)	-5.7 (-5.8)
Luleå	10.6 (24.4)	10.4 (25.1)	**5.5** (**17.4**)	6.9 (14.9)	6.6 (15.3)	**3.7** (**10.8**)	-4.4 (-4.2)	5.1 (4.9)	**-2.1** (**-2.1**)
Umeå	8.3 (16.4)	7.1 (23.0)	**5.3** (**13.2**)	6.1 (11.5)	4.4 (14.2)	**3.7** (**8.9**)	-3.2 (-3.5)	**2.0** (**2.1**)	-2.4 (-2.3)
Stockholm	6.8 (16.4)	7.0 (23.6)	**5.9** (**14.6**)	5.1 (11.5)	4.8 (15.7)	**4.2** (**9.8**)	**2.6** (**2.5**)	3.1 (3.1)	3.7 (3.6)
Göteborg	4.7 (14.9)	9.5 (26.1)	**4.6** (**14.6**)	**3.5** (10.5)	7.3 (17.0)	3.7 (**10.1**)	**1.6** (**1.8**)	6.9 (6.8)	2.8 (2.9)
SMHI locations	10.4 (20.3)	8.4 (24.4)	**6.8** (**15.8**)	6.3 (13.0)	5.6 (15.3)	**4.4** (**10.2**)	-2.1 (-2.3)	2.9 (2.9)	**-0.8** (**-0.8**)

Conclusion

Advantages
- CLARA cdr provides more accurate estimates in high latitude regions than other available datasets.
- The dataset is improving with each iteration

Shortcomings
- Missing data points
- Large errors on snow covers
- Lower temporal resolution as compared to other datasets, e.g. ERA5