Exploring the potential of thermal infrared remote sensing to improve a snowpack model through an observing system simulation experiment

Esteban Alonso-González[1], Simon Gascoin[1], Sara Arioli[2], Ghislain Picard[2]

[1]Centre d'Etudes Spatiales de la Biosphère, Université de Toulouse,CNRS/CNES/IRD/INRA/UPS, Toulouse, France
[2]Univ. Grenoble Alpes, CNRS, Institut des Géosciences de l'Environnement (IGE), Grenoble, France

Correspondence: Alonso-Gonzalez E.: esteban.alonso-gonzalez@univ-tlse3.fr

Abstract. The assimilation of data from Earth observation satellites into numerical models is considered as the path forward to estimate snow cover distribution in mountain catchments, providing accurate information on the mountainous snow water equivalent (SWE). The land surface temperature (LST) can be observed from space, but its potential to improve SWE simulations remains underexplored. This is likely due to the insufficient temporal or spatial resolution offered by the current thermal infrared (TIR) missions. However, three planned missions will provide global-scale TIR data at much higher spatio-temporal resolution in the coming years.
To investigate the value of TIR data to improve SWE estimation, we developed a synthetic data assimilation (DA) experiment at five snow-dominated sites covering a latitudinal gradient in the northern hemisphere. We generated synthetic true LST and SWE series by forcing an energy-balance snowpack model with the ERA5-Land reanalysis. We used this synthetic true LST to recover the synthetic true SWE from a degraded version of ERA5-Land. We defined different observation scenarios to emulate the revisiting times of Landsat 8 (16 days) and the Thermal infraRed Imaging Satellite for High-resolution Natural resource Assessment (TRISHNA) (3 days), while accounting for cloud cover. We replicated the experiments 100 times at each experimental site to assess the robustness of the assimilation process with respect to cloud cover under both revisiting scenarios.. We performed the assimilation using two different approaches: a sequential scheme (particle filter) and a smoother (particle batch smoother).
The results show that LST DA using the smoother reduced the normalized Root Mean Square Error (nRMSE) of the SWE simulations from 61% (open loop) to 17% and 13% for 16 day revisit and 3 day

revisit respectively, in the absence of clouds. We found similar but higher nRMSE values by removing observations due to cloud cover but with a substantial increase of the standard deviation of the nRMSE of the replicates, highlighting the importance of revisiting times in the stability of the assimilation performance. The smoother largely outperformed the particle filter algorithm, suggesting that the capability of a smoother to propagate the information along the season is key to exploit LST information for snow modeling. Finally, we have compared the benefit of assimilating LST with synthetic observations of fractional snow cover area (FSCA). LST DA performed better than FSCA DA in all the study sites, suggesting that the information provided by LST is not limited to the duration of the snow season. These results suggest that the LST data assimilation has an underappreciated potential to improve snowpack simulations and highlight the value of upcoming TIR missions to advance snow hydrology.

1 Introduction

The seasonal snowpack plays a key role in many ecological and hydrological processes worldwide (Barnett et al., 2005). Due to its high albedo and insulating capabilities, the extensive snow-covered area of the northern hemisphere influences the Earth climate system (Henderson et al., 2018). In mountain regions, the seasonal snowpack is also an important source of runoff during the summer when the water demand peaks. Hence, an accurate knowledge of the snowpack conditions has an important economic value (Sturm et al., 2017). The snow cover is also a source of natural hazards such as floods caused by rain-on-snow events or snow avalanches, events that are expected to increase due to the impacts of climate warming (Ballesteros-Cánovas et al., 2018; Musselman et al., 2018).

Despite its importance, monitoring the snowpack remains challenging for both scientists and water management agencies. The variable nature of the snowpack makes it challenging to deploy and maintain maintain dense enough ground based (Kinar & Pomeroy, 2015). Therefore, snow hydrologists have developed methods to take advantage of satellite remote sensing (Fayad et al., 2017; Condom et al., 2020) since the beginning of the Landsat programme (Rango & Martinec, 1979). Yet, the application of remote sensing in snow hydrology remains hindered by the lack of direct observations of the snow water equivalent (SWE) in mountain regions (Dozier et al., 2016). Numerical models allow simulating many snowpack state variables, including the SWE. However, their accuracy is primarily constrained by the large uncertainty of the meteorological forcing

(Raleigh et al., 2015). Recent studies suggest that data assimilation (DA) of remotely sensed products is the path forward to estimate the spatial distribution of relevant snowpack characteristics (Aalstad et al., 2018; Charrois et al., 2016; Cortés & Margulis, 2017; Margulis et al., 2016; Smyth et al., 2020; Stigter et al., 2017). Using DA techniques it is possible to fuse model simulations and multiple remote sensing datasets to improve the snowpack simulations. In particular, the snow cover area was assimilated in many case studies due to its widespread availability (e.g. Baba et al., 2018, Alonso-González et al., 2021). Yet, the extent of the snow cover provides no direct information on the internal state of the snowpack and is blind to snowpack changes when the pixel is fully snow covered (De Lannoy et al., 2012).

The ice-surface temperature (IST), is a key state variable for simulating the snowpack evolution. Physically-based snowpack models solve the energy balance equation iteratively along the time dimension, estimating the IST at each timestep of the model (Essery, 2015; Liston & Elder, 2006). Also it is a key parameter to estimate the emitted energy as outgoing longwave radiation. The estimation of the IST allows detecting the occurrence of surface melting events, as when the IST reaches 0°C all the added energy is converted to melt. Thus, assimilating IST may provide key information about the timing of the melting events. Also, the assimilation of the land surface temperature (LST) (i.e. the temperature of the earth surface independently if it is snow covered) may improve the snowpack simulations by different mechanisms. The IST is physically bound to the melting point temperature, since once the snow melts, the LST can exceed 0°C. Thus the assimilation of the LST may indirectly provide information about the snow cover area too. Also, it should be possible to improve the snow simulations by retrieving thermal information when there is no sunlight, like during the night time or during the polar night at high latitudes. A previous study showed that IST DA could potentially improve the surface ice mass balance simulations of the Greenland ice sheet (Navari et al., 2016), fusing synthetic IST estimations with the CROCUS snow model. On the other hand, previous research suggested little improvements in the mass simulations after assimilating LST simulations retrieved from the Meteosat Second Generation (MSG) (~6km spatial resolution) in the Alps (Piazzi et al. 2019). Also, the rapid variation of LST at hourly timescales can make it difficult to assimilate this variable using particle filters (Piazzi et al. 2018). However, the coarse resolution of the LST products of MSG prevents its use in complex terrain, and the rapid variations in the LST could be addressed using different algorithms. Thus, more research is needed to assess the potential of high resolution LST DA.

The thermal imagery already available only offers coarse resolution for the snow applications over

90 complex terrain (MODIS, Sentinel-3) or long revisiting times (Landsat). This has probably prevented the study of the impact of LST DA, although recent research suggests that LST can provide useful information to retrieve internal snowpack properties (Colombo et al., 2019), a capability that can be exploited from satellites (Colombo et al., 2023). The availability of high spatiotemporal resolution LST products will be improved in the short term with the appearance of new satellites,

95 such as the French-Indian mission Thermal infraRed Imaging Satellite for High-resolution Natural resource Assessment (TRISHNA) (Lagouarde et al., 2018). TRISHNA is expected to provide surface temperature measurements at 60 m spatial resolutions every 3 days at the equator, with an increasing revisiting time towards the poles. Also, given the agenda of the space agencies high resolution thermal infrared retrievals will be readily accessible in the near future. For instance the

100 LSTM (Copernicus Land Surface Temperature Monitoring) (Koetz et al., 2018) which will offer similar observations as TRISHNA but with improved spectral, spatial and temporal resolutions and the SBG satellite (Surface Biology and Geology)(Cawse-Nicholson et al., 2021) from NASA which will provide similar high resolution TIR images of the surface of the earth. The combination of these three missions may eventually provide close to bi-daily (day and night time) high resolution

105 thermal infrared observations of the earth surface. In this context the objectives of this work are; i) to test the potential of the LST to improve the snowpack simulations and ii) explore the effect of increasing the temporal resolution of the observations.

A convenient approach to emulate future remote sensing observations is to use an Observing System Simulation Experiment (OSSE). For example, Navari et al. (2016) assessed the feasibility to

110 integrate ice surface temperatures in a regional climate model estimate of Greenland ice sheet surface mass balance through an OSSE. Synthetic experiments were also used to explore the potential of data assimilation techniques in improving the snowpack simulations (Clark et al., 2006; Revuelto et al., 2021; Smyth et al., 2019).

Here, we designed an OSSE to evaluate the benefit of future remote sensing LST to simulate

115 seasonal SWE. In this experiment, synthetic LST and SWE data were generated in several climatic regions. Synthetic LST data were assimilated into a snowpack model under different cloud cover scenarios and satellite revisit times. The benefit of assimilating LST was studied by comparing the posterior (after assimilation) SWE to the synthetic SWE.

2 Data and methods

2.1 Synthetic observations

We selected five sites in snow-dominated regions of Europe, spanning 40° of latitude from the Pyrenees mountains to the Svalbard archipelago. The sites were chosen every 10° of latitude approximately to sample different climatic influences. The southern sites (Gerlachovský štít, Bigorre) are located in mountain regions (Tatra, Pyrenees), Finse is located on a elevated plateau (Hardangervidda), whereas Tromsø and Ny-Ålesund are in the polar circle. Gerlachovský štít is in Eastern Europe and its climate is influenced by its continental characteristics, Tromsø and Ny-Ålesund exhibit obvious polar climates and Bigorre shows a montane climate with Mediterranean influences.

Table 1: Geographical coordinates and elevation (m a.s.l.) of the ERA5-Land centroids used in the study

Centroid	Longitude	Latitude	Elevation
Ny-Ålesund (Norway)	12.0°	78.9°	124
Tromsø (Norway)	19.8°	69.5°	647
Finse (Norway)	07.5°	60.6°	147
Gerlachovský štít (Slovakia)	20.2°	49.2°	1479
Bigorre (France)	00.1°	42.9°	1843

We used ERA5-land surface reanalysis data (~9km spatial resolution) (Muñoz-Sabater et al., 2021) to force the Flexible Snow Model (FSM2) (Essery, 2015) over four consecutive hydrological years from 01/Sep/2017 to 31/Aug/2021. From this simulation, we retrieved the SWE and LST time series, which were used as the synthetic truth. The LST was exported at 13:00 local time, corresponding to foreseen TRISHNA overpass time. To mimic observational noise, we added to the LST time series a Gaussian noise with a zero mean and a standard deviation of 1.5 K. This standard deviation was chosen as an intermediate value between the reported root mean squared error (RMSE) obtained by the comparison of Landsat 8 with in situ measurements of the snow surface temperature (RMSE = 2.0 °K) (Robledano et al., 2022) and the expected performance of LST products delivered by the TRISHNA mission.

The synthetic true LST time series were downsampled with a period of 16 and 3 days, to emulate revisit times of Landsat 8 and TRISHNA respectively. We simulated the impact of cloud cover on data availability by further removing values in the synthetic LST time series at random dates selected from an uniform distribution. We defined four different cloud cover scenarios with probabilities of 0%, 25%, 50% and 75%. Following the same strategy we have generated fractional snow cover area (FSCA) synthetic true observations to be assimilated. The synthetic FSCA observations were degraded by adding random Gaussian noise with mean 0 and a standard deviation of 0.17 based on Aalstad et al. (2020).

For each site, we created a new degraded meteorological forcing to run FSM2. First we averaged the ERA5-Land data from the nearest 9 cells (i.e. resampling the spatial resolution from 10 km to 30 km approximately). Then, we added autocorrelated Brownian noise in 12h time windows (the data assimilation window of ERA5) using the standard deviation of the variable itself (Supplementary Figure 1). We further perturbed the precipitation field after aggregation dividing the precipitation by 2. This strong perturbation was chosen to emulate precipitation biases that are typically found in global reanalyses and large scale precipitation products (Beck et al., 2019) potentially leading to large underestimation of SWE in mountain regions (Wrzesien et al., 2019). A similar value was already used in previous snow data assimilation experiments (Deschamps-Berger et al., 2022).

2.2 Data assimilation experiments

The degraded meteorological forcing and synthetic LST were used to feed the Multiple Snow data Assimilation system (MuSA). MuSA is an open source ensemble based data assimilation toolbox built around the FSM2 model (Alonso-González et al., 2022). We used the same initial conditions to run FSM2 within MuSA (soil temperature profile, initial LST and absence of snow), therefore we did not perform a spin-up. To emulate the differences between the reality and modeling pipeline, we used a simplified parameterization scheme of FSM2 in MuSA (Table 2). The assimilation experiments were done using the Particle Batch Smoother (PBS) (Margulis et al., 2015) and a particle filter. Smoother algorithms are typically used to develop reanalyses as the whole time series of information are available, whereas filtering is rather used for operational forecasting where future observations respective to the analysis step are not available (Largeron et al., 2020). A description of these algorithms and the MuSA toolbox can be found in Alonso-González et al., (2022).

We performed a final experiment in which we assimilated fractional snow cover area (FSCA), using a PBS. This experiment has been performed using exactly the same set up as for the assimilation of LST. This includes running the simulations over the same geographical areas, using the same forcing and cloud cover scenarios, as well as using a different FSM2 parameterisation to generate the synthetic truth and MuSA. Using exactly the same setup allows a simple comparison between the performance of assimilating LST and FSCA. .

Table 2: FSM2 configuration chosen (and configuration number) to generate the synthetic truth and simulations.

Process	Syntetic true FSM2	MuSA FSM2
Snow albedo	[2] Diagnosed from snow age	[1] Diagnosed from LST
Snow thermal conductivity	[1] Estimated from density	[0] Constant thermal conductivity
Snow density	[2] Viscous compaction	[0] Constant density
Turbulent exchange	[1] Monin-Obukhov stability functions	[0] Neutral stability
Snow hydraulics	[2] Gravitational drainage	[1] Bucket storage
Snow cover fraction	[2] Asymptotic function	[1] Linear function

The prior ensemble of FSM2 simulations was composed of 300 particles that were generated by perturbing the air temperature (additive perturbation) and the precipitation (multiplicative perturbation). The perturbations were time-invariant and randomly drawn from a normal distribution of mean $\mu = 0$ and standard deviation $\sigma = 2$ (temperature) and a log-normal distribution with mean $\mu = 0.45$ and standard deviation $\sigma = 0.8$ of the underlying normal distribution (precipitation). These parameters were chosen to cover the expected differences between the "truth" forcing and the degraded forcing (Supplementary Figure 2) and were obtained by preliminary trial and error tests. Both PF and PBS are very prone to collapse. Although there are uncertainties in other variables, keeping a limited number of dimensions helps to prevent the collapse of the ensembles. In addition, not correcting for the other variables introduces errors in the simulations, which also helps to prevent the collapse of the ensembles. The observation errors prescribed for the synthetic true observations were the same as those defined to generate the degradation Gaussian noise.

While the PF performs the analysis sequentially, i.e. each time an observation occurs, the PBS is a smoother, hence it assimilates all the available observations in a single time window, propagating information from the observations forward and backward in time. Here the assimilation time window was defined as a hydrological year (i.e. one snow season). In both the PF and PBS, prior weights of ensemble members (particles) are updated based on the likelihood, i.e. a measure of

the distance between the predictions of each particle and the observations. The posterior weights are then used to estimate posterior statistics from the ensemble, typically its weighted mean and weighted standard deviation. In the case of the PF, we used the bootstrap resampling algorithm to eliminate particles with low weights and replicate particles with high weights by sampling with replacement randomly from the probability distribution of the updated weights. To prevent the filter from collapsing (all the weight is shared by few and eventually just one particle), new perturbation parameters were drawn from a normal approximation of the posterior from the previous analysis step at each new analysis step, instead of resampling both the states of the model and the parameters. A rigorous description of the algorithms, the underlying theory and implementation details can be found in Alonso-González et al., (2022).

2.3 Evaluation of the experiments

For each site and each cloud cover scenario, we ran MuSA and generated a posterior SWE. However, the output of MuSA is stochastic due to the random generation of the forcing perturbation parameters. Also the position of the gaps in the different cloud cover scenarios and the Gaussian and Brownian noises added to the observations and forcing respectively are random. Therefore, to increase the statistical robustness of our results, we repeated each assimilation experiment 100 times, drawing new gaps and noise for each replicate of the experiment. This created an ensemble of posterior SWE which was compared to the synthetic true SWE.

In total, for a given site, MuSA was run 2400 times (100 replicates x 4 cloud cover scenarios x 2 revisit times x 3 DA experiments). This corresponds to 720,000 FSM2 runs (300 particles by MuSA run), summing up to 3,600,000 FSM2 runs considering the 5 study areas, which is equivalent to simulating the snowpack over a period of 14,400,000 years. To compare the performance of data assimilation, in every site regardless of the absolute magnitude of the SWE, we used the normalized RMSE (nRMSE) as a performance score:

$$nRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Pred_i - Ref_i\right)^2}}{\overline{Ref}}$$

where n is the number of samples, $Pred$ the predicted and Ref the reference SWE values.

3 Results

As expected, the degradation of the ERA5-Land meteorological forcing had a large impact on the open loop SWE simulations (Fig. 2 and supplementary figures 3, 4 and 5). In comparison with the synthetic true SWE, the average normalized RMSE (nRMSE) nRMSE was 61% (after removing the
230 summer time months of July and August). The degradation of the ERA5-Land data caused an overall reduction in the simulated SWE, leading to a shorter snow seasons at all sites (Fig. 2). However, the LST assimilation with the PBS substantially improved the SWE simulations (Fig. 2). This improvement was evident at both revisit times, although the 3 day revisit scenario (nRMSE = 13%) outperformed the 16 day revisit scenario (nRMSE = 17%).
235 The posterior SWE series in Figure 2 were averaged from an ensemble of 100 replicates under clear sky conditions. The uncertainty in the replicates is caused mostly by the different random noises in the observations (see Sect. 2. Data and methods). Figure 3 shows every posterior SWE realization in the case of the Tromsø site under different cloud cover scenarios when assimilating LST at 3-day resolution. This figure shows that the spread of the replicates increased with the cloud
240 cover probability. The standard deviation of the obtained nRMSE values over each cloud cover scenario ranged from 4% to 10% in this particular case.

Figure 4 summarizes the results of the PBS from all experiments under every cloud cover and revisit scenarios. In all cases, the data assimilation significantly reduces the nRMSE in comparison with the open loop simulations. In most situations the nRMSE is always higher for the 16-day
245 revisit compared to the 3-day revisit, but the difference was more pronounced under the 50% and 75% cloud cover scenarios. In addition, the standard deviation of the nRMSE of the 100 replicates is higher for the 16-day compared to the 3-day revisit scenarios. Both the averaged nRMSE and standard deviations increase with the cloud cover with an average nRMSE for all the sites ranging between 13% and 16% in the case of the 3-day revisit experiments and 17% and 28% for the 16-
250 day revisit experiments. The results indicate a comparatively lower reduction in SWE nRMSE when FSCA is assimilated compared to LST assimilation ranging between 27% and 39% in the case of the 3-day revisit experiments and 41% and 45% for the 16-day revisit experiments (Figure 5), being comparable to results obtained in previous studies assimilating real observations (Aalstad et al., 2018) and thus confirming the robustness of the OSSE design. Furthermore, the greater dispersion
255 of nRMSE values suggests an   increased sensitivity to cloud distribution in the case of FSCA compared to LST.

Figure 6 shows the distribution of the mean of the posterior precipitation perturbation parameters

obtained from the 100 data assimilation runs using the PBS. It demonstrates that the assimilation of LST reduces the error on the precipitation forcing, since the posterior parameter distributions approximates the actual multiplicative perturbation factor of 2 to compensate the 0.5 scaling factor used to degrade the input precipitation. However, the difference between the true precipitation and the degraded precipitation should not be exactly equal to the scaling factor of 2 due to the perturbation strategy used, the fact that we do not correct all forcing variables, and the differences between the FSM2 configurations (Sect. 2), and other components of the forcing were also degraded and not included in the simulation to induce errors in the FSM2 model. As observed above in Figure 4, the standard deviation of the posterior perturbation parameters of the replicates increases when comparing the 3-day with the 16-day revisit scenarios, as well as with the cloud cover probability.

Whereas the above results show that the PBS algorithm clearly improved the SWE simulation, it was not the case with the PF. Figure 7 summarizes the results of the same experiments shown in Figure 4, but using the PF instead of the PBS. In this case the improvement in the average nRMSE of the posterior simulations is not as obvious with respect to the open loop as in the PBS experiments. Although, in the medium latitude cases the nRMSE shows a moderate improvement compared with the open loop on average, several runs among the 100 replicates had a higher nRMSE than the open loop run. The revisit or cloud cover scenarios had no clear effect on the nRMSE. The results in the high latitude areas yielded a higher nRMSE standard deviation than for the medium latitude regions. This is a consequence of the very cold conditions in this high latitude areas, causing that some particles became "glaciers" due to the perturbed forcing (non zero SWE at the end of the hydrological year).

4 Discussion

Navari et al. (2016) showed the potential of IST data assimilation to improve the surface mass balance of the Greenland ice sheet in a regional climate simulation with an Ensemble batch smoother. Our study also suggests that the assimilation of LST can improve seasonal snow simulations in sites with different climate contexts. With the PBS, the improvement was substantial independently of the site, i.e. the climatic context did not exhibit an obvious influence on the results. However, our results with the PF also supports the conclusions of Piazzi et al. (2019), who did not obtain obvious improvements in the posterior SWE simulations after assimilating LST using

an Ensemble Kalman Filter. Therefore, our study provide an explanation of the contrasting performances found by Navari et al. (2016) and Piazzi et al. (2019). While Navari et al. (2016) used smoothers, Piazzi et al. (2019) used a filter. A filter updates the simulations sequentially while smoothers update the whole season in batch. This in-batch assimilation allows the propagation of the information of the observations backwards in the simulation time. Also, the performance of the LST data assimilation reported by Piazzi et al. (2019) was probably hampered by the coarse resolution of the MSG LST products that were used to update snowpack simulations at the point scale. In the specific case of the LST, considering the observations of the whole snow season in a batch may be key to have a positive impact on the posterior SWE. The trajectory of the LST in seasonal snow dominated regions exhibits a characteristic pattern as the physical bounds of the IST are different from the LST. Once the snow melts, the LST can rise above the water melting point and therefore the trajectory of the LST may be a good indicator of the length of the snow season. However, this should not be the only reason, as Navari et al. (2016) experiments were developed over the Greenland ice sheet where there is a permanent ice cover. During the melting season the IST is fixed to the melting point temperature, providing information on the duration of the melting period. Also the occurrence of wintertime melt events should be visible in the TIR domain. The average performance of the LST assimilation was superior in all study areas in our experiments. This also suggests that LST may provide more information than FSCA. In addition, the lower dispersion of nRMSE values suggests that LST may be more resilient to cloud cover distribution than FSCA.The information of the whole seasonal trajectory of the LST is propagated to the posterior by using a smoother, but not by using a filter. This is highlighted at the high latitude study areas, where the polar conditions made snowmelt impossible at the end of the hydrological year for some of the replicates (Suplementary figure 2) leading to very high nRMSE using the PF (Figure 7). These results suggest that the LST may be less beneficial to snowmelt forecasting applications where the use of filters is more extended to update the model as new observations arise, but it should be valuable information to improve snow reanalyses which aim to reconstruct snow cover climatologies.

Our results also suggest that even the currently available thermal infrared estimations of the LST from Landsat missions have the potential to significantly improve SWE simulations despite a revisit time of 16 days. The emulated revisiting times of both Landsat and TRISHNA are the expected values at the equator, and can be lower in other latitudes. Here we did not study the effect of the spatial resolution but hypothesized that high resolution (i.e. Landsat-like) is needed for snow cover

320 simulations in the studied regions. Landsat TIR images have a 100 m resolution which makes them suitable to sample the slope scale in mountain terrain, hence homogeneous conditions in the energy balance budget (Baba et al., 2019). Despite the low revisiting times of the Landsat mission, Landsat TIR imagery may be useful to improve SWE simulations using a smoother data assimilation algorithm, an approach that to our knowledge has not been explored yet. More research should be

325 carried out on this topic, especially in the context of joint assimilation experiments where more than one variable is assimilated.

Nevertheless, the change in revisit from 3 to 16 days in our experiments translated into a higher posterior nRMSE. Therefore, we expect significant progress with TRISHNA observations, not to

330 mention the enhanced spatial resolution (approximately 60 m). The benefit of the 3 day revisit was particularly evident under 50% to 75% cloud cover. This should be considered, as previous global estimates of the cloud cover suggest values closer to our highest cloud cover scenario (Wylie et al., 2005). For instance, cloud cover probability in MODIS products reached 60% in the Alps and 50% in the Pyrenees (Gascoin et al., 2015; Parajka & Blöschl, 2008).

335 In any case, under both revisit scenarios, the cloud cover decreased the precision of the replicates of the posterior SWE, i.e. the variability between repeated experiments, but the average was only marginally affected. In other words, the cloud cover reduced the robustness of the data assimilation, but even regions with a persistent cloud cover could benefit from LST assimilation. The different replicates of each experiment exhibited different results, with a variance that

340 increased with the number of gaps introduced in the synthetic LST observations, suggesting that not all the combinations of observations are equally informative. This was also obvious regarding the posterior precipitation perturbation parameters, as the standard deviation of the different replicates increased with the percentage of cloud cover.

Despite the promising potential of the LST to improve SWE simulation, some limitations of the

345 current study inherent to the synthetic nature of the OSSE should be taken into consideration. The synthetic nature of the experiment could lead to an overestimation of the value of LST assimilation. This effect is mitigated in our experiment as typically done in OSSE's by (I) the degradation of all forcing variables, while only temperature and precipitation were corrected and (ii) the different FSM2 parameterisations used to create the synthetic truth. The simulation of the cloud cover

350 scenarios was generated by selecting random dates from an uniform distribution. However, in some regions the cloud cover exhibits marked seasonal patterns (Sudmanns et al., 2020), that may

challenge to update the snowpack simulations even with smoothers if cloud cover is more frequent during key periods in the snow season. Also, the surface temperature observation in complex terrain may differ from the simulated temperature, due to intra-pixel variability as a consequence

355 of variable snow and/or mixed pixel conditions (Robledano et. al., 2022, Lundquist et al., 2018). But this issue is greatly limited by the expected increase in resolution. In light of the results of this work, the next step is to conduct experiments using real remote sensing observations, although the general lack of data may complicate the interpretation of the results.

5 Conclusions

360 The motivation of this study on LST data assimilation is the upcoming launch of high resolution thermal infrared spatial missions with improved revisit time and resolution in the next few years. We implemented a synthetic data assimilation experiment to study the potential of the LST in improving SWE simulations along a latitudinal gradient in the northern hemisphere. The methodology was based on the generation of synthetic LST estimations and SWE true estimates,

365 and a prior ensemble of SWE simulations generated by forcing the FSM2 model with degraded meteorological fields. The MuSA snow data assimilation software was used to generate SWE posterior time series using the particle batch smoother and particle filter algorithms to be compared with synthetic true SWE.

The results suggest that the assimilation of LST has a great potential to improve seasonal

370 snowpack simulations across all the tested sites. Gap-free LST series, improved the average nRMSE of the open loop simulations from 61% to 17% and 13% for the 16 days and 3 day revisiting times respectively. However, a lower revisit frequency caused an increase in the variance of the nRMSE when the runs are replicated 100 times, showing that the performance of the assimilation would depend on cloud cover scenarios. This conclusion was more evident with high cloud cover

375 scenarios, highlighting the importance of the revisit time in thermal infrared remote sensing to reduce the uncertainty on the updated SWE.

The type of data assimilation was also key to explain the role of LST in improving SWE simulations. The particle batch smoother strongly improved the simulations, whereas the particle filter was much less performant and could even cause a degradation of the simulations. This effect could be

380 interpreted as a consequence of the strong seasonal signal of the LST, which reflects the duration of the snow season. But the lower performance shown by the FSCA assimilation suggests that the

LST assimilation provides more than just information on snow duration.

Overall, our results encourage a more systematic use of the current LST products within snow data assimilation studies, especially if the objective is to perform a snow reanalysis which can benefit
385   from observations acquired over an entire snow season.

Acknowledgements

This work was supported by the Centre National d'Etudes Spatiales (CNES) through the postdoctoral grant of Esteban Alonso-González, the PhD grant of Sara Arioli and the TOSCA program.

390   **Code and data availability.**

The MuSA v1.0 code is available at https://github.com/ealonsogzl/MuSA/tree/v1.0. The original FSM2 code is found at github.com/RichardEssery/FSM2, and in the MuSA repository with slight modifications from the original version. ERA5-Land data is available for download from the Copernicus Climate Data Store.

395

Author contributions. Conceptualization: EAG, SG. Data curation: EAG. Formal analysis: EAG, SG, SA, GP. Funding acquisition: EAG, SG. Investigation: EAG, SG. Methodology: EAG. Project administration: EAG, SG. Resources:EAG, SG. Software: EAG. Supervision: SG, GP. Validation: EAG.
400   Visualization: EAG. Writing - original draft preparation: EAG . Writing review & editing: EAG, SG, SA, GP.

Competing interests. The authors declare that they have no conflict of interest.

References

405   Aalstad, K., Westermann, S., Schuler, T. V., Boike, J., & Bertino, L. (2018). Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites. Cryosphere, 12(1), 247–270. https://doi.org/10.5194/tc-12-247-2018

Aalstad, K., Westermann, S., & Bertino, L. (2020). Evaluating satellite retrieved fractional snow-

45

covered area at a high-Arctic site using terrestrial photography. Remote Sensing of Environment, 239 (2020), Article 111618. https://doi.org/10.1016/j.rse.2019.111618.

Alonso-González, E., Gutmann, E., Aalstad, K., Fayad, A., Bouchet, M., & Gascoin, S. (2021). Snowpack dynamics in the Lebanese mountains from quasi-dynamically downscaled ERA5 reanalysis updated by assimilating remotely sensed fractional snow-covered area. Hydrology and Earth System Sciences, 25(8), 4455–4471. https://doi.org/10.5194/hess-25-4455-2021

Alonso-González, Esteban, Aalstad, K., Baba, M. W., Revuelto, J., López-Moreno, J. I., Fiddes, J., et al. (2022). MuSA: The Multiscale Snow Data Assimilation System (v1.0). Geoscientific Model Development Discussions, 1–43. https://doi.org/10.5194/gmd-2022-137

Baba, M. W., Gascoin, S., & Hanich, L. (2018). Assimilation of Sentinel-2 data into a snowpack model in the High Atlas of Morocco. Remote Sensing, 10(12), 1982. https://doi.org/10.3390/rs10121982

Baba, M. W., Gascoin, S., Kinnard, C., Marchane, A., & Hanich, L. (2019). Effect of Digital Elevation Model Resolution on the Simulation of the Snow Cover Evolution in the High Atlas. Water Resources Research, 55(7), 5360–5378. https://doi.org/10.1029/2018WR023789

Ballesteros-Cánovas, J. A., Trappmann, D., Madrigal-González, J., Eckert, N., & Stoffel, M. (2018). Climate warming enhances snow avalanche risk in the Western Himalayas. Proceedings of the National Academy of Sciences of the United States of America, 115(13), 3410–3415. https://doi.org/10.1073/pnas.1716913115

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. Nature, 438(7066), 303–309. https://doi.org/10.1038/nature04141

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., et al. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrology and Earth System Sciences, 23(1), 207–224. https://doi.org/10.5194/hess-23-207-2019

Cawse-Nicholson, K., Townsend, P. A., Schimel, D., Assiri, A. M., Blake, P. L., Buongiorno, M. F., et al. (2021). NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms. Remote Sensing of Environment, 257, 112349. https://doi.org/10.1016/j.rse.2021.112349

Charrois, L., Cosme, E., Dumont, M., Lafaysse, M., Morin, S., Libois, Q., & Picard, G. (2016). On the assimilation of optical reflectances and snow depth observations into a detailed snowpack model. Cryosphere, 10(3), 1021–1038. https://doi.org/10.5194/tc-10-1021-2016

Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B., & Leavesley, G. H. (2006). Assimilation of snow covered area information into hydrologic and land-surface models. Advances in Water Resources, 29(8), 1209–1221. https://doi.org/10.1016/j.advwatres.2005.10.001

Colombo, R., Garzonio, R., Di Mauro, B., Dumont, M., Tuzet, F., Cogliati, S., et al. (2019). Introducing Thermal Inertia for Monitoring Snowmelt Processes With Remote Sensing. Geophysical Research Letters, 46(8), 4308–4319. https://doi.org/10.1029/2019GL082193

Colombo, Roberto, Pennati, G., Pozzi, G., Garzonio, R., Di Mauro, B., Giardino, C., et al. (2023). Mapping snow density through thermal inertia observations. Remote Sensing of Environment, 284, 113323. https://doi.org/10.1016/j.rse.2022.113323

450 Condom, T., Martínez, R., Pabón, J. D., Costa, F., Pineda, L., Nieto, J. J., López, F., and Villacis, M.: Climatological and Hydrological Observations for the South American Andes: In situ Stations, Satellite, and Reanalysis Data Sets, Frontiers in Earth Science, 8, 2020.

Cortés, G., & Margulis, S. (2017). Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: Results from a high-resolution 31 year reanalysis: El Niño Effects on Andes Snow.
455 Geophysical Research Letters, 44(13), 6859–6867. https://doi.org/10.1002/2017GL073826

De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C., and Pauwels, V. R. N. (2012), Multiscale assimilation of Advanced Microwave Scanning Radiometer–EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, *Water Resour. Res.*, 48, W01522, doi:10.1029/2011WR010588.

460 Deschamps-Berger, C., Cluzet, B., Dumont, M., Lafaysse, M., Berthier, E., Fanise, P., & Gascoin, S. (2022). Improving the Spatial Distribution of Snow Cover Simulations by Assimilation of Satellite Stereoscopic Imagery. Water Resources Research, 58(3), e2021WR030271. https://doi.org/10.1029/2021WR030271

Dozier, J., Bair, E. H., & Davis, R. E. (2016). Estimating the spatial distribution of snow water
465 equivalent in the world's mountains. Wiley Interdisciplinary Reviews: Water, n/a-n/a. https://doi.org/10.1002/wat2.1140

Dunne, S., & Entekhabi, D. (2005). An ensemble-based reanalysis approach to land data assimilation. Water Resources Research, 41(2). https://doi.org/10.1029/2004WR003449

Essery, R. (2015). A factorial snowpack model (FSM 1.0). Geoscientific Model Development, 8(12),
470 3867–3876. https://doi.org/10.5194/gmd-8-3867-2015

Fayad, A., Gascoin, S., Faour, G., López-Moreno, J. I., Drapeau, L., Page, M. L., and Escadafal, R.: Snow hydrology in Mediterranean mountain regions: A review, Journal of Hydrology, 551, 374–396, https://doi.org/10.1016/j.jhydrol.2017.05.063, 2017.

Gascoin, S., Hagolle, O., Huc, M., Jarlan, L., Dejoux, J. F., Szczypta, C., et al. (2015). A snow cover
475 climatology for the Pyrenees from MODIS snow products. Hydrology and Earth System Sciences, 19(5), 2337–2351. https://doi.org/10.5194/hess-19-2337-2015

Henderson, G. R., Peings, Y., Furtado, J. C., & Kushner, P. J. (2018). Snow–atmosphere coupling in the Northern Hemisphere. Nature Climate Change, 8(11), 954–963. https://doi.org/10.1038/s41558-018-0295-6

480 Kinar, N. J., & Pomeroy, J. W. (2015). Measurement of the physical properties of the snowpack. Reviews of Geophysics, 53(2), 481–544. https://doi.org/10.1002/2015RG000481

Koetz, B., Bastiaanssen, W., Berger, M., Defourney, P., Del Bello, U., Drusch, M., et al. (2018). High Spatio- Temporal Resolution Land Surface Temperature Mission - a Copernicus Candidate Mission in Support of Agricultural Monitoring. In IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 8160–8162). https://doi.org/10.1109/IGARSS.2018.8517433

Lagouarde, J.-, Bhattacharya, B. K., Crébassol, P., Gamet, P., Babu, S. S., Boulet, G., et al. (2018). The Indian-French Trishna Mission: Earth Observation in the Thermal Infrared with High Spatio-Temporal Resolution (pp. 4078–4081). Presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. https://doi.org/10.1109/IGARSS.2018.8518720

Largeron, C., Dumont, M., Morin, S., Boone, A., Lafaysse, M., Metref, S., et al. (2020). Toward Snow Cover Estimation in Mountainous Areas Using Modern Data Assimilation Methods: A Review. Frontiers in Earth Science, 8. Retrieved from https://www.frontiersin.org/article/10.3389/feart.2020.00325

Liston, G. E., & Elder, K. (2006). A distributed snow-evolution modeling system (snowmodel). Journal of Hydrometeorology, 7(6), 1259–1276. https://doi.org/10.1175/JHM548.1

Lundquist, J. D., Chickadel, C., Cristea, N., Currier, W. R., Henn, B., Keenan, E., and Dozier, J.: Separating snow and forest temperatures with thermal infrared remote sensing, Remote Sensing of Environment, 209, 764–779, https://doi.org/10.1016/j.rse.2018.03.001, 2018.

Margulis, S. A., Girotto, M., Cortés, G., & Durand, M. (2015). A particle batch smoother approach to snow water equivalent estimation. Journal of Hydrometeorology, 16(4), 1752–1772. https://doi.org/10.1175/JHM-D-14-0177.1

Margulis, S. A., Cortés, G., Girotto, M., & Durand, M. (2016). A landsat-era Sierra Nevada snow reanalysis (1985-2015). Journal of Hydrometeorology, 17(4), 1203–1221. https://doi.org/10.1175/JHM-D-15-0177.1

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth System Science Data, 13(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., et al. (2018). Projected increases and shifts in rain-on-snow flood risk over western North America. Nature Climate Change, 8(9), 808–812. https://doi.org/10.1038/s41558-018-0236-4

Navari, M., Margulis, S. A., Bateni, S. M., Tedesco, M., Alexander, P., & Fettweis, X. (2016). Feasibility of improving a priori regional climate model estimates of Greenland ice sheet surface mass loss through assimilation of measured ice surface temperatures. The Cryosphere, 10(1), 103–120. https://doi.org/10.5194/tc-10-103-2016

Parajka, J., & Blöschl, G. (2008). Spatio-temporal combination of MODIS images - Potential for snow cover mapping. Water Resources Research, 44(3). https://doi.org/10.1029/2007WR006204

Piazzi, G., Thirel, G., Campo, L., and Gabellani, S. (2018.). A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an Alpine environment, The Cryosphere, 12, 2287–2306, https://doi.org/10.5194/tc-12-2287-2018,

520   Piazzi, G., Campo, L., Gabellani, S., Castelli, F., Cremonese, E., Cella, U. M. di, et al. (2019). An Enkf-Based Scheme for Snow Multivariable Data Assimilation at an Alpine Site. Journal of Hydrology and Hydromechanics, 67(1), 4–19. https://doi.org/10.2478/johh-2018-0013

Raleigh, M. S., Lundquist, J. D., & Clark, M. P. (2015). Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework.
525   Hydrology and Earth System Sciences, 19(7), 3153–3179. https://doi.org/10.5194/hess-19-3153-2015

Rango, A., & Martinec, J. (1979). Application of a Snowmelt-Runoff Model Using Landsat Data. Nordic Hydrology, 10(4), 225–238. https://doi.org/10.2166/nh.1979.0006

Revuelto, J., Cluzet, B., Duran, N., Fructus, M., Lafaysse, M., Cosme, E., & Dumont, M. (2021).
530   Assimilation of surface reflectance in snow simulations: Impact on bulk snow variables. Journal of Hydrology, 603, 126966. https://doi.org/10.1016/j.jhydrol.2021.126966

Robledano, A., Picard, G., Arnaud, L., Larue, F., & Ollivier, I. (2022). Modelling surface temperature and radiation budget of snow-covered complex terrain. The Cryosphere, 16(2), 559–579. https://doi.org/10.5194/tc-16-559-2022

535   Smyth, E. J., Raleigh, M. S., & Small, E. E. (2019). Particle Filter Data Assimilation of Monthly Snow Depth Observations Improves Estimation of Snow Density and SWE. Water Resources Research, 55(2), 1296–1311. https://doi.org/10.1029/2018WR023400

Smyth, E. J., Raleigh, M. S., & Small, E. E. (2020). Improving SWE Estimation With Data Assimilation: The Influence of Snow Depth Observation Timing and Uncertainty. Water Resources Research,
540   56(5). https://doi.org/10.1029/2019WR026853

Stigter, E. E., Wanders, N., Saloranta, T. M., Shea, J. M., Bierkens, M. F. P., & Immerzeel, W. W. (2017). Assimilation of snow cover and snow depth into a snow model to estimate snow water equivalent and snowmelt runoff in a Himalayan catchment. The Cryosphere, 11(4), 1647–1664. https://doi.org/10.5194/tc-11-1647-2017

545   Sturm, M., Goldstein, M. A., & Parr, C. (2017). Water and life from snow: A trillion dollar science question.

Sudmanns, M., Tiede, D., Augustin, H., & Lang, S. (2020). Assessing global Sentinel-2 coverage dynamics and data availability for operational Earth observation (EO) applications using the EO-Compass.   International   Journal   of   Digital   Earth,   13(7),   768–784.
550   https://doi.org/10.1080/17538947.2019.1572799

Wrzesien, M. L., Pavelsky, T. M., Durand, M. T., Dozier, J., & Lundquist, J. D. (2019). Characterizing Biases in Mountain Snow Accumulation From Global Data Sets. Water Resources Research, 55(11), 9873–9891. https://doi.org/10.1029/2019WR025350

Wylie, D., Jackson, D. L., Menzel, W. P., & Bates, J. J. (2005). Trends in Global Cloud Cover in Two
555   Decades   of   HIRS   Observations.   Journal   of   Climate,   18(15),   3021–3031. https://doi.org/10.1175/JCLI3461.

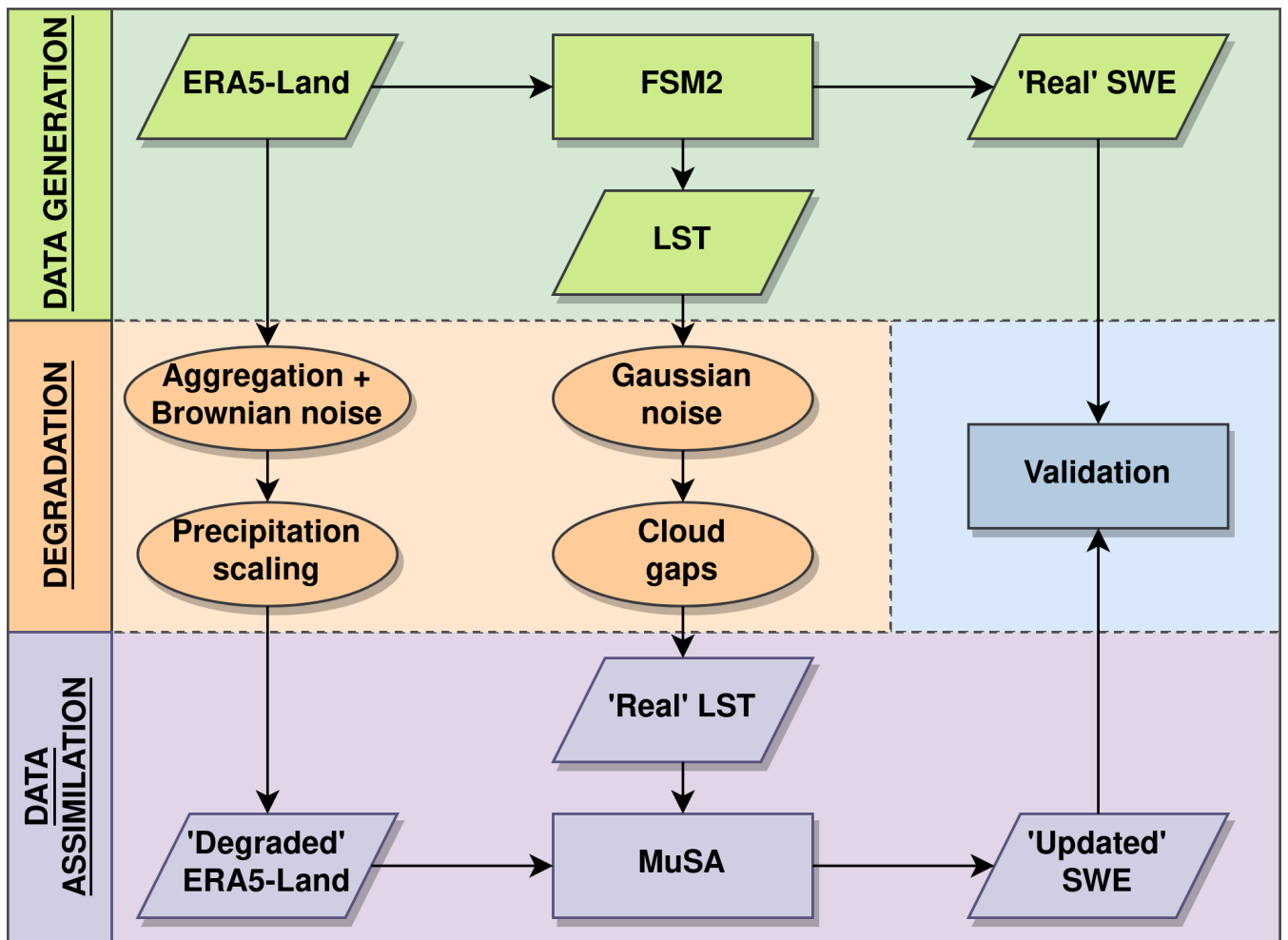



Figure 1: Main workflow of the OSSE.

560

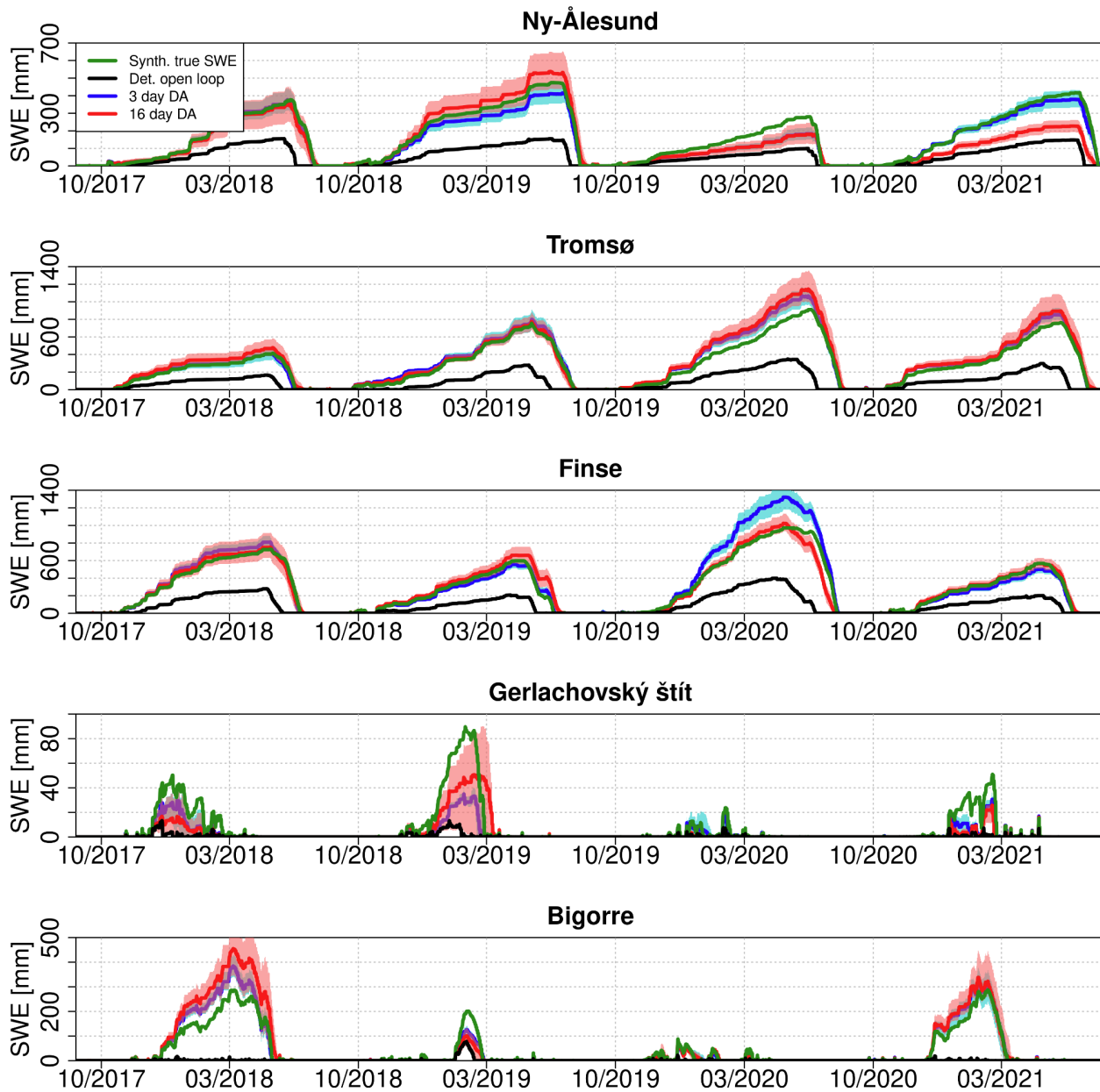



Figure 2: Comparison of hourly time series of synthetic true SWE with the open loop simulations and the posterior SWE after assimilating LST with a revisit time of 16 days or 3 days (0% cloud cover scenario) using the PBS. Here, the posterior SWE is the average of the 100 replicates and the shaded areas represent the 95 to 5th quantile range.

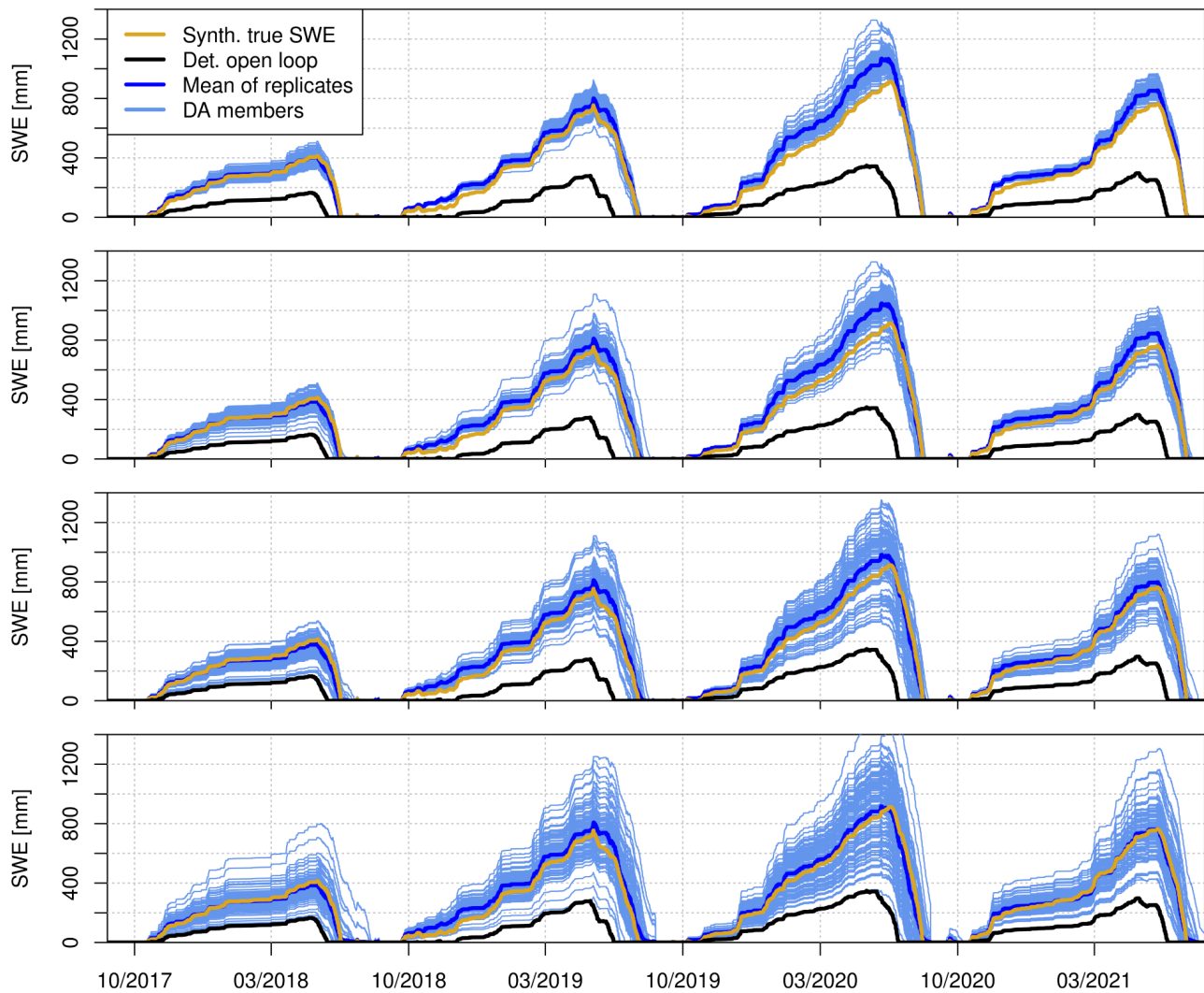



Figure 3: Comparison of hourly time series of synthetic true SWE with the open loop simulation, and the 100 replicates of posterior SWE after assimilating LST with a revisit time of 3 days and varying cloud cover scenarios and the average of the experiments in Tromsø.

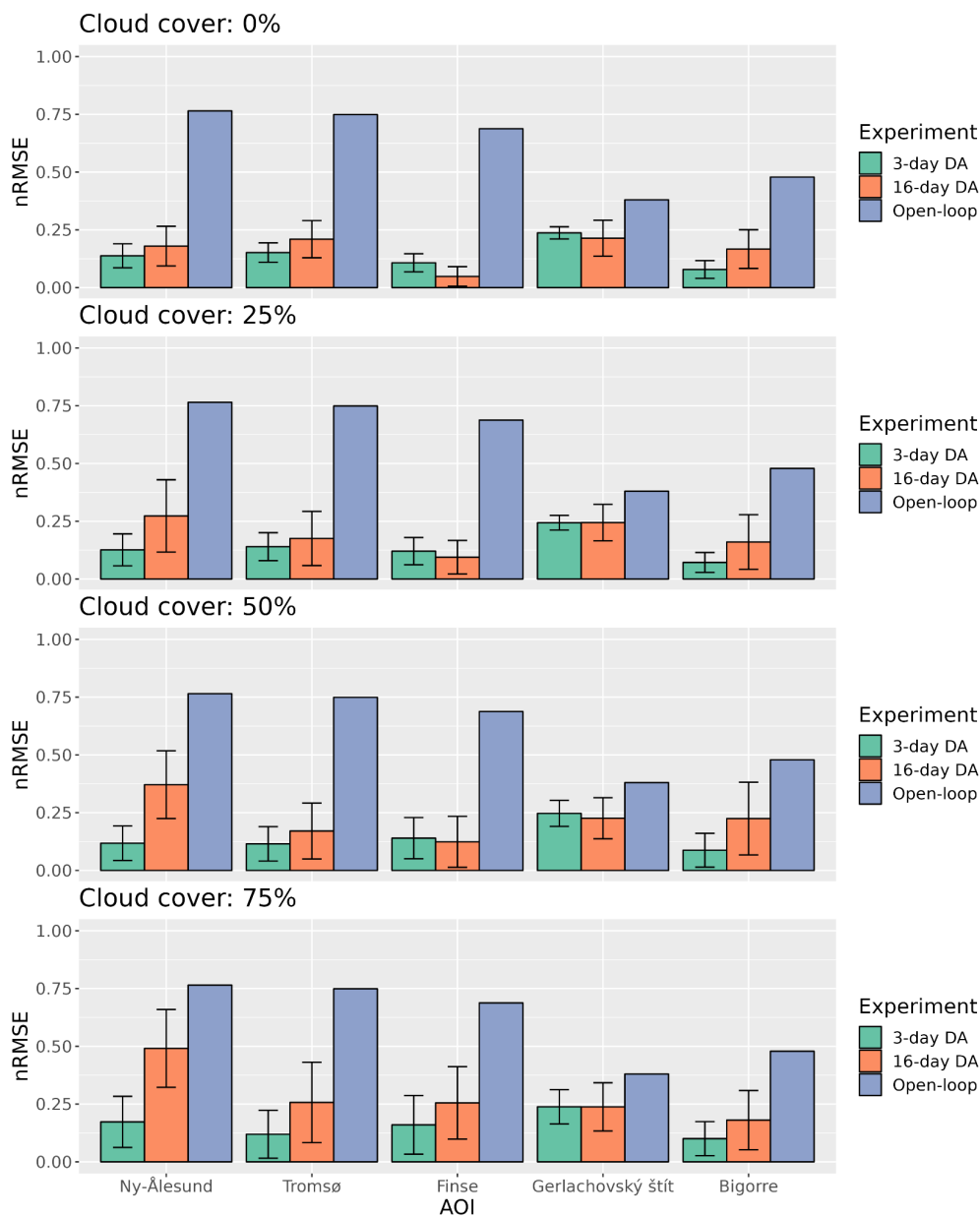



Figure 4: Normalized root mean square error (nRMSE) of the posterior SWE after assimilating LST using the PBS compared with the synthetic true SWE for each experiment. The bars indicate the mean nRMSE while the error bars indicate the standard deviation in the 100 replicates.

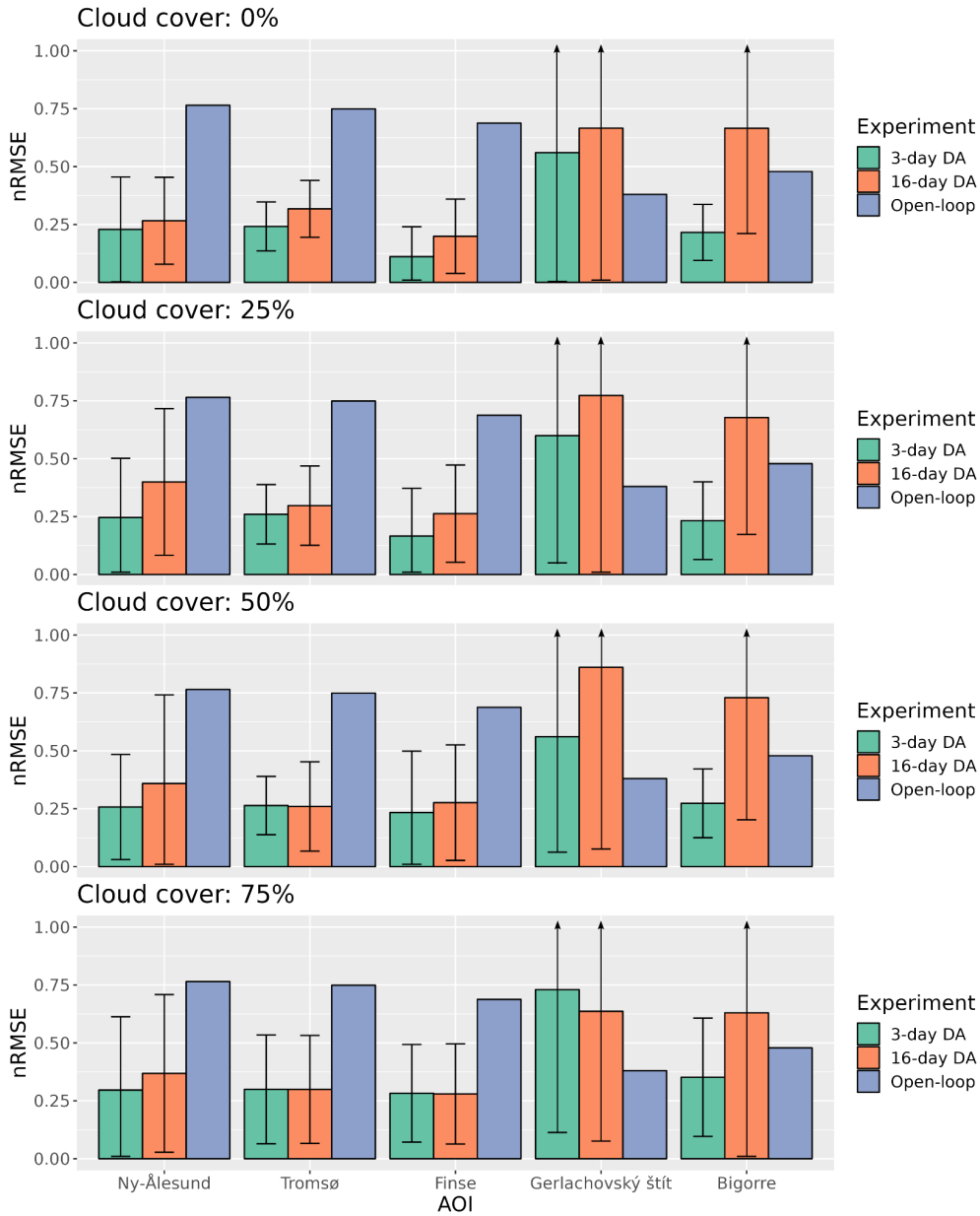



Figure 5: Normalized root mean square error (nRMSE) of the posterior SWE after assimilating FSCA using the PBS compared with the synthetic true SWE for each experiment. The bars indicate the mean nRMSE while the error bars indicate the standard deviation in the 100 replicates.

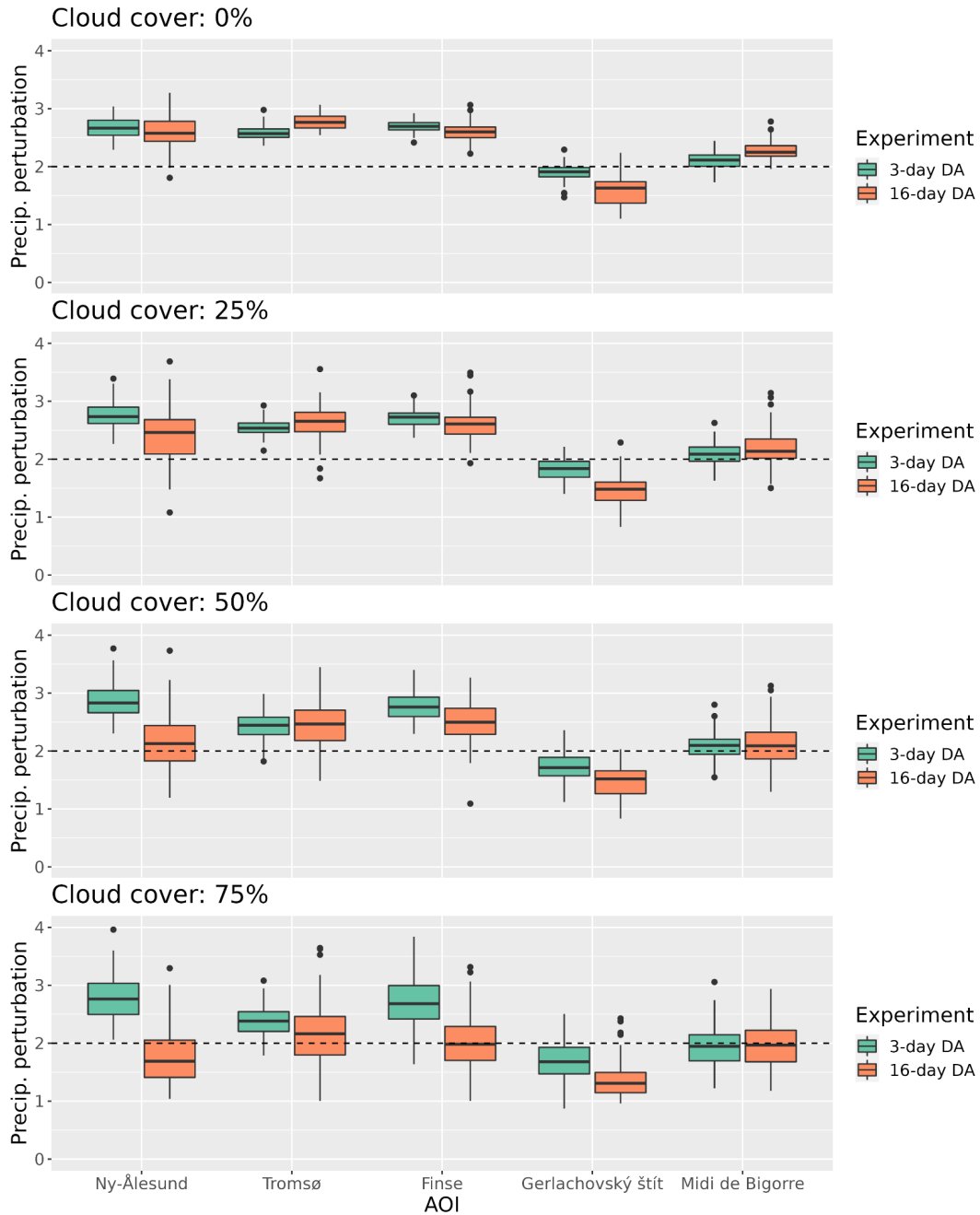



Figure 6: Boxplots showing the distribution of the posterior precipitation perturbation parameter for each experiment estimated by the PBS. The dashed line indicates the true perturbation that was applied to the forcing.

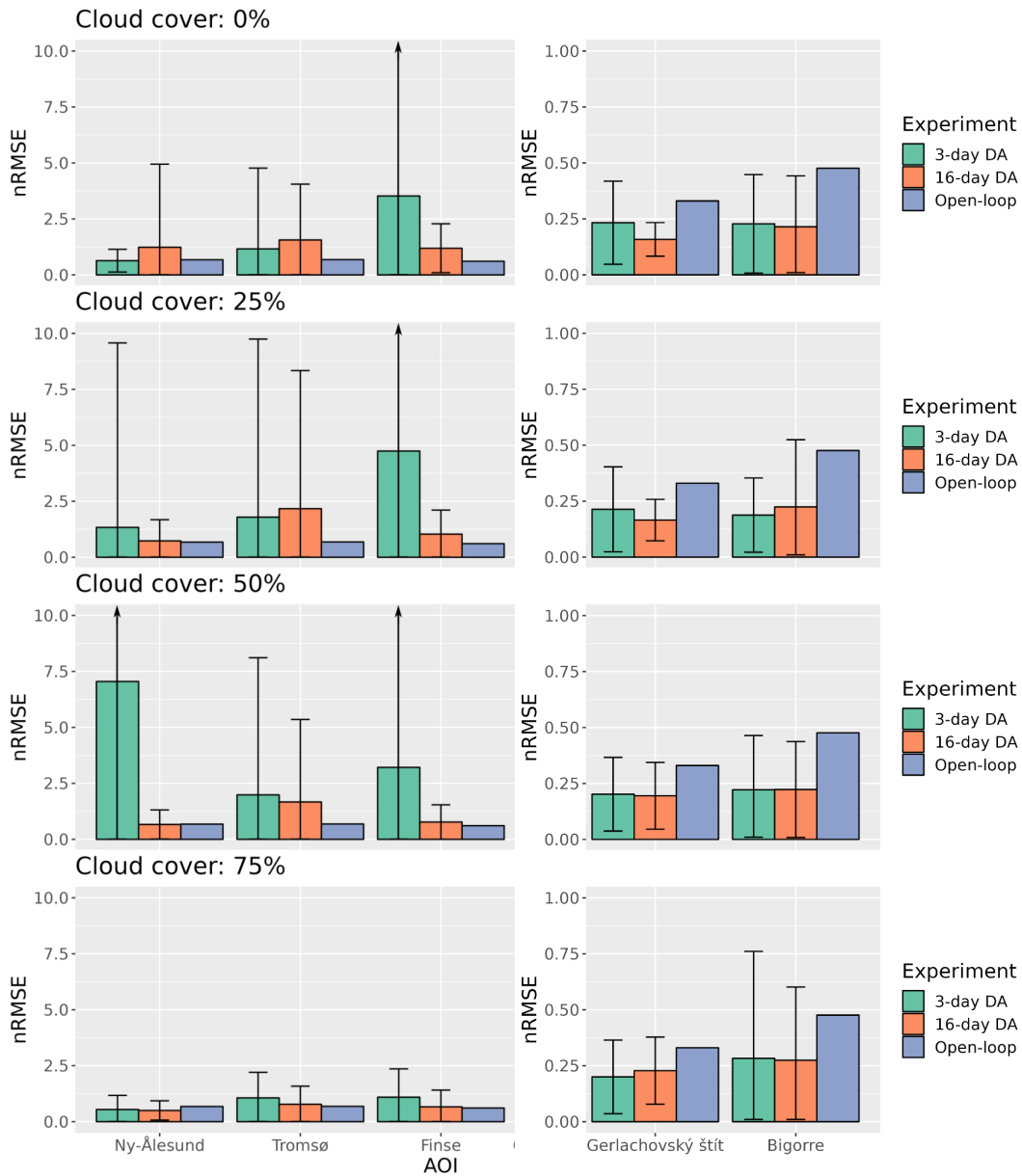
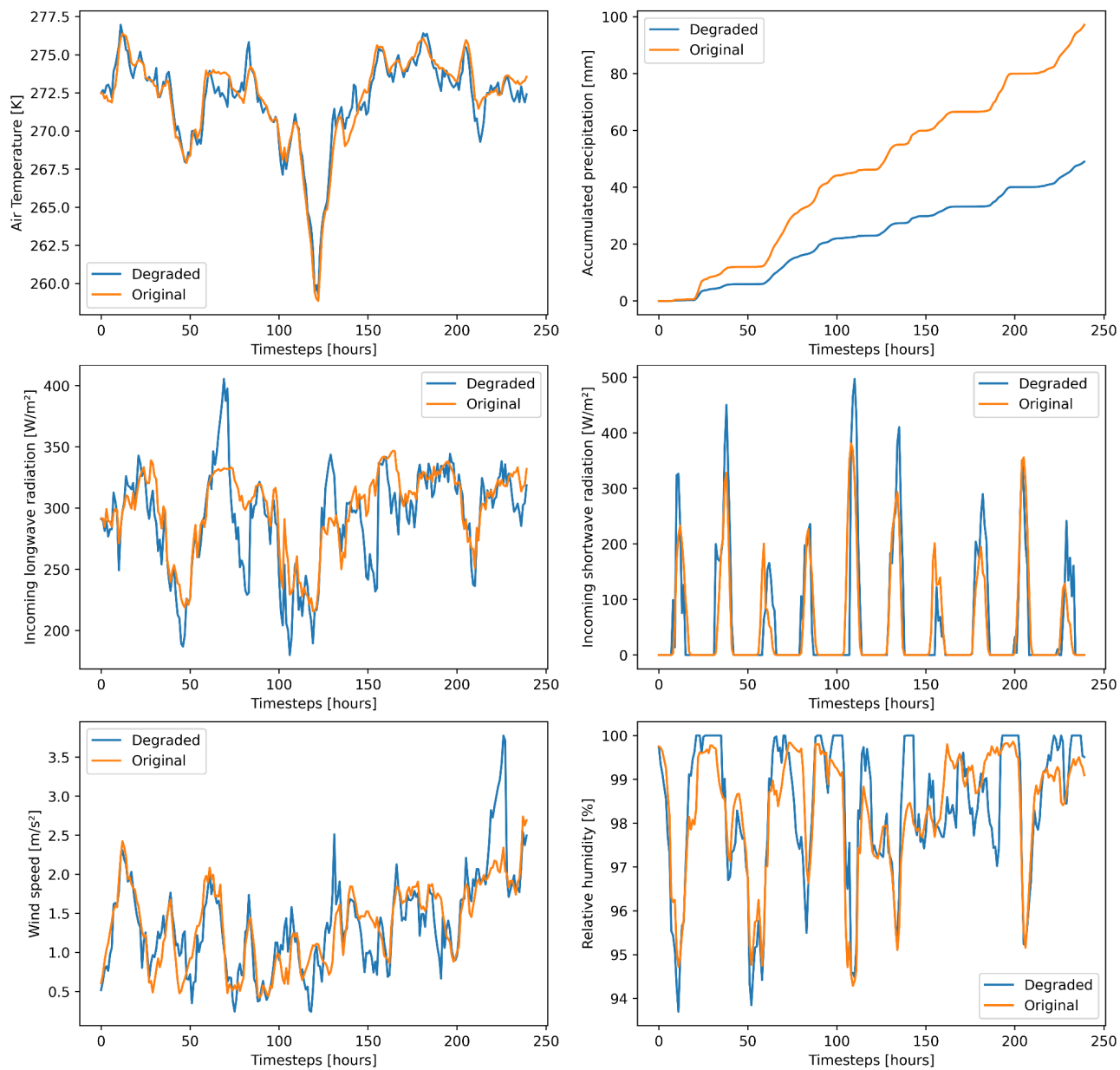
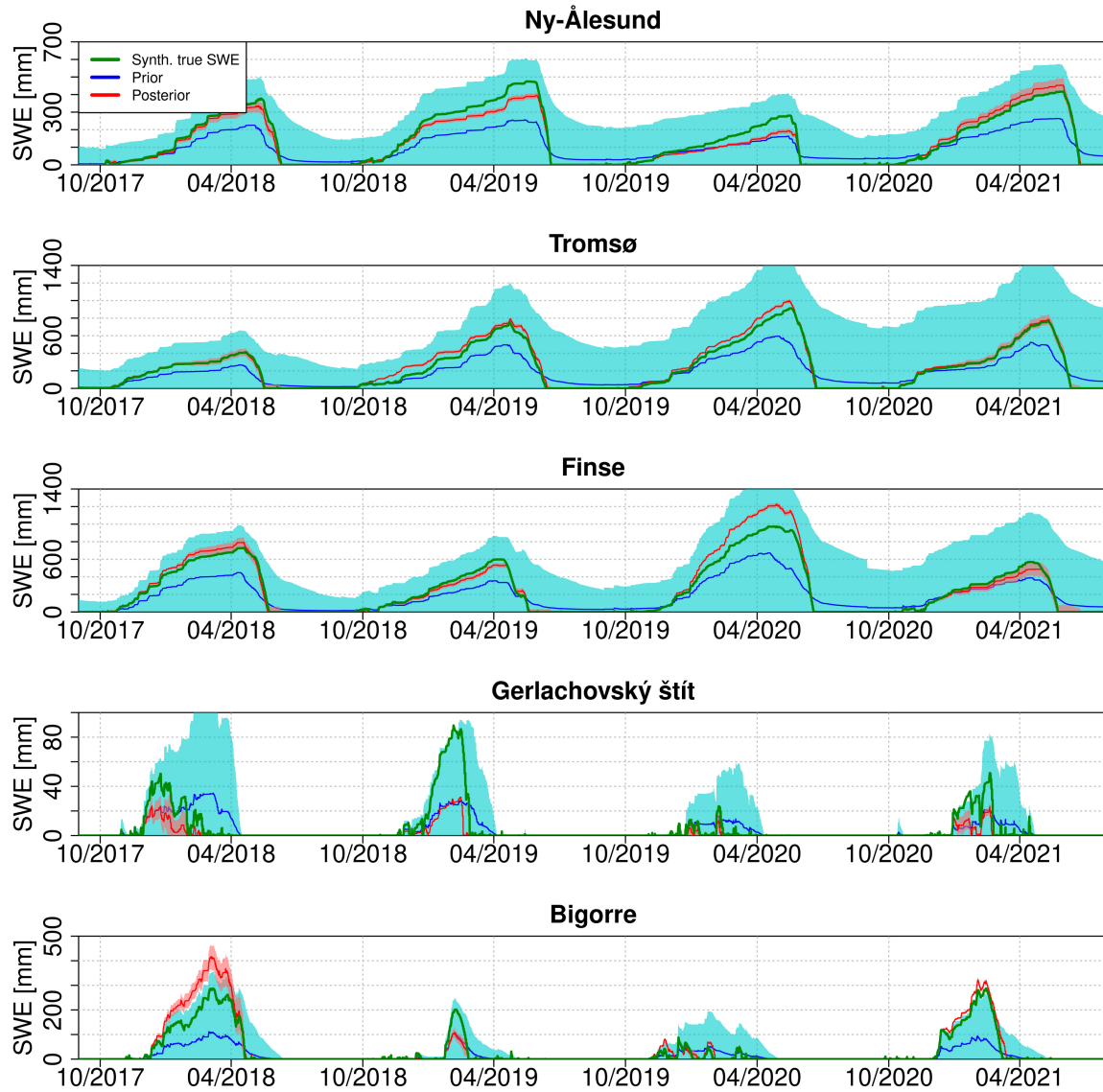
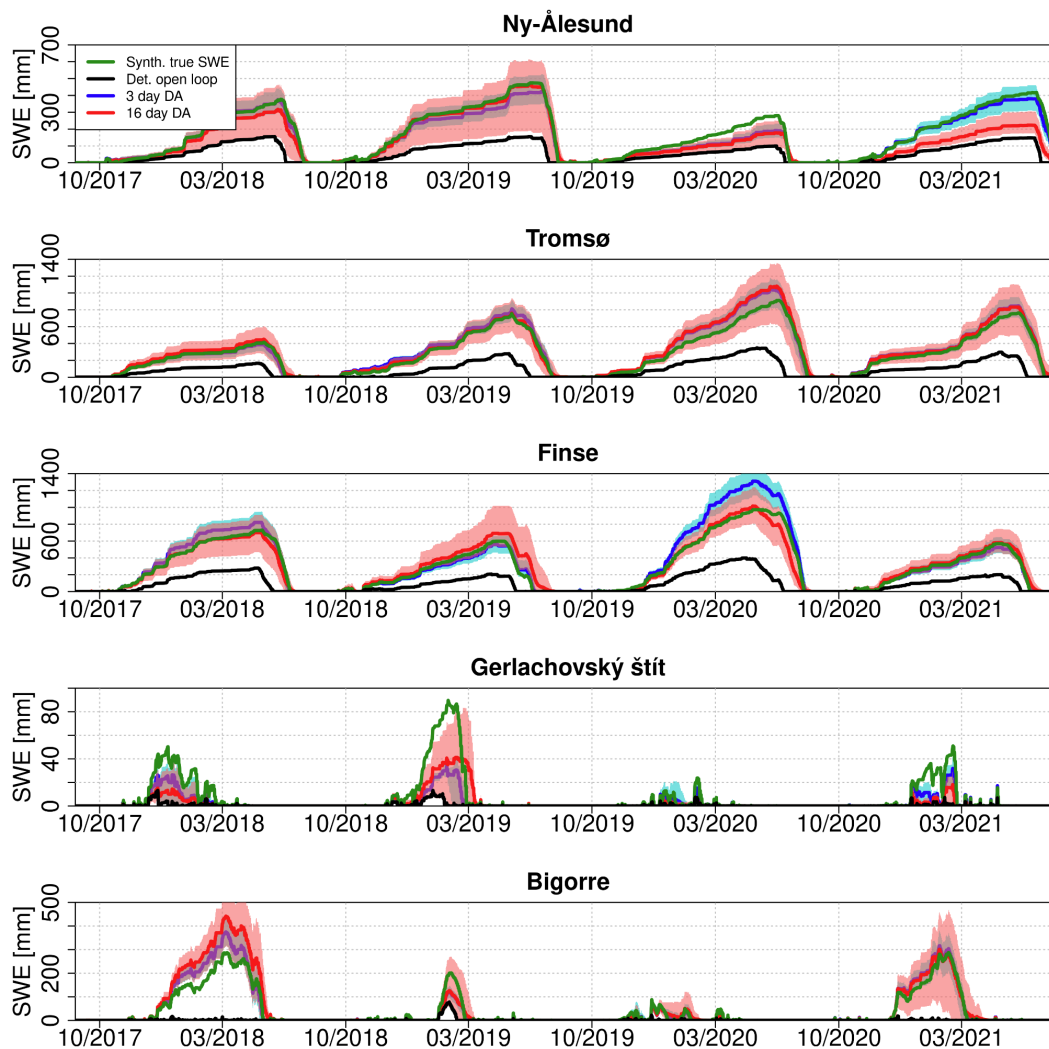
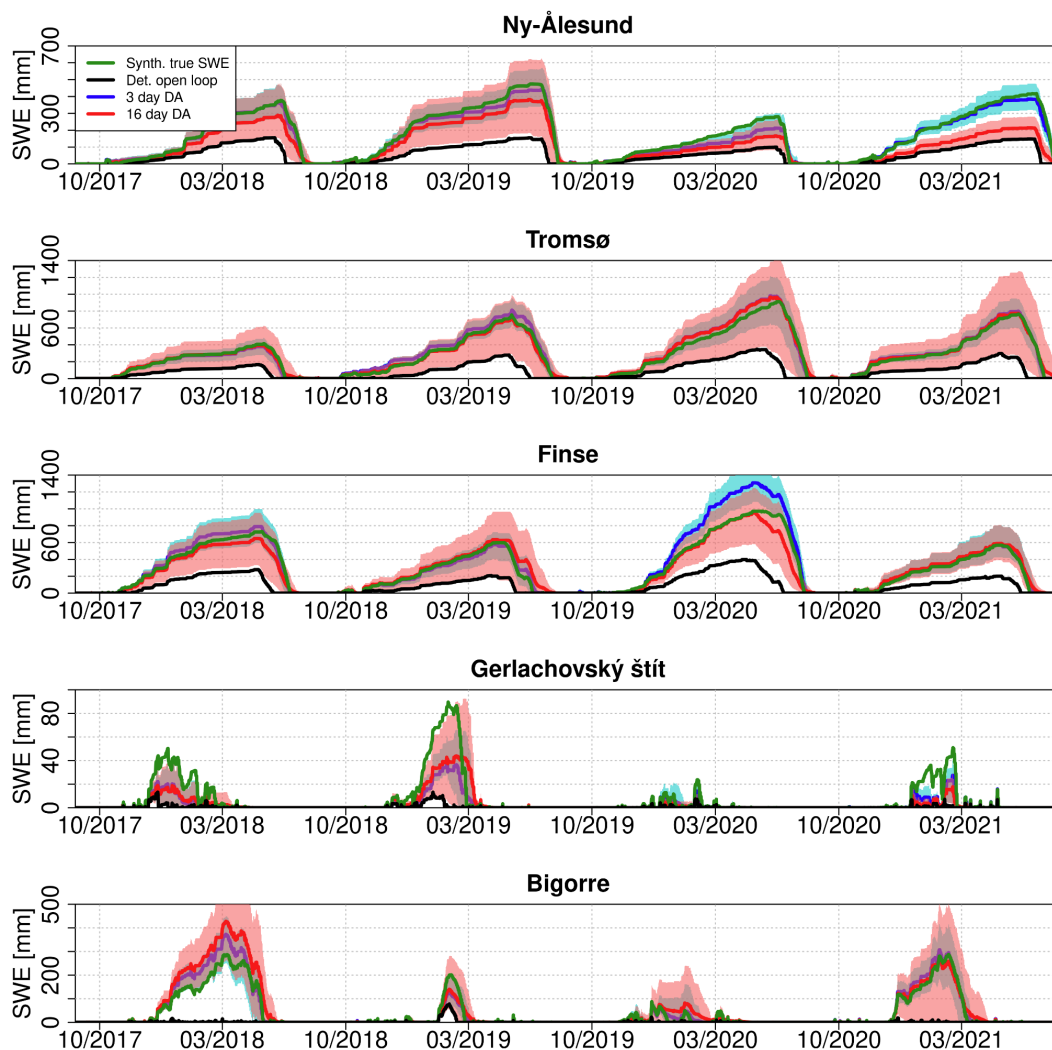
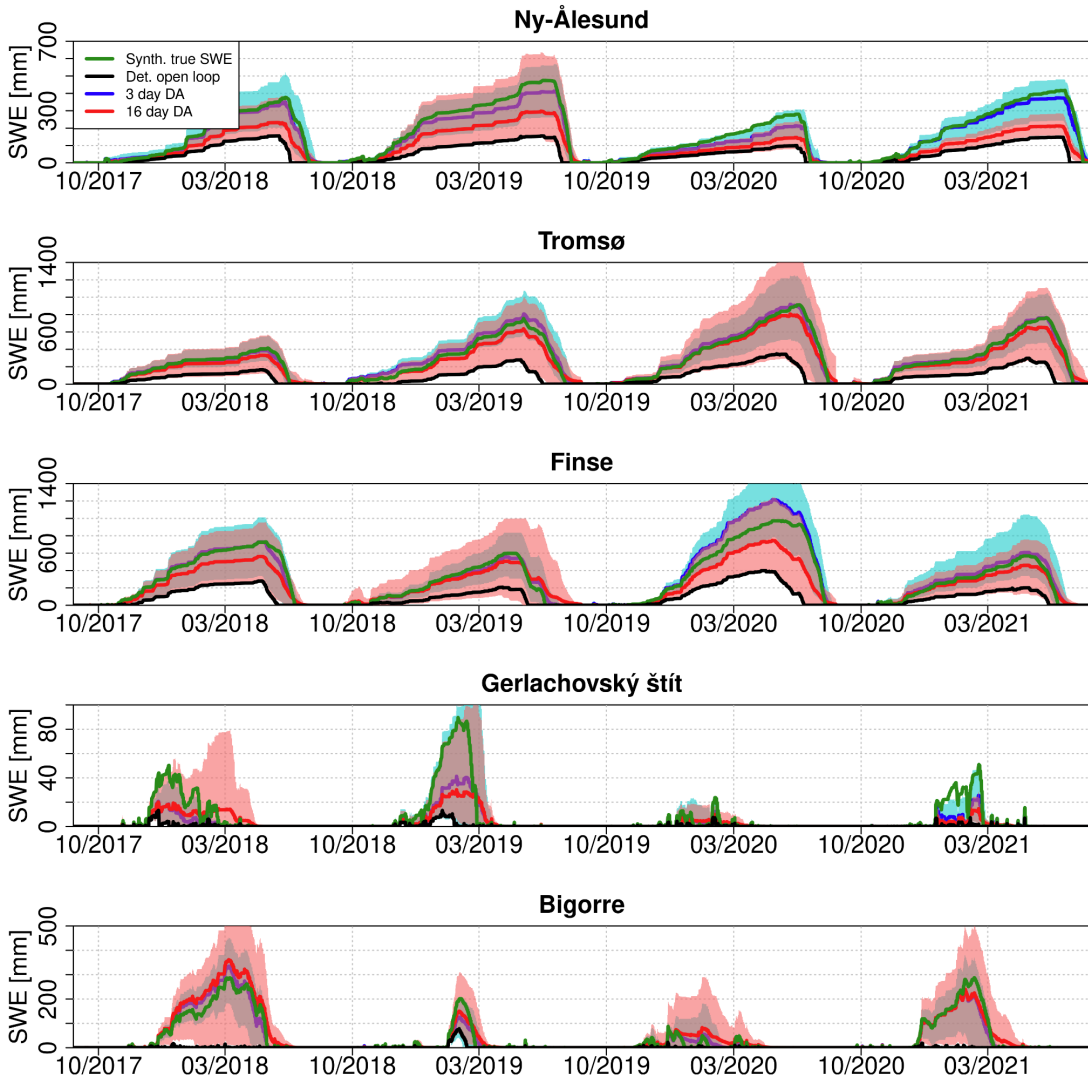



Figure 7: Normalized root mean square error (nRMSE) of the posterior SWE after assimilating LST from the PF compared with the synthetic true SWE for each experiment. The bars indicate the mean nRMSE while the error bars indicate the standard deviation in the 100 replicates.

575

90

Supplementary 1: Example of the forcing degradation covering 10 days at hourly timesteps

Supplementary 2: Example of prior and posterior SWE ensembles for a single replicate assimilating LST at a time resolution of 3 days in the absence of clouds. The shaded areas show the prior and posterior standard deviation.

Supplementary 3: Comparison of hourly time series of synthetic true SWE with the open loop simulations and the posterior SWE after assimilating LST with a revisit time of 16 days or 3 days (25% cloud cover scenario) using the PBS. Here, the posterior SWE is the average of the 100 replicates and the shaded areas represent the 95 to 5th quantile range

Supplementary 4: Comparison of hourly time series of synthetic true SWE with the open loop simulations and the posterior SWE after assimilating LST with a revisit time of 16 days or 3 days (50% cloud cover scenario) using the PBS. Here, the posterior SWE is the average of the 100 replicates and the shaded areas represent the 95 to 5th quantile range.

585

Supplementary 5: Comparison of hourly time series of synthetic true SWE with the open loop simulations and the posterior SWE after assimilating LST with a revisit time of 16 days or 3 days (75% cloud cover scenario) using the PBS. Here, the posterior SWE is the average of the 100 replicates and the shaded areas represent the 95 to 5th quantile range.