Supplement of Atmos. Meas. Tech., 15, 5701–5717, 2022
https://doi.org/10.5194/amt-15-5701-2022-supplement

Atmospheric Measurement Techniques
Open Access
EGU

Supplement of

Ice water path retrievals from Meteosat-9 using quantile regression neural networks

Adrià Amell et al.

Correspondence to: Adrià Amell (amell@chalmers.se)

This document present supplementary material to the paper "ice water path retrievals from Meteosat-9 using quantile regression neural networks". All figures herein are in vector format, so that one can zoom to read the small font sizes and details.

S1 The collocations database

The collocations consist of nearly all Meteosat-9 SEVIRI images and DARDAR-cloud version 2.1.1 profiles between 6 May 2008 and 31 March 2011, collocated as described in the paper. We state 'nearly all' instead of 'all' as few available DARDAR observations failed to download at the time the database was constructed. The upper time bound is given by the CloudSat battery anomaly in April 2011, which forced it to be daylight-only operations (Sect. 2.1). Therefore, DARDAR data after this date only exists for daytime. The lower time bound is due to the files provided by the EUMETSAT Data Store: files before 6 May 2008 were processed with another algorithm version than those posterior to this date. We have not found a citable reference that clearly explains this. This information is scattered on the Internet and can be noticed from the SEVIRI file names. Therefore, our database contains nearly all possible day and night collocations for Meteosat-9 SEVIRI images processed with the same algorithm version before the CloudSat battery anomaly. The temporal coverage of these collocations in the region of interest (ROI) is given in Fig. S1.

Figure S1: Temporal coverage of the collocations in the database. Samples refer to the 32×32 images.

Figure S2 shows the diversity in the amount of DARDAR profiles used to obtain the collocated values for the SEVIRI pixel over the ROI, as well as the frequency of each number of profiles per pixel. Fig. S2 relates to Fig. 1 in the paper. Note that here we use the term profile to identify both the original and replicated DARDAR profiles at the swath edges (blue and green marks,

respectively, in Fig. 1). Towards the lower right corner of the ROI, up to 12 DARDAR profiles can be used to obtain a collocated value, which is a consequence of the projection used, as it is the furthest area from the centre of the projection. However, up to 9 DARDAR profiles are used more uniformly over the ROI.

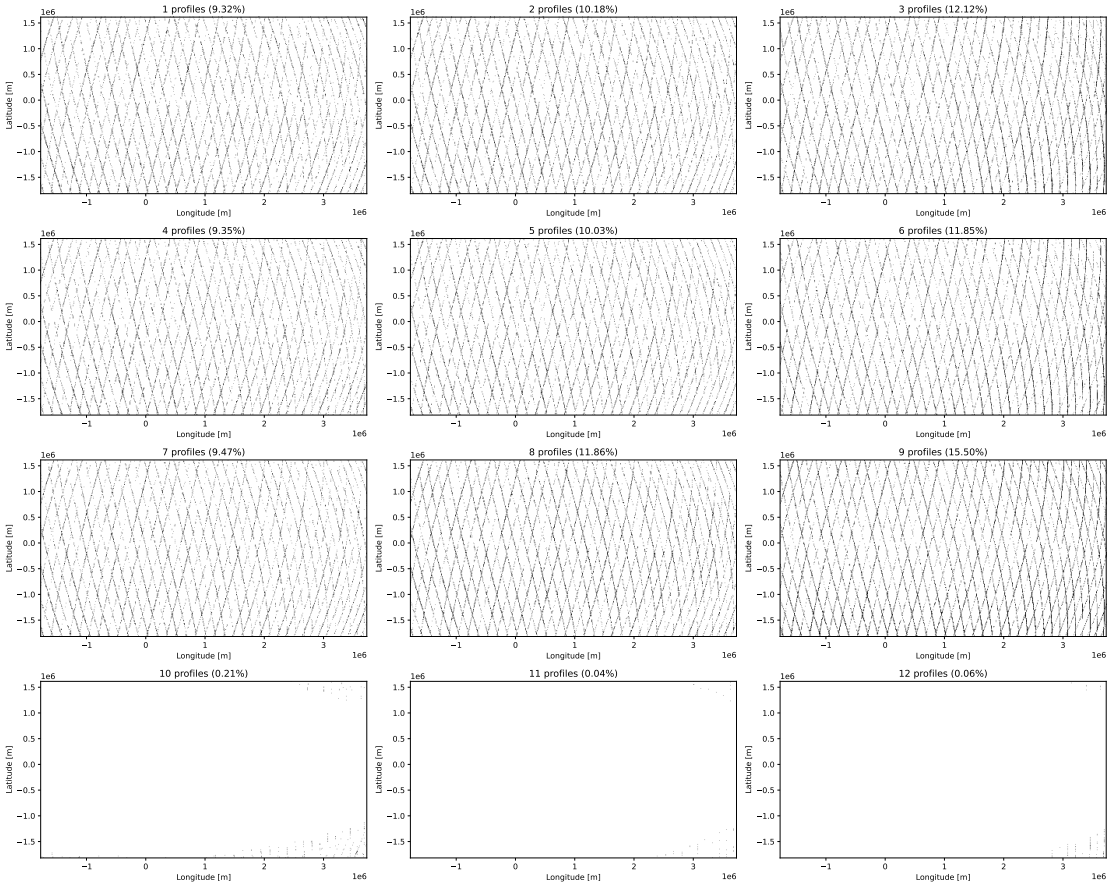
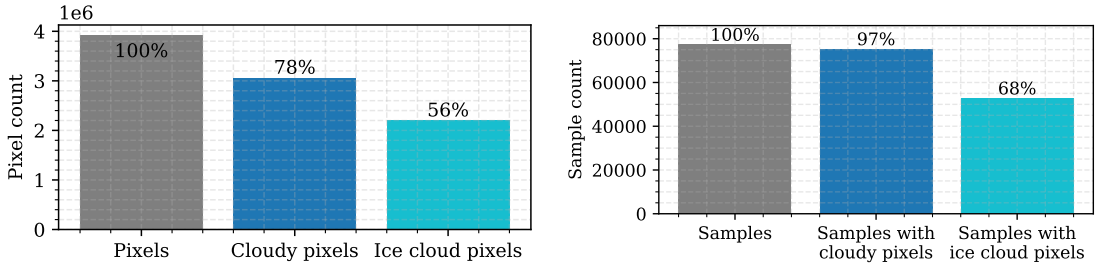


Figure S2: Number of DARDAR profiles collocated with each SEVIRI pixel. Each dot represents a SEVIRI pixel (not to scale). The subplot titles indicate how many DARDAR profiles, as described in text, are collocated in the SEVIRI pixel and therefore used to obtain the collocated value. The subplot title also indicates how often the given number of profiles are collocated in a SEVIRI pixel.

DARDAR provides a flag for each DARDAR vertical bin, describing its content. Among these flags, exist the values `ice`, `ice + supercooled`, `liquid warm`, and `supercooled`. We tagged profiles with any of these flags present as cloudy, and with the flags `ice` and `ice + supercooled` as ice clouds. Any SEVIRI pixel with a DARDAR profile with a cloudy flag is tagged as such, and analogously for ice clouds. Any sample (the 32×32 pixels images) with a cloudy or ice cloud pixel are also tagged as such. Figure S3 summarizes the frequency of each of these tags in the collocations database.

Figures S4 and S5 show the relationships between the collocated DARDAR IWP and the visible and infrared radiances for daytime and nighttime, respectively. Concerning the thermal infrared, there are small differences between the daytime and nighttime relationships, but this



(a) Frequency in SEVIRI pixels.

(b) Frequency in the samples.

Figure S3: Frequencies of SEVIRI pixels or samples containing no cloud, cloud or ice cloud flags.

is not the case for the channels with a solar contribution, as is expected. Note that, while the relationship for channel 4 presents discontinuities for nighttime, this channel is never used for nighttime retreivals in the models presented. On the other hand, it can be observed that the distributions are reasonable: for infrared channels, the colder a SEVIRI pixel is, the higher the IWP; for solar channels in daytime, the higher the reflectance is, the higher the IWP.

Finally, the training, validation and test distributions are nearly identical, independently of whether daytime-only data or daytime and nighttime data from each set is considered. Figure 9 in the paper shows the test distribution for daytime data, from which the remaining distributions can be extrapolated. Nevertheless, Table S1 provides different quantiles for each subset of the collocations database used in the paper, as well as the respective mean.

Table S1: Quantiles and mean of the DARDAR collocations of the different data sets used, all in $\mathrm{kg\,m^{-2}}$. Daytime indicates only daytime data, and fulltime daytime and nighttime data.

	Set	Training		Validation		Test	
		Daytime	Fulltime	Daytime	Fulltime	Daytime	Fulltime
Percentile	5%	0.00	0.00	0.00	0.00	0.00	0.00
	16%	0.00	0.00	0.00	0.00	0.00	0.00
	25%	0.00	0.00	0.00	0.00	0.00	0.00
	50%	0.00	4.29×10^{-4}	0.00	3.85×10^{-4}	0.00	4.03×10^{-4}
	75%	8.51×10^{-3}	1.23×10^{-2}	8.33×10^{-3}	1.17×10^{-2}	8.50×10^{-3}	1.28×10^{-2}
	84%	2.67×10^{-2}	3.13×10^{-2}	2.66×10^{-2}	3.08×10^{-2}	2.70×10^{-2}	3.21×10^{-2}
	95%	2.92×10^{-1}	3.00×10^{-1}	3.10×10^{-1}	2.99×10^{-1}	2.94×10^{-1}	3.11×10^{-1}
Mean		1.14×10^{-1}	1.18×10^{-1}	1.18×10^{-1}	1.20×10^{-1}	1.14×10^{-1}	1.19×10^{-1}

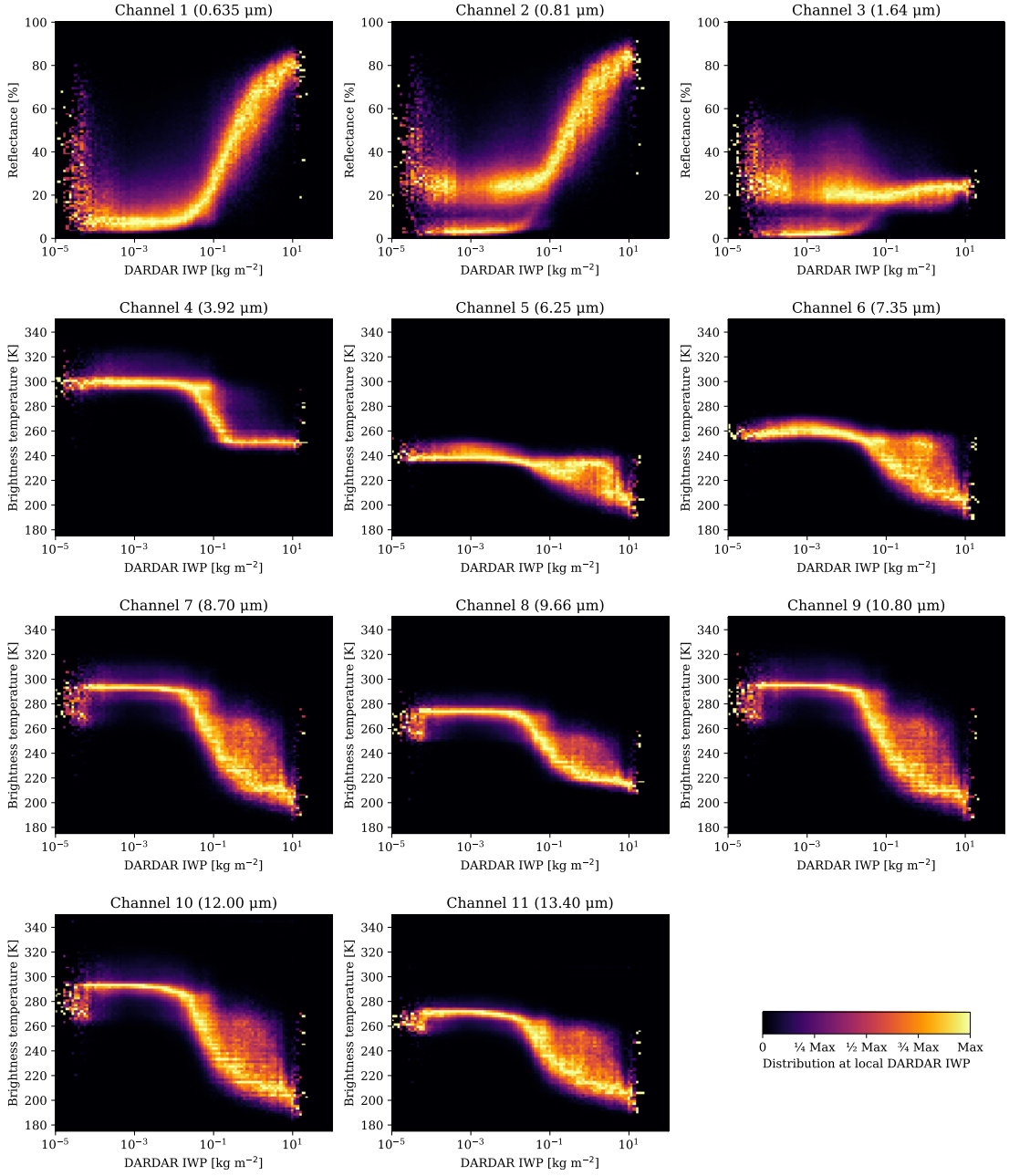


Figure S4: Relationships for the SEVIRI channels and the DARDAR collocated IWP for daytime data only.

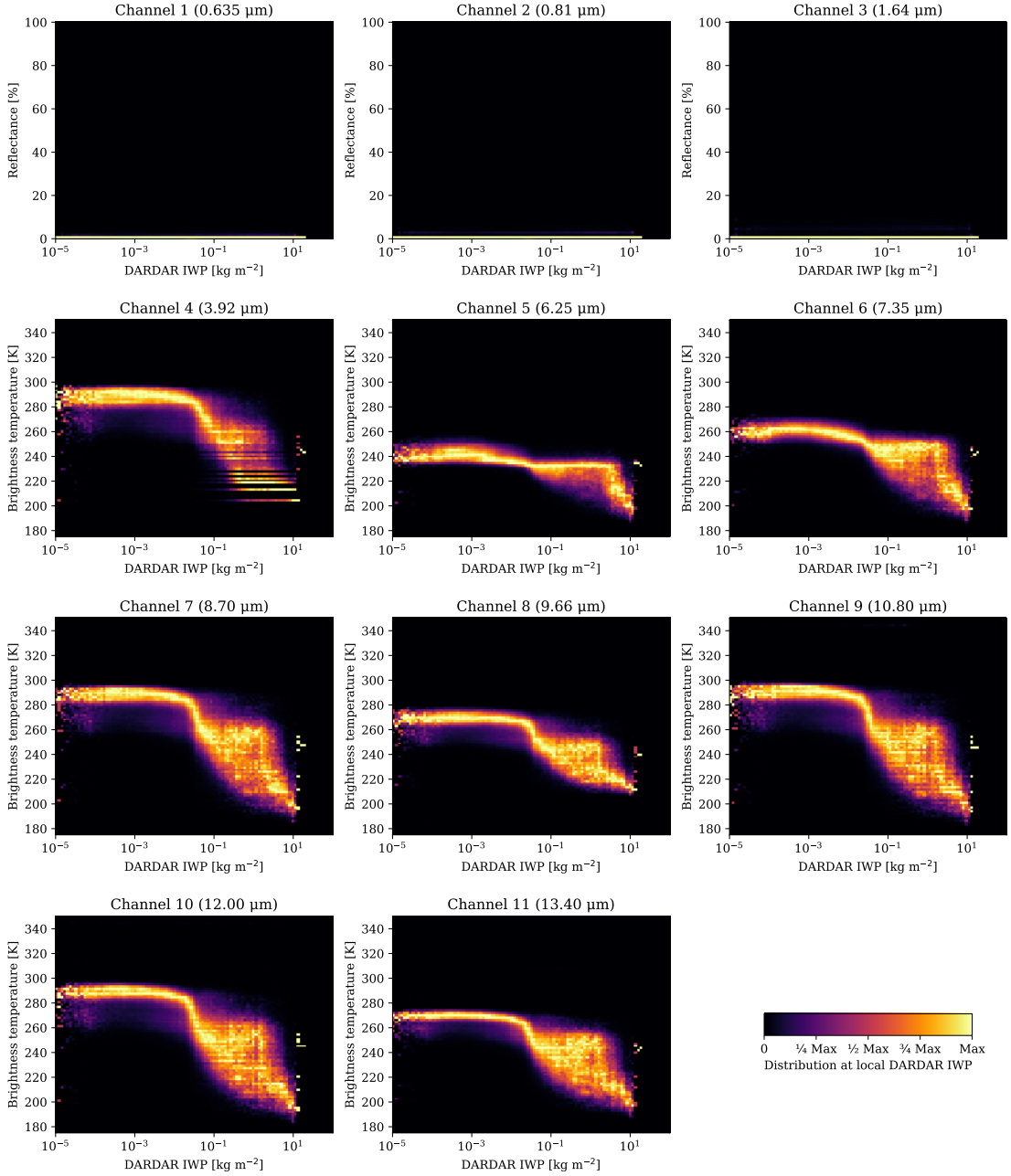


Figure S5: Relationships for the SEVIRI channels and the DARDAR collocated IWP for nighttime data only.

S2 Other Figures

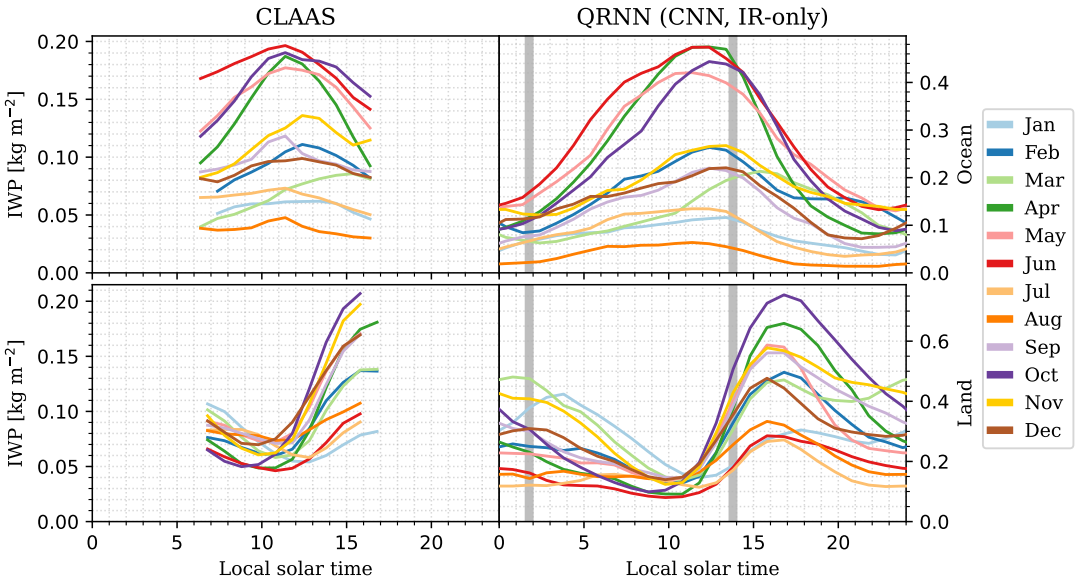
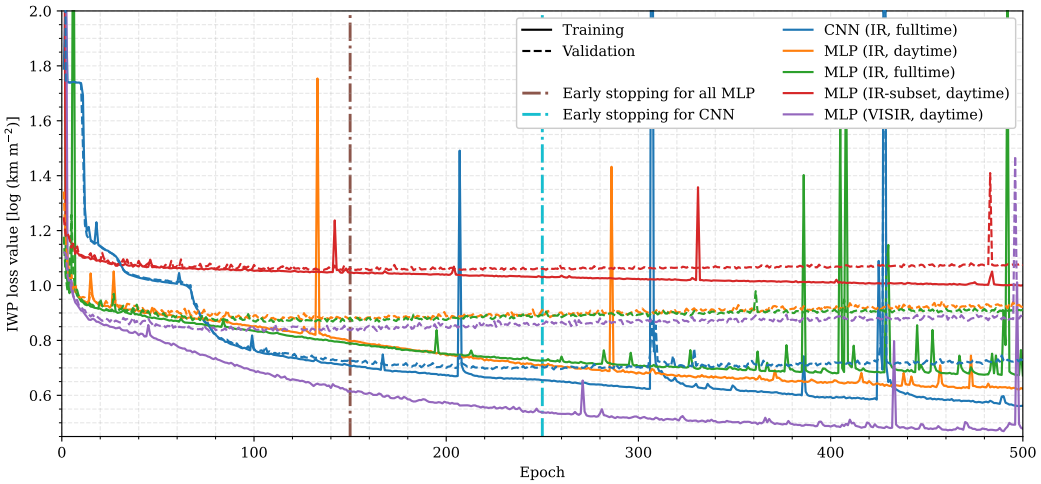
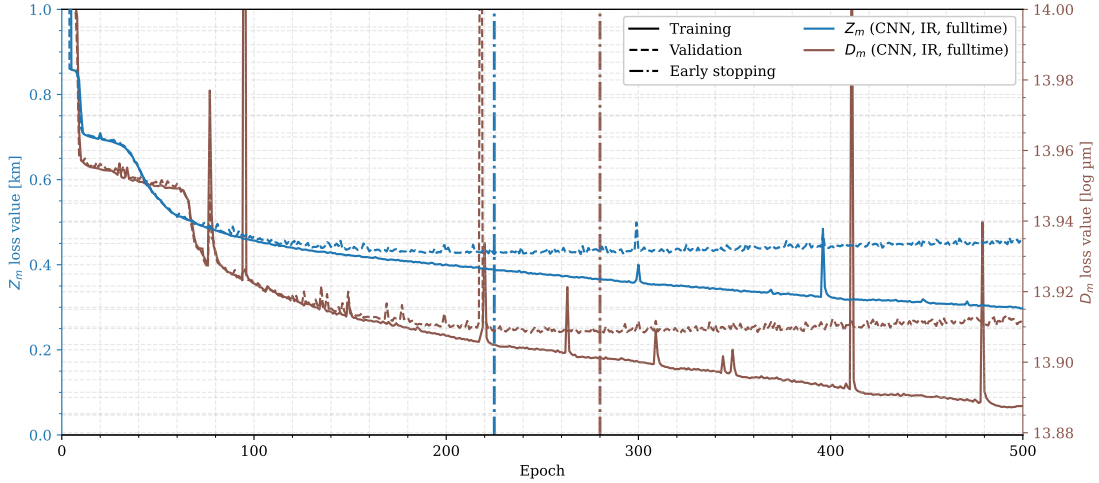


Figure S6: Monthly mean diurnal cycles for 2012 from CLAAS and the CNN, IR-only QRNN retrieval for the areas in Fig. 2. The QRNN expected value is used as the retrieval value. Local solar time (LST) approximated from UTC as LST = UTC + 12 hours/180° · longitude. The grey areas indicate the LST coverage of the DARDAR profiles in the training set. Note the different vertical axes.



(a) Learning curves for the IWP.



(b) Learning curves for the Z_m and D_m.

Figure S7: Learning curves from all networks presented in the paper. Fulltime indicates that the network is trained and validated on daytime and nighttime data. Note that the loss units are a consequence of the transforms applied to the data. The bursts in the loss values were identified to be likely related to the optimizer chosen and the values of its hyperparameters. Given the satisfactory results, we considered any further analysis out of the scope of this work.