

Magnusson Jan (Orcid ID: 0000-0003-0257-1862)

Mazzotti Giulia (Orcid ID: 0000-0003-3857-7449)

Saloranta Tuomo Mikael (Orcid ID: 0000-0001-8436-0342)

Influence of spatial resolution on snow cover dynamics for a coastal and mountainous region at high latitudes (Norway)

Magnusson, Jan (1)

Eisner, Stephanie (2)

Huang, Shaochun (1)

Lussana, Cristian (3)

Mazzotti, Giulia (4), (5)

Essery, Richard (6)

Saloranta, Tuomo (1)

Beldring, Stein (1)

(1) Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway

(2) Norwegian Institute of Bioeconomy Research (NIBIO), Ås, Norway

(3) Norwegian Meteorological Institute (METNO), Oslo, Norway

(4) WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

(5) Laboratory of Hydraulics, Hydrology and Glaciology VAW, ETH Zurich, Zurich, Switzerland

(6) School of GeoSciences, University of Edinburgh, Edinburgh, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019WR024925

Abstract

Climate models show that global warming will disproportionately influence high-latitude regions and indicate drastic changes in, amongst others, seasonal snow cover. However, current continental and global simulations covering these regions are often run at coarse grid resolutions, potentially introducing large errors in computed fluxes and states. To quantify some of these errors, we have assessed the sensitivity of an energy-balance snow model to changes in grid resolution using a multi-parametrization framework for the spatial domain of mainland Norway. The framework has allowed us to systematically test how different parametrizations, describing a set of processes, influence the discrepancy, here termed the scale-error, between the coarser (5 to 50 km) and finest (1 km) resolution. The simulations were setup such that liquid and solid precipitation was identical between the different resolutions, and differences between the simulations arise mainly during the ablation period. The analysis presented in this study focuses on evaluating the scale-error for several variables relevant for hydrological and land surface modelling, such as SWE and turbulent heat exchanges. The analysis reveals that the choice of method for routing liquid water through the snowpack influences the scale-error most for SWE, followed by the type of parametrizations used for computing turbulent heat fluxes and albedo. For turbulent heat exchanges, the scale-error is mainly influenced by model assumptions related to atmospheric stability. Finally, regions with strong meteorological and topographic variability show larger scale-errors than more homogenous regions.

1 Introduction

The snow cover simulated by weather forecasting and climate models influences, for example, their predictions of air temperature since snow affects the surface energy balance [Van den Hoof et al., 2013]. For continental and global simulations, these models typically operate on horizontal scales ranging from approximately 10 to 100 km. When using such coarse resolutions, considerable heterogeneity is averaged out in land surface properties and meteorological conditions. Neglecting subgrid heterogeneities through averaging typically introduces errors in any model with nonlinear governing equations [Kirchner et al., 1993]. It is thus important to assess the likely magnitude of errors arising from neglecting heterogeneities in land surface parameters and meteorological forcings in land surface modelling. Indeed, for snow simulations, it has long been recognized that increasing the spatial resolution influences the model results, e.g. Essery [2003].

Using effective parameter values or introducing subgrid parametrizations in land surface models can only partly reduce scale-errors (i.e. the errors that arise from coarsening grid resolution and neglecting subgrid variability). Even when applying so-called effective parameters that should capture the influence of subgrid heterogeneity, land surface models tend to, for example, overestimate evapotranspiration and sensible heat fluxes compared to schemes taking such variability into account (see Rouholahnejad & Kirchner [2017]). To circumvent such issues when addressing critical water science questions, it has been argued that we need so-called hyper-resolution land surface models that resolve processes across very fine spatial scales, i.e. not coarser than 1 km, even at the global scale [Wood et al., 2011]. However, running models on very high resolution at the global scale is very computationally demanding. Likely, an even stronger limitation is the paucity of reliable, high-resolution input data in many regions, e.g. meteorological forcings that capture the actual spatial variability of the single variables. For such regions, the effective grid spacing of hyper-resolution models can be much coarser than that technically specified since the forcing data may not depict local features sufficiently well (e.g. convective precipitation, temperature inversions or variability in wind speed). Thus, even hyper-resolution models may be prone to the same scale-errors as models using a coarser grid.

For large-scale simulations (e.g. continental or global scale), several studies have examined how different grid resolutions affect snow simulations performed by various land surface models. Singh et al. [2015] found that high resolution (1 km) soil and topographic information induced changes in several simulated variables such as snow and surface energy fluxes compared to coarse scale simulations (100 km) using the National Center for Atmospheric Research (NCAR) Community Land Model. Dutra et al. [2011] compared snow simulations

performed at 25, 80 and 200 km grid resolutions using the land surface model HTESSEL (Tiled ECMWF Scheme for Surface Exchange over Land) applied in the European Centre for Medium-Range Weather Forecasts (ECMWF) system. The authors found that the horizontal resolution played an important role for the snow cover simulations over complex terrain (e.g. coastal and mountain regions). In the Rhone-AGGregation intercomparison project of land surface schemes, snow water equivalent was reduced greatly when upscaling the horizontal resolution for all models except the Variable Infiltration Capacity (VIC) model that included an elevation dependent tiling of the snow scheme and therefore performed well even at coarse grid resolutions [Boone et al., 2004]. For the same catchments, similar conclusions were drawn by Stöckli et al. [2007]. Even though VIC uses elevation bands for snow, Haddeland et al. [2002] found a slight underestimation of snow water equivalent when increasing the spatial resolution from 1/8 to 2° for the Columbia River basin. To summarize, most of the previous large-scale studies show that coarser grid resolutions reduce snow amounts. However, the applied models show different responses to the coarsening of the grid resolution likely due to, for example, varying process parametrizations and numerical implementations. Therefore, previous studies give limited insight into why the models respond differently to changes in horizontal grid resolution.

In this study, we assess the sensitivity of snow simulations for mainland Norway to changes in spatial resolutions from fine (1 km) to coarse scale (5 to 50 km) using a multi-parametrization framework (i.e. a model that allows the user to choose from alternative parametrization for different steps in the modelling chain, e.g. computation of surface albedo and turbulent heat fluxes). The multi-parametrization setup allows us to test how various parametrizations of different processes influence the sensitivity of the simulations to changes in grid resolution. Our main aims are (a) to assess which process parametrizations influence the scale-errors most, (b) to quantify the scale-error between the coarse and fine scale simulations for different model configurations, and (c) to assess in which physiographic and climatic settings the scale-errors are largest. We focus our analysis on differences in simulated snow water equivalent (SWE), latent heat fluxes (LATMO), sensible heat fluxes (HATMO) and net radiation (RNET) between the fine and coarse scales. We have focused on these variables since SWE is important for water-related questions, whereas the other variables are more relevant for weather forecasting and climate model simulations.

2 Data

2.1 Site description

The study area covers mainland Norway and small regions of Sweden and Finland (Figure 1). These additional areas are situated along the eastern Norwegian border, with runoff draining to Norway. The study region is completely located on the Scandinavian peninsula, and covers an area of approximately 345000 km². Latitudes range from approximately 58 to 71° North and the highest peak rises 2469 m above sea level. Large parts of the study region are characterized by highly variable topography and a rugged coastline in the south, west and north (Figure 1a). When coarsening the digital elevation model from 1 to 10 km resolution, many valleys disappear, but the main features are still visible (Figure 1b). However, when coarsening the digital elevation map further to 50 km resolution, only the very large-scale features are still apparent (Figure 1c).

2.2 Meteorological data

As mentioned above, one of the main difficulties for high-resolution simulations is to provide the model with spatially distributed parameter sets and forcing data. For this study, we used daily gridded data of air temperature, precipitation and wind speed with 1 km horizontal resolution for the period from 2008-9-1 to 2011-9-1. The daily temperature and precipitation gridded datasets were derived with statistical interpolation methods based on hundreds of observations stored in the climatological database of the Norwegian Meteorological Institute. The observational data were collected by meteorological stations managed by public institutions and an automatic data quality control procedure were applied to all data. The minimum and maximum two-meter air temperature grids were generated using the data and interpolation procedure outlined by Lussana [2017]. The daily precipitation grids were produced using the methods presented in Lussana et al. [2018]. The observed values were corrected for undercatch using the wind speed grids along with the methods reported in Wolff et al. [2015], adapted to daily accumulated values as described in Lussana et al. [2019]. The 10-meter daily averaged wind speed grids are based on numerical model output because the spatial distribution of stations measuring wind is too sparse for statistical interpolation to be applied. The grids were produced on the 1 km grid by the quantile mapping technique presented by Gudmundsson et al. [2012] that adjusts the historical archive of the 10 km Norwegian hindcast dataset (NORA10, [Reistad et al., 2011]) to better match the climatology of the high-resolution operational numerical weather prediction model

AROME [Müller et al., 2017]. Note that the AROME wind speeds have been downscaled from the original 2.5 km grid resolution to 1 km by means of a nearest neighbor approach, where the elevation differences are considered in addition to the horizontal distance.

Mean annual precipitation ranges from approximately 300 mm in the inland regions in the southeast and north to nearly 4000 mm along the southwestern coastline (Figure 2a). Notice the strong precipitation gradient in the south when moving eastwards from the coast towards inland regions. Average wind speeds are highest in the northernmost areas and on the mountain ridges in the south (Figure 2b). Mean annual air temperature is highest along the southern coastline and all the way towards mid Norway, where Foehn effects often influence temperatures (Figure 2c). The high mountains ranges in southern and northern Norway experience the coldest temperatures. Our gridded observational fields of meteorological conditions show large spatial variations, likely higher than in many other regions of the world with flatter topography and situated further away from the sea.

To better illustrate the spatial variability in the meteorological fields, Figure 3a to c shows the subgrid standard deviations in the forcings for the 10 km resolution computed using the time-averaged 1 km gridded data presented above (Figure 2). Figure 3d to e shows the corresponding results for the 50 km resolution. For precipitation, the highest subgrid variability emerges along the coastline in the south and the associated transition zone from high to low precipitation when moving inland (Figure 3a and d). Subgrid variability in precipitation is much higher for the 50 than 10 km resolution. The highest variability in wind speed occurs in the southern and mid-latitude regions of the study domain (Figure 3b and e). In contrast to precipitation, already the 10 km resolution shows rather high subgrid variability compared to the 50 km results. Finally, for air temperature, the highest standard deviations are visible in the southwestern regions (Figure 3c and f). Similar to wind speed, the subgrid variability in air temperature can be rather high for the 10 km resolution, comparable with the 50 km results. In summary, it is evident that much information is already lost when aggregating the forcings from 1 to 10 km resolution, and even more smoothing occurs when moving to the 50 km results. Noteworthy, the areas with highest variability do not necessarily coincide with the highest topographic complexity (compare with Figure 1a), but are also influenced by, for example, gradients from coastal towards inland regions.

2.3 Forcing data generation

In many regions, generating forcing data sets for high resolution (1 km and smaller) land surface modeling is very challenging due to limited data availability on such fine scales [Singh et al., 2015]. In our case, several input variables (e.g. radiation components) required

by an energy-balance snow model are missing in the available input dataset (see section 2.2). Since there is no high-resolution re-analysis dataset available for our study region containing all required forcing variables, we used a version of the so-called mountain microclimate simulation model (MT-CLIM) presented by Hungerford et al. [1989] implemented in the VIC model (see Bohn et al. [2013] for details) to create the necessary meteorological input fields on subdaily resolution. The model was driven by 1 km gridded daily data sets of minimum air temperature, maximum air temperature, wind speed and precipitation (see section 2.2). Based on these variables, the VIC implementation of MT-CLIM uses a set of algorithms to estimate mean air temperature, rainfall, snowfall, relative humidity, longwave radiation, shortwave radiation and wind speed at subdaily resolution (in our case 3h). Clear-sky longwave radiation was computed using the Tennessee Valley Authority [TVA, 1972] algorithms, and adjusted for clouds using the methods presented by Deardorff [1978]. Overall, MT-CLIM has been shown to reproduce global patterns of incoming shortwave and longwave radiation as well as humidity reasonably well, but for stations situated close to coasts, the algorithms occasionally show poor performance [Bohn et al., 2013].

2.4 Land-use data

The land-use distribution in the study domain was derived from the high-resolution (30 m), remote sensing-based forest resources map SAT-SKOG [Gjertsen & Nilsen, 2012], which was classified into 12 structural forest types using the scheme developed by Majasalmi et al. [2018] for Fennoscandian forests. This classification scheme differentiates between three species groups (spruce, pine, and deciduous dominated), and each group is further divided into four structural subgroups reflecting differences in, amongst others, stand height and leaf area index. For each forest type, a look-up table provided by Majasalmi et al. [2018] gave values for leaf area index and vegetation height. These look-up tables provide maximum growing-season leaf area index only. Since we focused on snow-season processes, the leaf area in deciduous forest was assumed to equal zero. Unforested areas were not further classified into subtypes since the multi-parametrization framework we have used for simulating snow processes does not distinguish between different non-forested land-use types (e.g. urban and glacier areas).

2.5 Snow observations

For evaluating model performance, we used a comprehensive SWE data set, consisting of snow measurements made by various hydropower companies since 1914. The SWE in this

dataset is computed from manual measurements of snow depth and density. In most cases, the reported snow depth is an average of multiple observations along a snow course, while the density observation is normally measured at one snow pit along the snow course. The measurements were made once each winter, and mostly recorded around the time of maximum annual SWE just before spring snowmelt begins. The data were quality-checked and values considered as outliers were corrected or omitted. In total, 2636 SWE observations were available for model evaluation within the simulation period stretching from 2008-9-1 to 2011-9-1. Most of the observations originate from mountain areas in southern, eastern and western Norway. This snow data set has previously been used among others in evaluation of the operational snow map model in Norway [Saloranta, 2012], as well as in calculation of long-term decadal trends in snow depth and SWE [Skaugen et al. 2012, Dyrrdal et al. 2013].

3 Methods

3.1 Model description

The snow simulations have been carried out using an updated version of the Flexible Snow Model (FSM) developed by Essery [2015], which additionally includes a one-layer canopy model [Essery et al., 2003] utilizing the land-use information outlined in section 2.4. FSM is a multi-parametrization framework of energy-balance snow models with intermediate complexity. The framework gives the user the ability to choose from two alternative parametrizations for five snowpack processes, namely the representation for simulating the surface albedo of snow, thermal conductivity of snow, density of snow, turbulent heat exchange at the snow surface, and routing of liquid water through the snowpack (Table 1). The parametrizations were all taken from established land surface models such as CLASS [Bartlett et al., 2006], CLM [Oleson et al., 2010] and ISBA [Noilhan & Mahfouf, 1996]. For each of the snowpack processes listed above, we have applied two different parametrizations denoted as Option 0 and Option 1 in this study.

This results in 32 possible combinations of parametrizations, yet not all possible combinations are necessarily unique. For example, the thermal conductivity of snow will be constant even if we use Option 1 for this process if the snow density is set constant (Option 0 for Snow compaction). In the setup of the multi-parametrization framework used in this study, the conductivity given by Yen [1981] for the fixed density was set such that it matches the fixed conductivity. As a result, there are only 24 distinct configurations. However, for

simplicity, we present results from all 32 available combinations. The model was run using 3 h time steps. FSM solves the mass and energy exchanges for up to three individual snow layers. The surface heat balance equation is solved iteratively and the vertical temperature profile in the snowpack is solved by the Crank-Nicolson method. FSM also simulates the snow covered fraction as a simple function of total snow depth. When applied over forested areas, the updated model version solves coupled energy balances between the canopy and the snowpack. This includes the effect of the canopy on energy and mass balance processes such as canopy radiative transfer and snow interception. For more details about the framework and available parametrizations, see Essery et al. [2003] and Essery [2015].

3.2 Description of simulations

All 32 configurations were run at 1, 5, 10, 25 and 50 km resolution. For the finest resolution, the model was driven using the original meteorological forcing data as described in section 2.2 and 2.3. For the coarser scale simulations, the 1 km forcings were upscaled using the arithmetic mean, to ensure mass-conserving scaling. Note that in the upscaling only land points were considered in the upscaling, and that areas covered by, for example, oceans or located outside of the fine scale model domain were ignored. For land-use, each 1 km grid cell can contain up to 12 different forest types (see section 2.4) in addition to one land-use class representing open areas. Due to computational constraints for the 1 km simulations, we only considered the three land-use classes with largest areal coverage. Consequently, each grid cell was divided into three tiles describing the separate land-use types. For the coarse scale simulations (5 to 50 km), we first computed the average areal coverage of each land-use class within each coarse scale grid cell from the 1 km grids. Subsequently, the three classes with the largest coverage were proportionally scaled to total cell area. Note that precipitation phase was determined on 1 km resolution, and rainfall and snowfall amounts were separately aggregated to the coarser grid resolutions. We chose this approach to minimize differences between the simulations during the accumulation phase, in order to being able to better attribute differences between the different resolutions during the ablation phase.

3.3 Evaluation statistics for characterizing scale-errors

For the comparison of the simulation results given by the different resolutions, the results from the fine scale simulation (1 km) were aggregated to the coarser resolutions (5 to 50 km) by arithmetic averaging. Thus, in all analysis presented below the finest scale simulations (1 km) were always upscaled to the coarser resolutions. We then computed the discrepancy,

which we term the scale-error, between the results for the coarse scale y_n and upscaled fine resolution \hat{y}_n :

$$\varepsilon_n^i = y_n^i - \hat{y}_n^i \quad (\text{Eq. 1})$$

for grid cell i and time n of N available steps. For the i -th grid cell, the root-mean-square-error is given by:

$$RMSE^i = \sqrt{\frac{\sum_{n=1}^N \varepsilon_n^i{}^2}{N}} \quad (\text{Eq. 2})$$

and the bias and mean-absolute-bias by:

$$Bias^i = \frac{\sum_{n=1}^N \varepsilon_n^i}{N} \quad (\text{Eq. 3})$$

$$MAB^i = \frac{\sum_{n=1}^N |\varepsilon_n^i|}{N} \quad (\text{Eq. 4})$$

To compute the average scale-error for the whole simulation domain, we used a weighted average of $RMSE^i$ and MAB^i . For this calculation, the weights were given by:

$$\omega^i = \frac{NG^i}{L^2} \quad (\text{Eq. 5})$$

where NG^i denotes the number of available 1 km grid cells within one coarse resolution grid cell, and L represents the grid resolution. We used a weighted average since not all coarse resolution grid cells have complete coverage of fine resolution land-use and meteorological data. Since the length of the snow season varies between regions, we computed the scale-error for the whole study period to facilitate the comparison of this error between regions. This choice likely reduces the scale-error, in particular in regions with long snow-free periods.

4 Results

4.1 Evaluation against observations

Figure 4 shows the model performance for SWE. For this evaluation, the model was run on the 1 km grid excluding vegetation effects since the measurements were taken in open areas. These simulations were then evaluated using the snow observations presented in section 2.5 using the bias and normalized root-mean-square-error (NRMSE) as performance metrics. The normalization was performed using the average of all SWE observations. We used standard procedures to compute these error statistics, and they should not be confused with those used for characterizing the scale-errors as outlined in section 3.3. First, most model configurations underestimate SWE compared to the observations. The high NRMSE shows that the discrepancy can be large between the 1 km resolution simulations and the point observations. The latter can be influenced by local effects such as wind drift giving rise to large errors when comparing point scale observations with gridded model results. Second, we find that the model performance is sensitive to the choice of method for computing turbulent heat fluxes (Exchng) and routing of liquid water through the snowpack (Hydraul). For both of these processes, the more physically-based parametrizations (Option 1) give both lower bias and lower NRMSE than their simpler counterparts (Option 0). Finally, the method for computing albedo does not seem to affect the model performance notably. The remaining processes (snow compaction and thermal conductivity, see section 3.1) follow the behavior of the albedo and are not shown here.

4.2 Influence of parametrization on scale-errors

Figure 5 ranks the different model configurations from lowest to highest scale-error in terms of RMSE (see section 3.3 for details about the computations of these error statistics). When considering SWE, we first find that configurations not taking liquid water in the snowpack (Option 0) into account mostly show lower scale-errors than those using a bucket formulation (Option 1). Second, given the parametrization for snow hydraulics, the method for computing turbulent heat fluxes also influences the scale-error systematically. RMSE is typically smaller when applying a stability correction scheme (Option 1) than when assuming neutral atmospheric stability (Option 0). Third, given the parametrization for snow hydraulics and turbulent fluxes, the diagnostic parametrization (Option 0) for computing surface albedo often shows a smaller scale-error than the more physically based alternative (Option 1). Summarizing the results for SWE, the different process parametrizations seem to influence

the scale sensitivity in a cascading order where some processes are more influential than others.

For net radiation (RNET), the more complex parametrizations (Option 1) for turbulent heat fluxes, snow hydraulics and surface albedo tend to produce smaller scale-errors than the simpler parametrizations (Option 0). Apart from this tendency, no clear pattern between the processes and parametrizations emerges such as found for SWE. For sensible (HATMO) and latent heat fluxes (LATMO), the scale-error is systematically smaller when applying a stability correction scheme (Option 1) than when assuming neutral atmospheric stability (Option 0). The bucket approach (Option 1) to simulate liquid water content of the snowpack usually produces smaller scale-errors in LATMO than the simpler parametrization, which contradicts the pattern found for SWE. In summary, the results show that (a) for some variables, e.g. SWE, rather clear patterns emerge on how the different parametrizations influence scale-errors whereas for others, e.g. LATMO, the patterns are less clear, (b) the same parametrization, e.g. Option 1 for turbulent heat exchanges, can have the opposite effect on the ranking for different variables, e.g. compare SWE and LATMO, and (c) some of the parametrizations, e.g. snow compaction, are less important for the ranking presented here and consequently do not introduce large scale-errors when coarsening the grid resolution.

4.3 Model sensitivity to variations in spatial resolution

From the results above, it is evident that, depending on the variable of interest, different parametrizations dominate the scale-error between the 1 km and the 50 km resolutions. Figure 6 to 9 shows how the scale-error in terms of domain-averaged root-mean-square-error and mean-absolute-bias increases from the 1 to 50 km resolution (both equaling 0 at 1 km resolution). MAB was chosen to avoid cancellation effects during averaging.

For all variables, we find that the RMSE increases the most from 1 to 5 km resolution. Differentiated by turbulent heat flux parametrization, we find that the two parametrizations systematically influence the scale-error across all resolutions. Thus, when comparing the results to those presented in

Figure 5 for the 50 km resolution, the ranking of the configurations from lowest to highest scale-error apparently remains similar across all grid sizes. For SWE, MAB and RMSE show similar behavior, but Option 0 for turbulent heat fluxes typically produces larger biases than Option 1 (Figure 6). MAB for RNET is very small, and increases more strongly between 1 and 5 km resolution if Option 1 for turbulent heat exchanges is selected instead of Option 0,

but then remains almost stable from 5 to 50 km grid size (Figure 7). For both turbulent heat fluxes (HATMO, LATMO), we find a rather linear increase between different resolutions, although the largest increase still occurs from 1 to 5 km resolution (Figure 8 and 9). To summarize the results presented in these figures, we note that the scale-error increases most at the fine resolutions and that the parametrization of turbulent heat fluxes influences the results systematically across the different scales.

4.4 Spatial patterns in scale-errors

The bias between the coarse (50 km) and upscaled fine (1 km) resolution simulations differs throughout our study region depending on both the variable of interest and the choice of model configuration, and is in general small for the different variables (Figure 10 and 11). In the following, we assess how the choice of turbulent heat flux parametrizations influences the results since those two options influence all variables, in contrast to, for example, the available parametrizations for snow hydraulic processes that mainly affects SWE (Figure 5). In many regions, the bias in SWE is close to zero, in particular in the southeastern and northern parts of our study domain for both options of turbulent heat fluxes (compare panel A in both figures). However, along the coastline, the bias is predominantly negative, with larger discrepancies between the two resolutions when assuming a neutral atmosphere (Option 0) instead of applying a stability correction scheme (Option 1). In the remaining regions of the study area, the bias is slightly positive, with somewhat higher values for the Option 1 than 0. For the turbulent heat fluxes, Option 1 shows lower biases than Option 0 (compare panels C and D in both figures, respectively). For the latter, the biases are almost exclusively negative for the sensible and positive for the latent fluxes. Only in a few areas in the southern and northern parts of the study domain, the maps show almost unbiased results. In most areas, the bias for sensible and latent heat fluxes show opposite signs. Finally, for net radiation, the biases are overall slightly positive for Option 1, whereas more centered around zero for Option 0 (compare panels B in both figures). For all variables, we find that the coarse scale runs both over- or underestimate the fine resolution simulations although with regional differences. Overall, the neutral stability assumption shows larger biases than the simulations applying a stability correction scheme.

Figure 12 shows the RMSE between the coarse (50 km) and upscaled fine (1 km) resolution simulations. The largest scale-errors for SWE occur in the regions along the western coastline, and the errors are much smaller in the inland regions (see panel A). For the three surface energy fluxes, the largest scale-errors occur in the southern mountain regions,

followed by the regions at mid latitudes (see panels B to D). For the northern and southeastern most regions, the errors are much smaller.

4.5 Topographic influence on scale-errors

For all variables, we find that the scale-error correlates well with the subgrid topographic variability for all configurations (Figure 13). These correlations were computed using (a) the RMSE between the 50 and upscaled 1 km simulation results, and (b) the standard deviation of the 1 km digital elevation model within each of the 50 km resolution grid cells. For SWE, the topographic variability explained in median 71 % of the variance in the scale-error, and the corresponding values for latent heat fluxes was 81 % and for sensible heat fluxes 75 %. For net radiation, subgrid topographic variability explained approximately 80 % of the variance in the scale-error in median. For all the variables, the spread in squared correlation coefficient between the model configurations is rather low. To summarize, regions with large topographic variability are prone to larger discrepancies between coarse and fine scale simulations than flatter regions irrespective of model configuration.

5 Discussion

For our study region and evaluation data, FSM performed best for SWE when using the configurations applying Option 1 for turbulent heat exchanges and snow hydraulic processes (Figure 5). However, at the coarser grid resolutions (5 to 50 km) model configurations relying on those parametrizations may not necessarily be the best choice anymore since the upscaling introduces errors, and the scale-errors are larger for some configurations than others (Figure 4). For example, the SWE simulations are more sensitive to changes in model resolution when using the bucket formulation (Option 1) instead of the simpler approach (Option 0) for routing liquid water through the snowpack. Thus, when changing model resolution the question arises whether one should choose the configuration showing the best performance, or the one with the lowest scale sensitivity. Preferably, model performance should be evaluated at the (coarser) simulation scale to avoid such dilemmas. However, with point observations as used here, such an evaluation will introduce even larger representativeness errors than already present between the SWE observations and the 1 km grid cells. In fact, practically all reliable snow observations are made at much smaller scales than those continental and global land surface models are run at (10 to 100 km).

The subgrid variability of the forcings increases with resolution, and is influenced by topographic effects and weather gradients from ocean towards inland regions (Figure 3). The MT-CLIM algorithm we used for generating the high-resolution forcing fields likely underestimates the true variability in the forcings, in particular for derived variables such as incoming shortwave radiation and relative humidity. The lower than true variability in some input variables likely leads to an underestimation of the scale-error for simulated variables such as snow water equivalent, but should not affect the sign of potential biases. Nevertheless, acquiring high resolution forcing fields that fully replicate the true spatial variability is likely unfeasible, in particular for large domains with very complicated meteorological conditions such as our study region. However, recent numerical weather prediction systems operate on higher resolution (e.g. 2.5 km), and their outputs may be useful for enhancing the understanding of scale related sensitivities in snow models since they produce physically consistent sets of meteorological forcings.

For most variables, we find that the largest increase in scale-error already occurs from the 1 to 5 km resolutions and afterwards flattens out for the coarser grid sizes (Figure 6 to 9). We also observe that the scale-error is larger for some parametrizations than others, e.g. when computing turbulent heat fluxes using a neutral stability assumption (Option 0) instead of applying a stability correction scheme (Option 1). Thus, the question arises why some parametrizations induce larger scale-errors than others. To assess this behavior in more detail, we compare SWE simulated by two configurations only differing in the choice of method for computing turbulent heat fluxes (Figure 14a and b). The figure displays results for one 50 km grid cell where the difference in SWE is particularly large compared to the 1 km resolution results.

During snow accumulation, the difference in SWE is small between the coarse (blue line) and upscaled fine scale resolution (green lines) results for both configurations. The small difference is due to the consistency in rainfall and snowfall amounts between the scales (see section 3.2). The simulations applying Option 0 show lower peak SWE than Option 1 due to occasional melt during snow cover buildup. Note that if we would have computed precipitation phase for the different resolutions, the scale-errors would probably have been much larger for SWE. The approach used here is, thus, an obvious and a computationally cheap solution to reduce scale-errors provided that high resolution forcings for air temperature and precipitation exist.

During snowmelt, the discrepancy in SWE between the two scales is larger than in the accumulation period, especially for Option 0. Furthermore, in particular for sensible heat

fluxes (panels e and f), but also for latent heat exchanges (panels g and h), we find that the configurations relying on a neutral stability assumption give larger scale-errors than the stability correction scheme. For net radiation, on the other hand, the scale-errors is rather similar between the two configurations (panels c and d). For the ablation phase, reducing the scale-error seems much more difficult than during the accumulation period. The scale-error in SWE appears to foremost depend on discrepancies in simulated turbulent heat fluxes between the two scales, which are larger for Option 0 than 1 (compare panel c and d as well as e and h). Whether such scale-errors can be reduced using subgrid parametrizations is difficult to judge due to complex interactions between different heat and energy fluxes, as well as time dependencies between different variables in energy-balance snow models, such as FSM used in this study.

As shown above, we find that the choice of parametrization for turbulent heat fluxes influences the simulations considerably, foremost for turbulent fluxes themselves but also for SWE (

Figure 5). For the 1 km resolution, the variability in simulated sensible heat fluxes is larger when applying a neutral stability assumption (see panel 15a) than when using the stability correction scheme (see panel 15b). Thus, the stability correction scheme dampens the turbulent heat fluxes in most situations through a smaller exchange coefficient over snow covered areas, which are typically dominated by stable rather than neutral or unstable conditions. The reduced exchange coefficient will also lead to a more linear function describing the turbulent heat fluxes when using Option 1 instead of 0 during the simulations (see Equation 20 in Essery [2015]). When upscaling the forcings (i.e. aggregation), the introduced error depends on (a) how non-linear governing equation is, and (b) how variable the high-resolution forcings are (for more details, see Figure 1 in Kirchner et al. [1993] and corresponding description). This effect likely causes the larger scale sensitivity of the configurations using Option 0 instead of 1 for computing turbulent heat fluxes. Likewise, the bucket formulation for routing liquid water through the snowpack introduces an additional nonlinearity that likely contributes to the scale-error. For a detailed discussion of how non-linear relationships influence models running on different resolutions, see also Rouholahnejad F & Kirchner [2017] who observed that simulated evapotranspiration increases when using larger grid resolutions.

Several earlier studies have found that SWE is reduced when coarsening the grid resolution for several land surface models (e.g. Boone et al. [2004], Dutra et al. [2011]). For the coastal areas, our simulations show similar behavior (Figure 10a and 11a). However, for the inland regions the coarse resolution simulations instead overestimate the snow amounts, which

was not reported in the studies mentioned above. In these studies, precipitation was given as liquid and solid phases directly to the models similar to this study (see section 3.2). Consequently, differences in SWE between the scales occur due to differences in melting and not accumulation of snow as noted above. The coherent spatial patterns in SWE bias indicate that, for example, the meteorological and topographic conditions along the coastline are such that they lead to an underestimation of SWE for the coarse compared to the fine grid resolution, whereas the opposite occurs in the inland regions. We have, however, not found any strong relationship between, for example, the skewness of the elevation distribution within the coarse scale grid cells and the observed biases. It is possible that several interacting factors cause these spatial patterns.

We find that topographic variability strongly influences the scale-error (Figure 13), which is in line with several earlier studies (e.g. Boone et al. [2004], Dutra et al. [2011], Rouholahnejad F & Kirchner [2017]). In mountainous regions, meteorological conditions can change considerably over short distances since variables such as air temperature and humidity largely depend on altitude. However, subgrid topographic variability did not explain all of the variance in the scale-error indicating that additional factors influence the discrepancy between the coarse and fine scale simulations. For example, precipitation shows a strong gradient from west to east in southern Norway due to the prevailing wind direction and orographic effects (Figure 2a). Likely, also the vicinity to the ocean introduces variability in local meteorological conditions along the coastline. Thus, further analysis of relationships between scale-errors and, for example, variability in wind speed and land-use properties such as forest coverage can give additional insight. In the coarse scale simulations, many of these local effects are smoothed out in the averaging process, which introduces errors due to the nonlinear governing equations used in FSM.

Further, we likely underestimate the scale-error due to the following omissions: (a) any variability below the 1 km resolution is not taken into account apart from snow covered fraction through a simple subgrid parametrization, (b) local effects caused by e.g. ice covered and urban areas are neglected, and (c) not all important processes have been taken into account (e.g. topographic shading, as well as redistribution of snow through avalanches and wind). Nevertheless, our study has shown which processes influence the scale-error strongest, and in which regions those errors are largest. One critical question is how to minimize these errors. One option is to use subgrid parametrizations (e.g. Helbig et al. [2015]), another is to run the models on very high resolution (e.g. Singh et al. [2015]). As outlined above, running models using a very fine grid requires large computational resources and is hampered by low quality forcing data in many regions. However, the last option might

still be more reliable than the subgrid parametrizations since these will hardly capture all combinations of peculiarities around the global mean caused by, for example, ocean influences, differing weather patterns and land-use types.

Parametrization procedures attempt to take into account all of the local scale heterogeneity of vegetation, topography, surface roughness, water stress and meteorological inputs that influence the integrated processes at a certain spatial scale. When local variations are sufficiently well integrated above a certain threshold scale, describing the variability of processes in the landscape is assumed to be sufficient to provide realistic predictions. However, comparing model simulations to observations may be misleading as neither of these correspond to the true process scale. The spatial resolution, extent and support (integration volume) of the observation and model scales act as a filter relative to the scale of natural variability. Processes with larger spatial scale than the spatial extent of observations appear as trends in the data, whereas processes with smaller spatial than the resolution appear as noise. This scale triplet will be different for the various processes influencing land surface atmosphere interactions and will also vary between different geographic regions and landscape types, e.g. below and above the tree level. Although it may be concluded that coarse scale models may perform well, it is nevertheless better to apply models with fine spatial (and temporal) resolution if possible.

6 Conclusions

In this study, we have assessed the sensitivity of an energy-balance snow model to changes in grid resolution using a multi-parametrization framework. The framework has allowed us to systematically test how different parametrizations describing a set of processes influence the discrepancy between the coarser (5 to 50 km) and finest (1 km) resolutions, here termed the scale-error. The simulations were setup such that liquid and solid precipitation was identical between the different resolution, and differences between the simulations arise mainly during the ablation period. We have focused our analysis on evaluating the scale-error for snow water equivalent, which is important for water-related questions, as well as surface heat fluxes and net radiation, both relevant for weather forecasting and climate model simulations.

This study confirms results from other studies showing that the grid resolution is an important aspect in land surface modelling, and that the simple choice of grid size introduces errors in the results (e.g. Boone et al., [2004], Dutra et al. [2011], Haddeland et al. [2002]).

However, unlike earlier studies, we have been able to assess in detail which process parametrizations introduce the largest errors for different variables. For snow water equivalent, we find that the choice of method for routing liquid water through the snowpack influences the scale sensitivity most, followed by the parametrizations for computing turbulent heat fluxes and albedo. For turbulent heat exchanges, on the other hand, only one process parametrization seems to influence the scale-error largely, that is whether one chooses to assume a neutral stability of the atmosphere or includes a stability correction scheme for the flux estimates. For net radiation, the choice of method for computing surface albedo seems to influence the results most, whereas the other process parametrizations seem less influential. Our analysis reveals which type of parametrizations introduces errors between different grid resolutions, and that complex interactions can arise between different processes that determine the magnitude of the scale-error. All these findings are very important since most models contain various user defined settings, e.g. for computing turbulent fluxes, and small changes in the options can introduce large scale-errors in coarse resolution simulations.

Only few studies have evaluated the model sensitivity for grid sizes ranging from so-called hyper-resolution models to global climate simulations (e.g. Singh et al. [2015]). In this study, we have therefore chosen to cover this range using grid resolutions from 1 to 50 km. When considering the whole study domain, we find that the scale-error increases with resolution, with the largest error between the coarsest (50 km) and finest (1 km) grid sizes for all variables. However, already coarsening the grid resolution from 1 to 5 km introduces a substantial scale-error when comparing to the error for the 50 km resolution. The ranking of parametrizations sorted from lowest to highest scale-error mostly remains consistent for the different resolutions. For example, for the simulated turbulent heat fluxes the application of a stability correction scheme shows lower scale-errors than the neutral stability assumption for all model resolutions. Likely, the latter parametrization shows a stronger nonlinearity than the former and therefore induces larger scale-errors. However, we cannot select models based on the criteria of low scale-errors alone since those with lowest scale sensitivity may not show the best performance, as our case study has shown.

For all variables, we find that the scale-error increases with subgrid topographic variability, which is inline with earlier findings (Boone et al., 2004). Such effects induce coherent spatial patterns in the scale-error including biases between the coarse and fine resolution results. However, we also observe that the scale-error likely depends on subgrid variability in, for example, meteorological conditions induced by coastal and orographic effects. Thus, potential scale-errors between coarse and fine resolution simulations can take on rather

complex patterns, as we have observed for snow water equivalent. For this variable, the coarse scale simulations show an underestimation along the coastline, whereas the bias is positive in mountainous regions.

In recent years, hyper-resolution modelling has been much discussed in the hydrological and meteorological communities (e.g. Bierkens et al. [2015], Singh et al. [2015]). Nowadays, several models at the continental scale are indeed operated with 4 km grid spacing or finer (e.g. Liu et al. [2017], Milbrandt et al. [2016]). For snow models, it seems obvious to run them at highest possible resolution because this is the scale we typically can evaluate them at and, even more important, any upscaling will introduce errors that regionally can be large due to nonlinearities in the energy balance formulations. The main challenge to minimize such errors will be to develop reliable methods for generating the forcing data needed for high resolution simulations. From a practical perspective, it might be of advantage to establish collaborations for sharing forcing data and computational resources such that the simulations of various models can be performed on the same platform as where the data is stored.

7 Acknowledgments

This work has been supported by the Norwegian Research Council under grant no. 243803/E10 (I:CAN - Impacts: Climate, Anthroposphere and Nature). Development of FSM2 is supported by NERC grant NE/P011926/1. Meteorological data were provided by METNO (<http://thredds.met.no/thredds/catalog.html>) and land cover/forest structure data by Majasalmi et al. (2018) are available from NIBIO (<https://doi.org/10.21350/c9j08bz3>). We thank the three reviewers for their constructive suggestions. Simulations used in this study can be made available upon request. The MTCLIM code for generating the surface forcing fields was available as part of VIC 4.2.d (<https://doi.org/10.5281/zenodo.56058>) and the snow simulation were performed based on FSM v2.0.0 (<https://github.com/RichardEssery/FSM2>).

8 References

Bartlett, P. A., MacKay, M. D., & Verseghy, D. L. (2006). Modified snow algorithms in the Canadian land surface scheme: Model runs and sensitivity analysis at three boreal forest stands. *Atmosphere-Ocean*, 44(3), 207–222.

<https://doi.org/10.3137/ao.440301>

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., et al.

(2015). Hyper-resolution global hydrological modelling: what is next?: “Everywhere and locally relevant.” *Hydrological Processes*, 29(2), 310–320.

<https://doi.org/10.1002/hyp.10391>

Bohn, T. J., Livneh, B., Oyster, J. W., Running, S. W., Nijssen, B., & Lettenmaier, D. P.

(2013). Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. *Agricultural and Forest Meteorology*, 176, 38–49.

<https://doi.org/10.1016/j.agrformet.2013.03.003>

Boone, A., Habets, F., Noilhan, J., Clark, D., Dirmeyer, P., Fox, S., et al. (2004). The Rhône-Aggregation Land Surface Scheme Intercomparison Project: An Overview. *Journal of Climate*, 17, 22.

Deardorff, J.W., 1978. Efficient prediction of ground surface temperature and moisture, with an inclusion of a layer of vegetation. *J. Geophys. Res.* 83 (N64), 1889–1903, <http://dx.doi.org/10.1029/JC083iC04p01889>.

Dutra, E., Kotlarski, S., Viterbo, P., Balsamo, G., Miranda, P. M. A., Schär, C., et al. (2011). Snow cover sensitivity to horizontal resolution, parameterizations, and atmospheric forcing in a land surface model. *Journal of Geophysical Research: Atmospheres*, 116(D21). <https://doi.org/10.1029/2011JD016061>

Dyrørdal, A.V, Saloranta, T., Skaugen, T & Stranden, H.B. (2013). Changes in snow depth in Norway during the period 1961-2010. *Hydrology Research* 44, 169-179. <https://doi.org/10.2166/nh.2012.064>

Essery, R. (2003). Aggregated and distributed modelling of snow cover for a high-latitude basin. *Global and Planetary Change*, 38(1–2), 115–120. [https://doi.org/10.1016/S0921-8181\(03\)00013-4](https://doi.org/10.1016/S0921-8181(03)00013-4)

Essery, R., Pomeroy, J., Parvianen, J., Storck, P. (2003). Sublimation of snow from boreal forests in a climate model. *Journal of Climate*, 16, 1855-1864.

Essery, R. (2015). A factorial snowpack model (FSM 1.0). *Geoscientific Model Development*,

8(12), 3867–3876. <https://doi.org/10.5194/gmd-8-3867-2015>

Gjertsen, A. K., & Nilsen, J.-E. (2012). *Et skogkart basert på tolking av satellittbilder* (No.

23/2012). NO- 1431 Ås: Norsk institutt for skog og landskap. Retrieved from

<https://brage.bibsys.no/xmlui/bitstream/handle/11250/2453917/SoL-Rapport-2012-23.pdf?sequence=2&isAllowed=y>

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Technical

Note: Downscaling RCM precipitation to the station scale using statistical

transformations - a comparison of methods. *Hydrology and Earth System Sciences*, 16(9), 3383–3390. <https://doi.org/10.5194/hess-16-3383-2012>

Haddeland, I., Matheussen, B. V., & Lettenmaier, D. P. (2002). Influence of spatial resolution

on simulated streamflow in a macroscale hydrologic model. *Water Resources*

Research, 38(7), 29-1-29–10. <https://doi.org/10.1029/2001WR000854>

Helbig, N., van Herwijnen, A., Magnusson, J., & Jonas, T. (2015). Fractional snow-covered

area parameterization over complex topography. *Hydrology and Earth System*

Sciences, 19(3), 1339–1351. <https://doi.org/10.5194/hess-19-1339-2015>

Hungerford, R. D., Nemani, R. R., Running, S. W., & Coughlan, J. C. (1989). *MTCLIM: a*

mountain microclimate simulation model (No. INT-RP-414). Ogden, UT: U.S.

Department of Agriculture, Forest Service, Intermountain Forest and Range

Experiment Station. <https://doi.org/10.2737/INT-RP-414>

Kirchner, J. W., Dillon, P. J., & Lazerte, B. D. (1993). Predictability of geochemical buffering

and runoff acidification in spatially heterogeneous catchments. *Water Resources*

Research, 29(12), 3891–3901. <https://doi.org/10.1029/93WR02202>

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., Chen, F., Chen,

L., Clark, M., Dai, A., Dudhia, J., Eidhammer, T., Gochis, D., Gutmann, E., Kurkute,

S., Li, Y., Thompson, G., Yates, D. (2017). Continental-scale convection-permitting

modeling of the current and future climate of North America. *Climate Dynamics*, 49(1-2), 71-95.

Lussana, C. (2017). *Spatial Interpolation of daily minimum, maximum and mean temperature*

(No. 02/2017). Norwegian Meteorological Institute.

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., & Andersen, J.

(2018). seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. *Earth System Science Data*, 10(1), 235–249.

<https://doi.org/10.5194/essd-10-235-2018>

Lussana, C., Tveito, O. E., Dobler, A., & Tunheim, K. (2019). seNorge_2018, daily

precipitation and temperature datasets over Norway. *Earth System Science Data Discussion*, <https://doi.org/10.5194/essd-2019-43>

Majasalmi, T., Eisner, S., Astrup, R., Fridman, J., & Bright, R. M. (2018). An enhanced forest classification scheme for modeling vegetation–climate interactions based on national forest inventory data, 14.

Milbrandt, J. A., Bélair, S., Faucher, M., Vallée, M., Carrera, M. L., & Glazer, A. (2016). The pan-Canadian high resolution (2.5 km) deterministic prediction system. *Weather and Forecasting*, 31(6), 1791–1816.

Müller, M., Homleid, M., Ivarsson, K.-I., Køltzow, M. A. Ø., Lindskog, M., Midtbø, K. H., et al. (2017). AROME-MetCoOp: A Nordic Convective-Scale Operational Weather Prediction Model. *Weather and Forecasting*, 32(2), 609–627.

<https://doi.org/10.1175/WAF-D-16-0099.1>

Noilhan, J., & Mahfouf, J.-F. (1996). The ISBA land surface parameterisation scheme. *Global and Planetary Change*, 13(1–4), 145–159. [https://doi.org/10.1016/0921-8181\(95\)00043-7](https://doi.org/10.1016/0921-8181(95)00043-7)

Oleson, K. W., Lawrence, D. M., Flanner, M. G., Kluzek, E., Levis, S., Swenson, S. C., et al. (2010). Technical Description of version 4.0 of the Community Land Model (CLM), 266.

Reistad, M., Breivik, Ø., Haakenstad, H., Aarnes, O. J., Furevik, B. R., & Bidlot, J.-R. (2011). A high-resolution hindcast of wind and waves for the North Sea, the Norwegian Sea, and the Barents Sea. *Journal of Geophysical Research*, 116(C5).

<https://doi.org/10.1029/2010JC006402>

- Rouholahnejad F, E., & Kirchner, J. W. (2017). A Budyko framework for estimating how spatial heterogeneity and lateral moisture redistribution affect average evapotranspiration rates as seen from the atmosphere. *Hydrology and Earth System Sciences*, 21(1), 217–233. <https://doi.org/10.5194/hess-21-217-2017>
- Saloranta, T. M. (2012). Simulating snow maps for Norway: description and statistical evaluation of the seNorge snow model. *The Cryosphere*, 6(6), 1323–1337. <https://doi.org/10.5194/tc-6-1323-2012>
- Skaugen, T., Stranden, H. B., Saloranta, T. (2012). Trends in snow water equivalent in Norway (1931-2009). *Hydrology Research* 43, 489-499. <https://doi.org/10.2166/nh.2012.109>
- Singh, R. S., Reager, J. T., Miller, N. L., & Famiglietti, J. S. (2015). Toward hyper-resolution land-surface modeling: The effects of fine-scale topography and soil texture on CLM4.0 simulations over the Southwestern U.S. *Water Resources Research*, 51(4), 2648–2667. <https://doi.org/10.1002/2014WR015686>
- Stöckli, R., Vidale, P. L., Boone, A., & Schär, C. (2007). Impact of Scale and Aggregation on the Terrestrial Water Exchange: Integrating Land Surface Models and Rhône Catchment Observations. *Journal of Hydrometeorology*, 8(5), 1002–1015. <https://doi.org/10.1175/JHM613.1>
- Tennessee Valley Authority, 1972. Heat and mass transfer between a water surface and the atmosphere. Tennessee Valley Authority, Norris, TN. Laboratory report no. 14. Water resources research report no. 0-6803.
- Van den Hoof, C., Vidale, P. L., Verhoef, A., & Vincke, C. (2013). Improved evaporative flux partitioning and carbon flux in the land surface model JULES: Impact on the simulation of land surface processes in temperate Europe. *Agricultural and Forest Meteorology*, 181, 108–124. <https://doi.org/10.1016/j.agrformet.2013.07.011>
- Verseghy, D. L. (1991). Class-A Canadian land surface scheme for GCMS. I. Soil model. *International Journal of Climatology*, 11(2), 111–133. <https://doi.org/10.1002/joc.3370110202>

Wolff, M. A., Isaksen, K., Petersen-Øverleir, A., Ødemark, K., Reitan, T., & Brækkan, R.

(2015). Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study. *Hydrology and Earth System Sciences*, 19(2), 951–967. <https://doi.org/10.5194/hess-19-951-2015>

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., et al.

(2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, 47(5).

<https://doi.org/10.1029/2010WR010090>

Yen, Y.-C. (1981). *Review of Thermal Properties of Snow, Ice and Sea Ice* (No. CRREL

Report 81-10) (p. 37). Hanover, NH: US Army Core of Engineers Cold Regions Research and Engineering Laboratory.

9 Tables

Table 1. Description of snowpack processes with available options for parametrizations in FSM.

Process	Description
Snow albedo	The variations in reflectivity of snow depending on, for example, grain types and the incident angle of shortwave radiation. Option 0 - a diagnostic parametrization that computes the albedo from snow surface temperature. Option 1 - a prognostic parametrization using the snow age and snowfall events.
Thermal conductivity of snow	The thermal conductivity of snow, varying with factors such as snow density, affecting the heat flux through the snowpack. Option 0 - the snow conductivity is set constant. Option 1 - the snow conductivity is computed using the density dependent formulation presented by Yen [1981].
Snow compaction	The increase in snow density due to, for example, metamorphosis and weight of overlying snow. Option 0 - the snow density is assumed constant. Option 1 - the snow density increases with time following the parametrization given by Verseghy [1991].
Turbulent heat exchange	The turbulent exchange of heat and moisture between the snow or soil surface and the atmosphere. Option 0 - turbulent heat exchanges are computed assuming neutral atmosphere stability. Option 1 - turbulent heat exchanges are calculated using the atmosphere stability correction. See Essery [2015] for more details.
Snow hydraulics	The process for routing liquid water through the snowpack. Option 0 - the snowpack cannot hold any liquid water and melt or rainwater drains immediately. Option 1 - a so-called bucket formulation is used where liquid water flow occurs when the irreducible water content is exceeded. See Essery [2015] for more details.

10 Figures

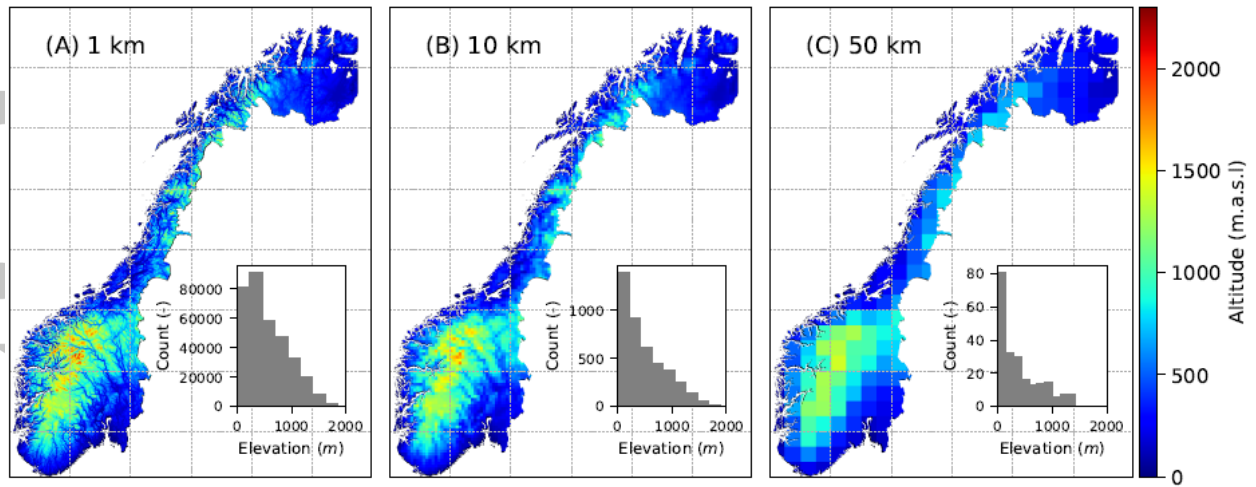


Figure 1. Topography of the study region depicted in three different resolutions. The 50 km resolution only captures the large-scale features of the 1km map. Note that we cutoff grid cells of the coarser resolutions, 10 and 50 km, at country borders and water bodies since we lack high resolution land-use data for those regions. Averages are thus only displayed for available land areas for those grid cells.

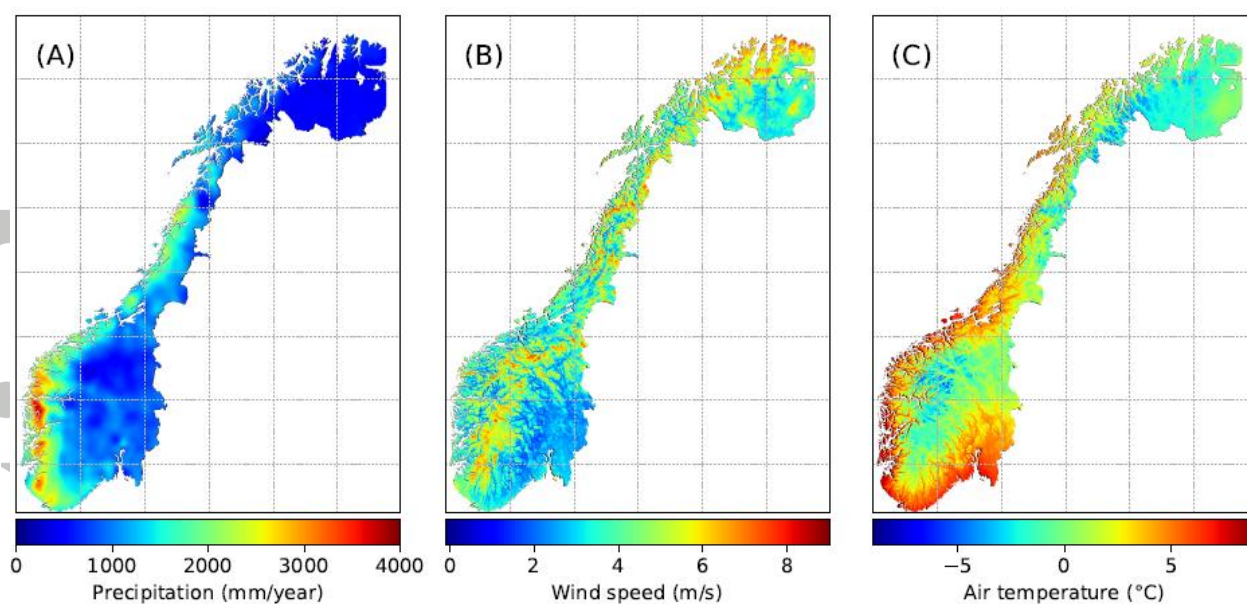


Figure 2. Mean annual precipitation, wind speed and air temperature for the study region and period depicted on the 1 km resolution grid. The averages were computed using the data for the period from 2008-9-1 to 2011-9-1.

Accepted

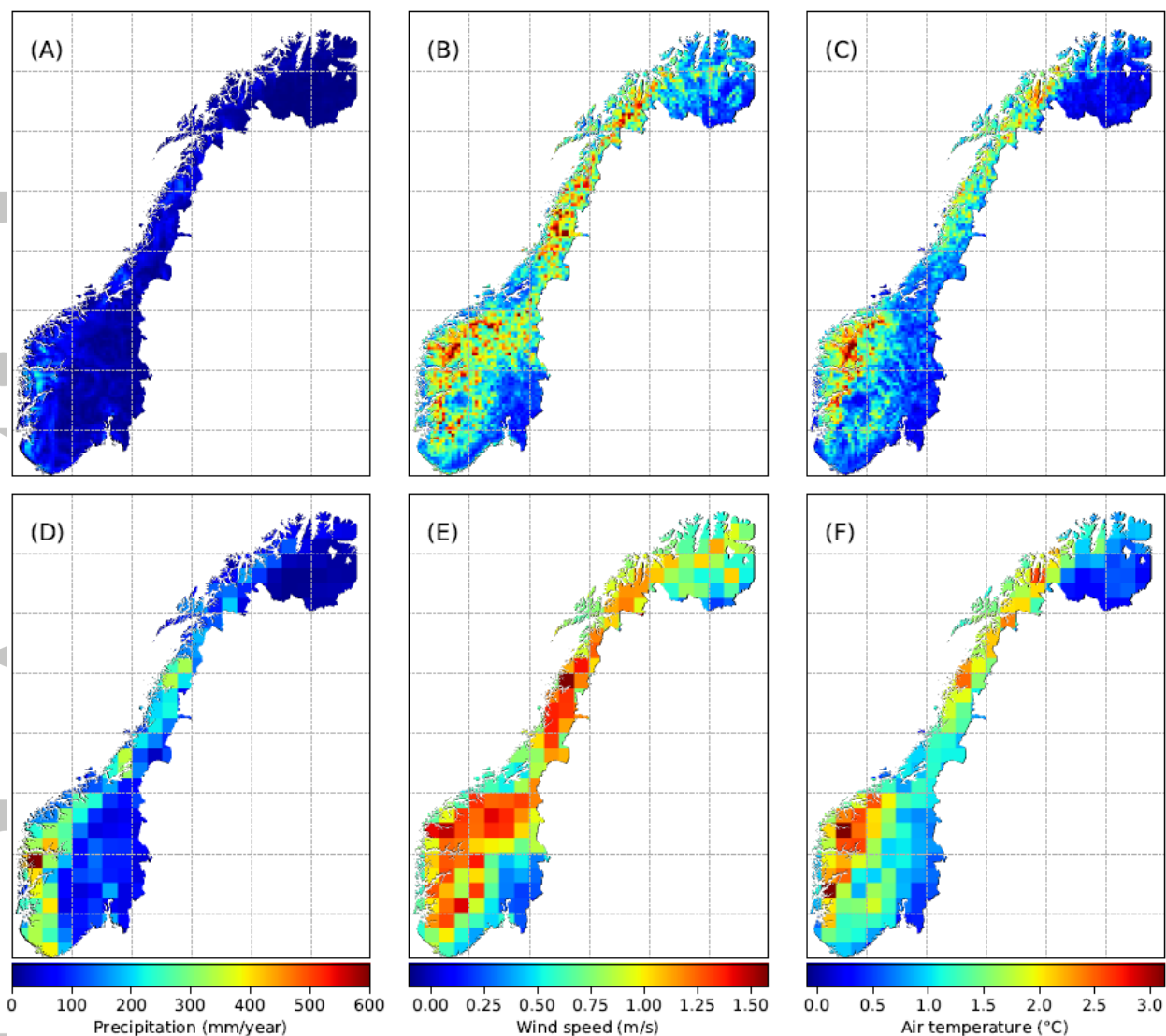


Figure 3. Standard deviation of precipitation (left panels), wind speed (mid panels) and air temperature (right panels) within 10 by 10 km (upper row) and 50 by 50 km (lower row) large grid cells computed using the time-averaged data of the 1 km grid (see results in Figure 2).

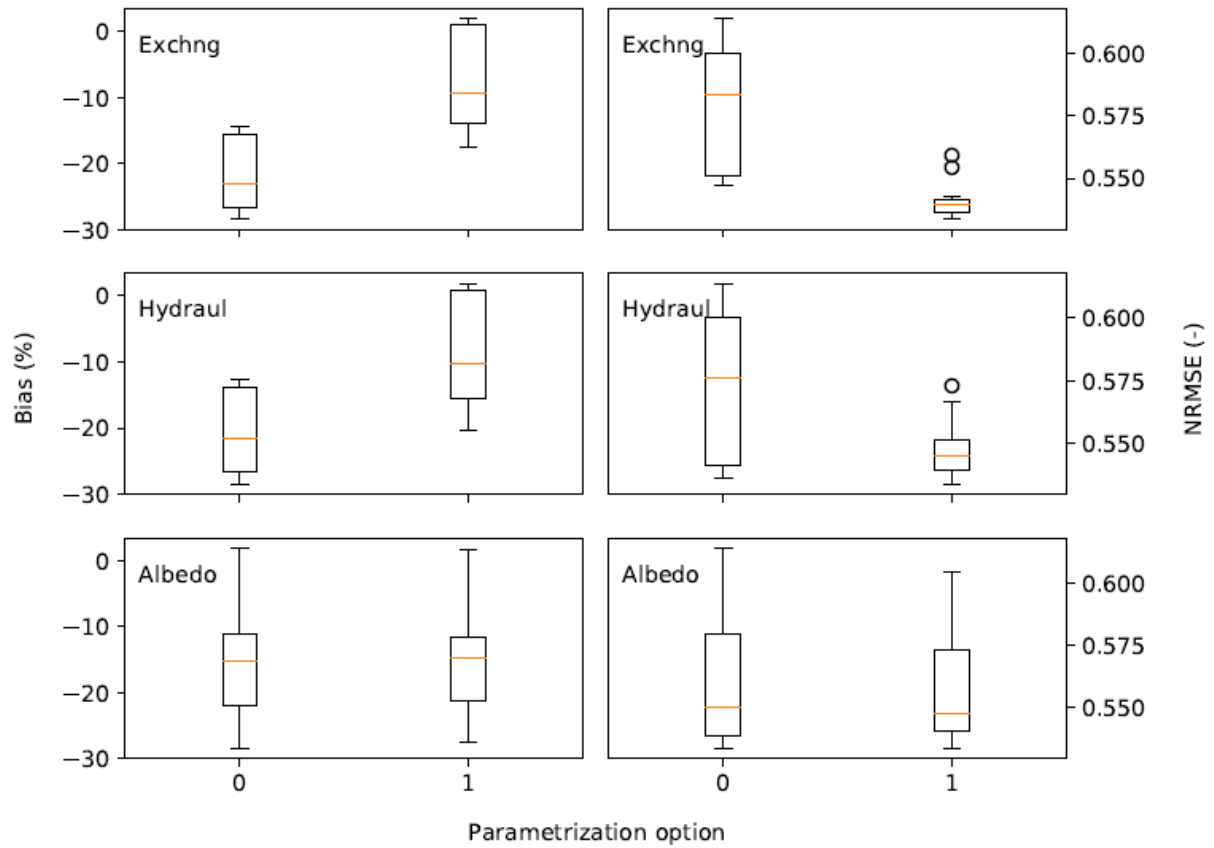


Figure 4. Model performance for SWE in terms of percentage bias and normalized root-mean-square-error (NRMSE) for three different processes (turbulent heat exchange, snow hydraulics and snow albedo) and depending on the choice of parametrization (Option 0 or 1). The figure shows data from all 32 possible model configurations, and RMSE was normalized using the average of all SWE observations. The boxes extend from the lower to upper quartile values of the data, with the line indicating the median. The position of the whiskers is set to 1.5 times the interquartile range from the edges of the box, and points denote outliers.

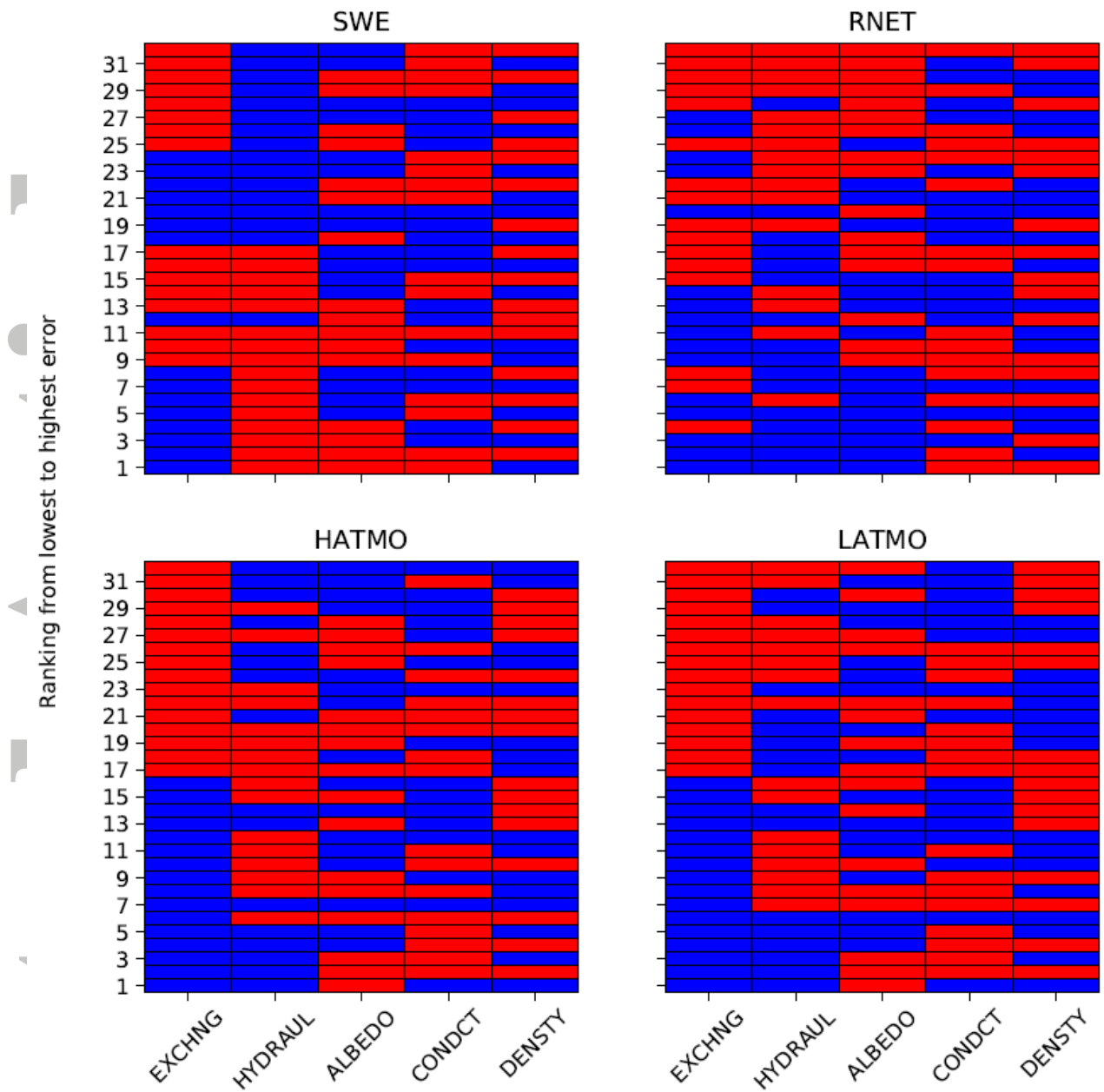


Figure 5. Ranking of the model configurations from lowest to highest scale-error for snow water equivalent (SWE), net radiation (RNET), latent heat fluxes (LATMO) and sensible heat fluxes (HATMO). Here the scale-error was given by the RMSE between the upscaled fine resolution (1 km) and coarse resolution (50 km) simulations. This error statistic was computed as outlined in section 3.3. Red colors denote Option 0 and blue colors Option 1 for the parametrizations.

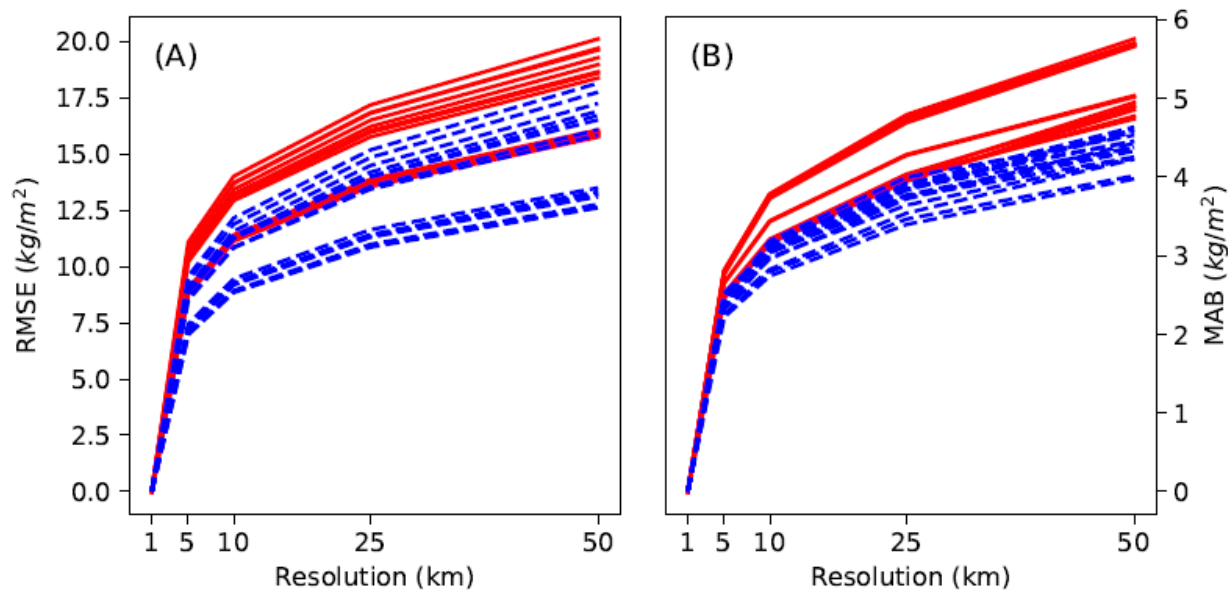


Figure 6. RMSE and MAB between the 1 to 50 km and 1 km gridded simulations for SWE. These statistics for the scale-error were computed using the methods outlined in section 3.3. Red colors denote Option 0 and blue colors Option 1 for parametrization of the turbulent heat fluxes.

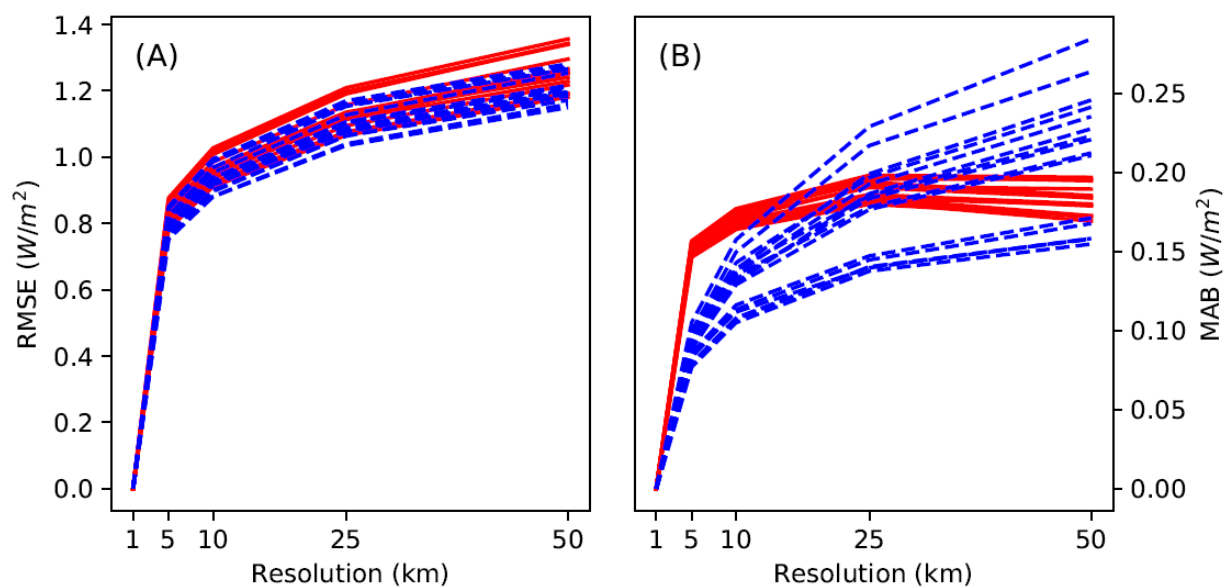


Figure 7. RMSE and MAB between the 1 to 50 km and 1 km gridded simulations for RNET. These statistics for the scale-error were computed using the methods outlined in section 3.3. Red colors denote Option 0 and blue colors Option 1 for parametrization of the turbulent heat fluxes.

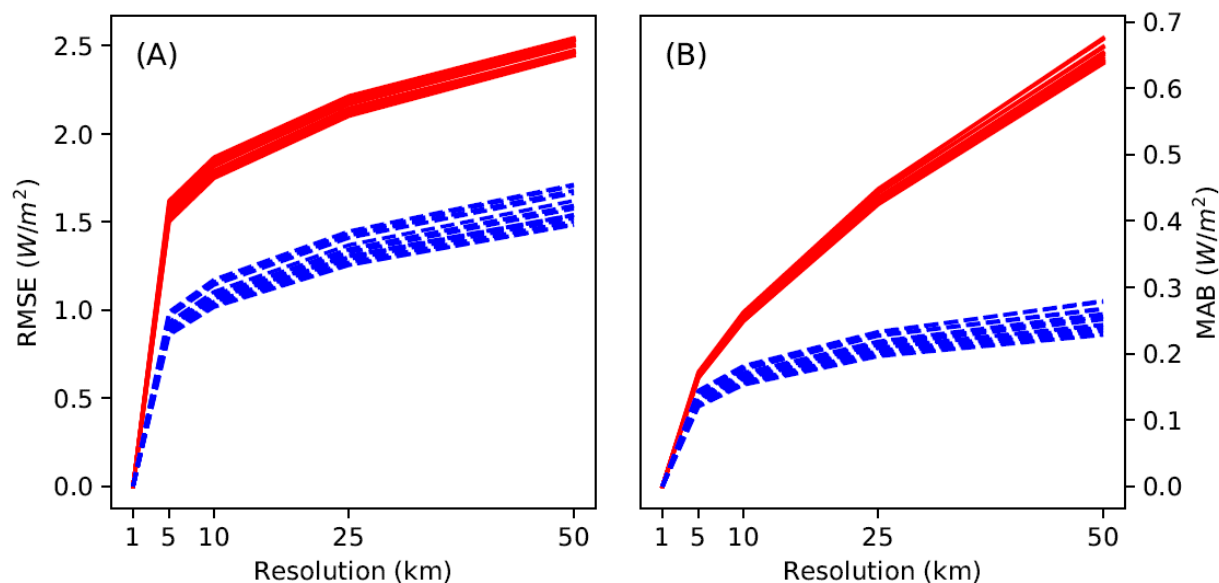


Figure 8. RMSE and MAB between the 1 to 50 km and 1 km gridded simulations for HATMO. These statistics for the scale-error were computed using the methods outlined in section 3.3. Red colors denote Option 0 and blue colors Option 1 for parametrization of the turbulent heat fluxes.

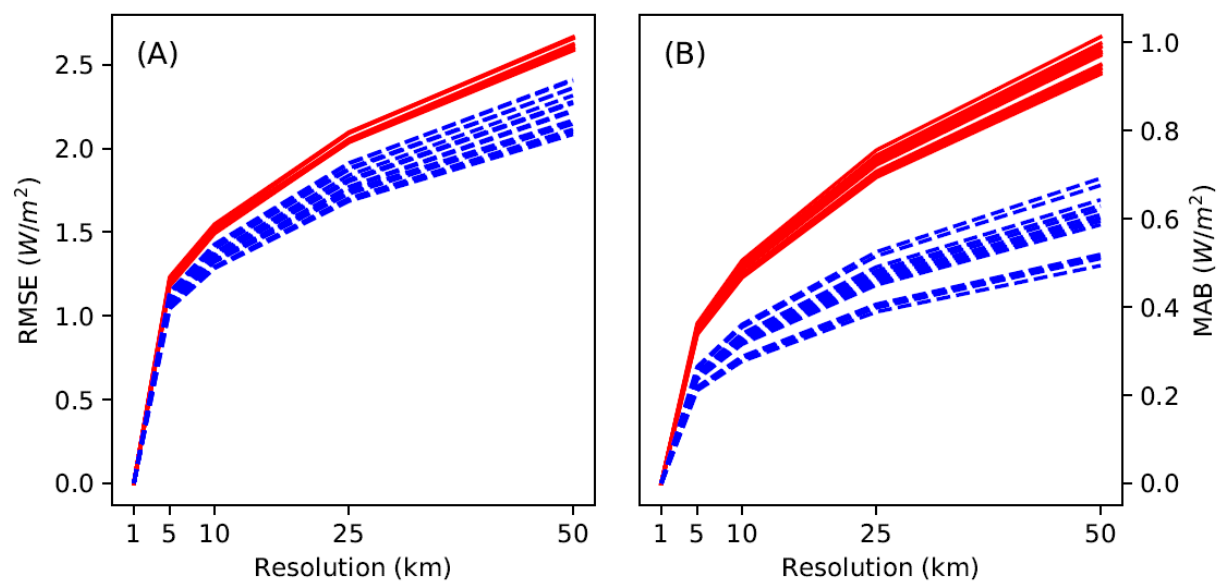


Figure 9. RMSE and MAB between the 1 to 50 km and 1 km gridded simulations for LATMO. These statistics for the scale-error were computed using the methods outlined in section 3.3. Red colors denote Option 0 and blue colors Option 1 for parametrization of the turbulent heat fluxes.

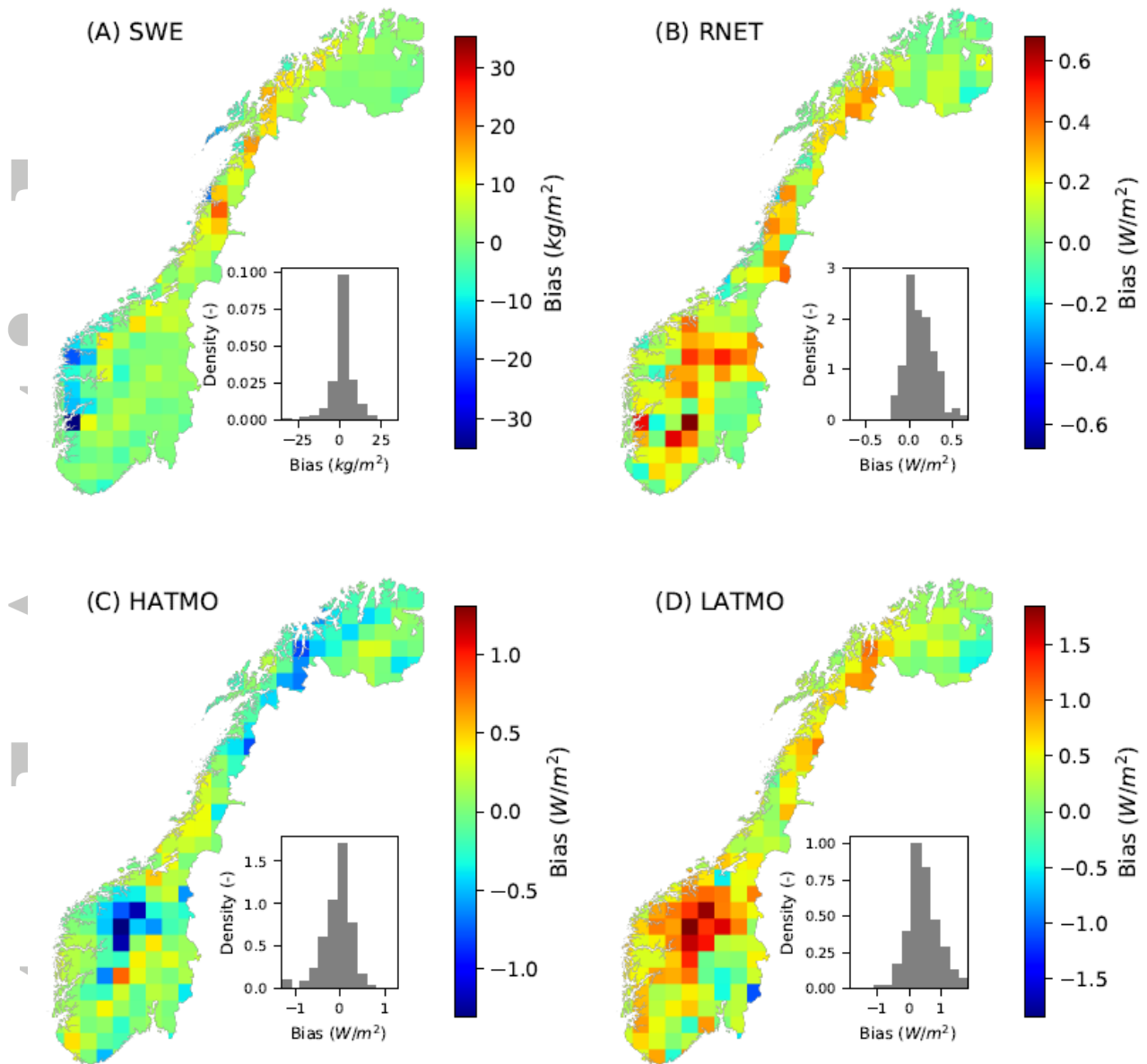


Figure 10. Scale-error in terms of bias between the coarse (50 km) and fine (1 km) scale simulations for snow water equivalent (A), net radiation (B), sensible heat fluxes (C) and latent heat fluxes (D). The figure shows simulations where all process parametrizations are set to Option 1, also for turbulent heat exchanges. The biases are computed per grid cell and for the whole simulation period.

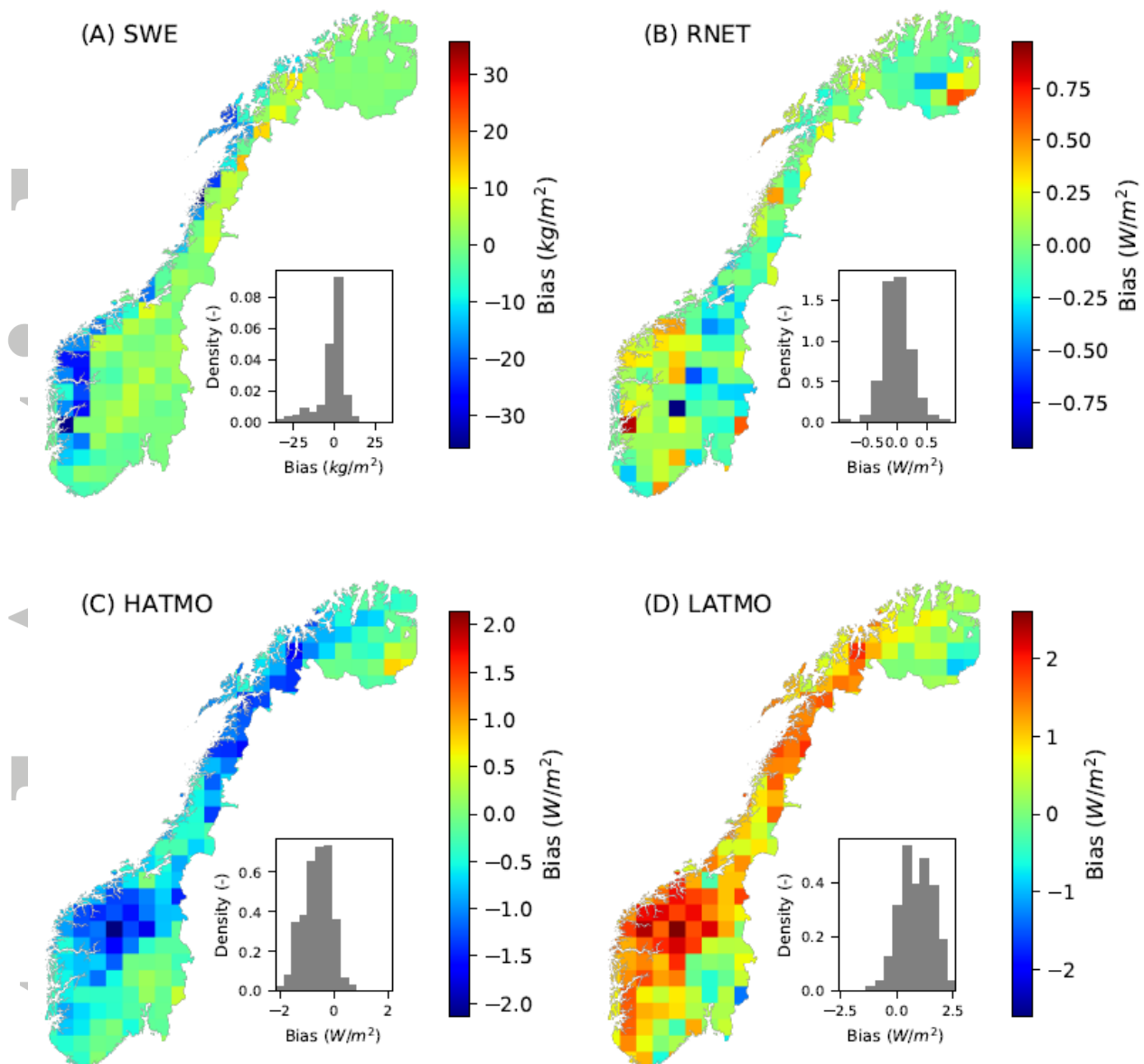


Figure 11. Scale-error in terms of bias between the coarse (50 km) and fine (1 km) scale simulations for snow water equivalent (A), net radiation (B), sensible heat fluxes (C) and latent heat fluxes (D). The figure shows simulations where all process parametrizations are set to Option 1, except for turbulent heat exchanges that was set to 0 (assuming neutral atmospheric conditions). The biases are computed per grid cell and for the whole simulation period.

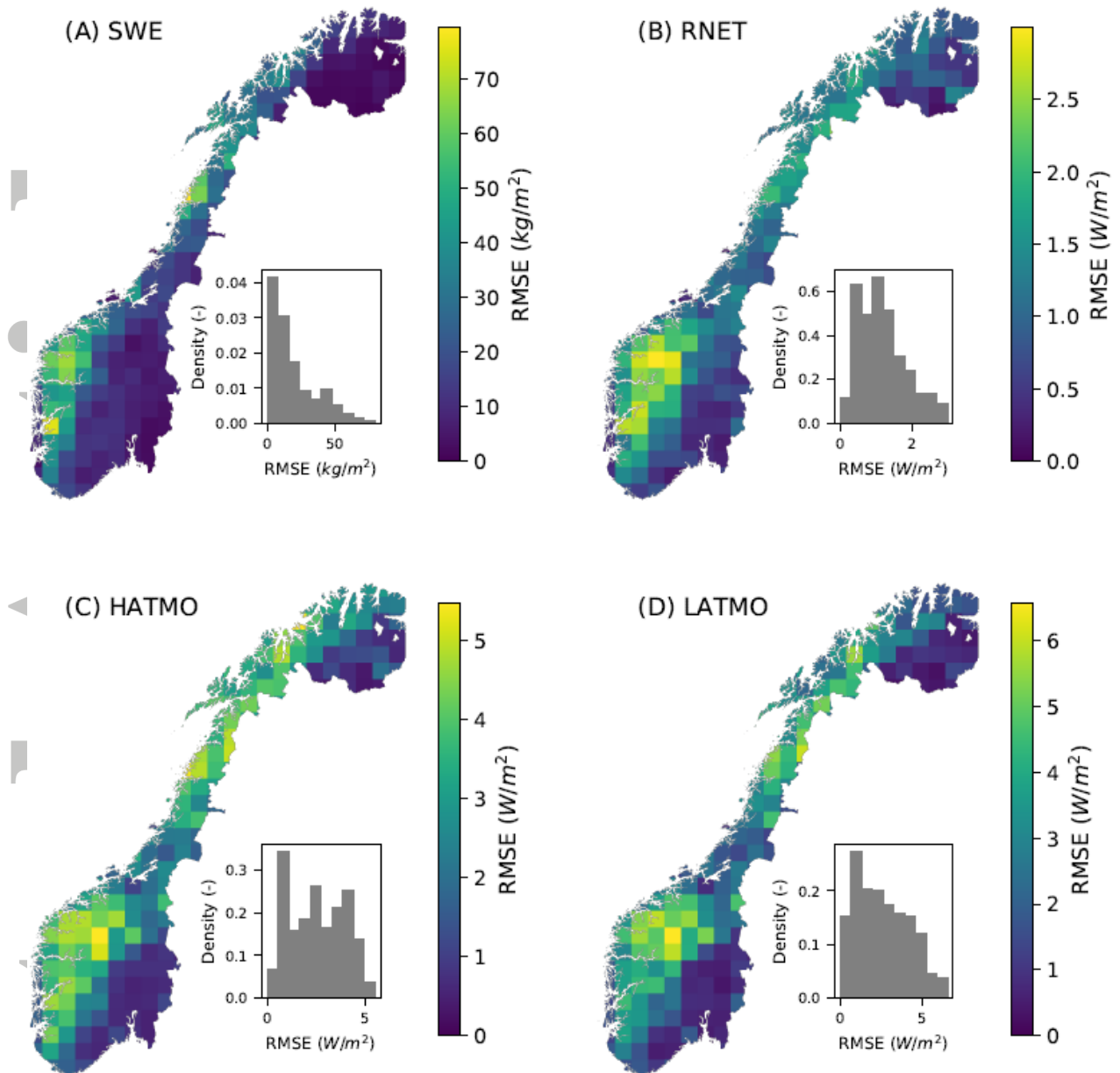


Figure 12. Scale-error in terms of RMSE between the coarse (50 km) and fine (1 km) scale simulations for snow water equivalent (A), net radiation (B), sensible heat fluxes (C) and latent heat fluxes (D). The figure shows simulations where all process parametrizations are set to Option 1.

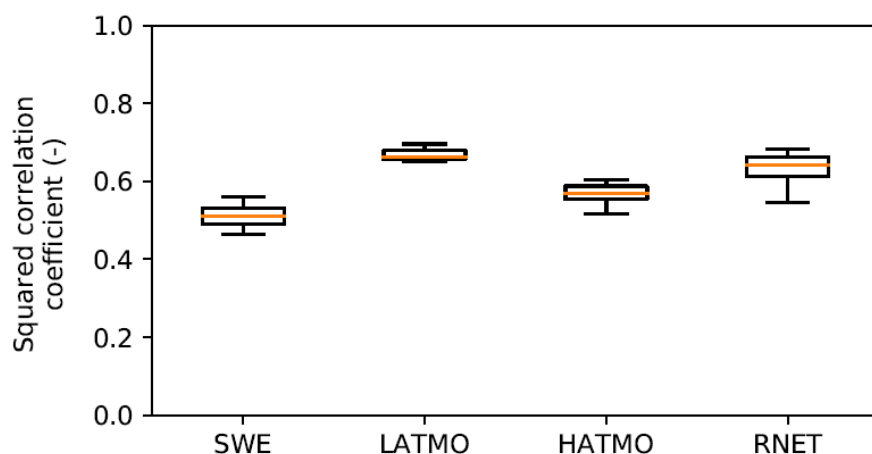


Figure 13. Squared correlation coefficient between the scale-error and the subgrid topographic variability for snow water equivalent, latent heat fluxes, sensible heat fluxes and net radiation. Here, we defined the scale-error as the RMSE between the 50 and upscaled 1 km resolution simulations. The topographic variability within each 50 km grid cell was computed using the standard deviation of the 1 km elevation model. The boxes show the spread for the 32 different model configurations.

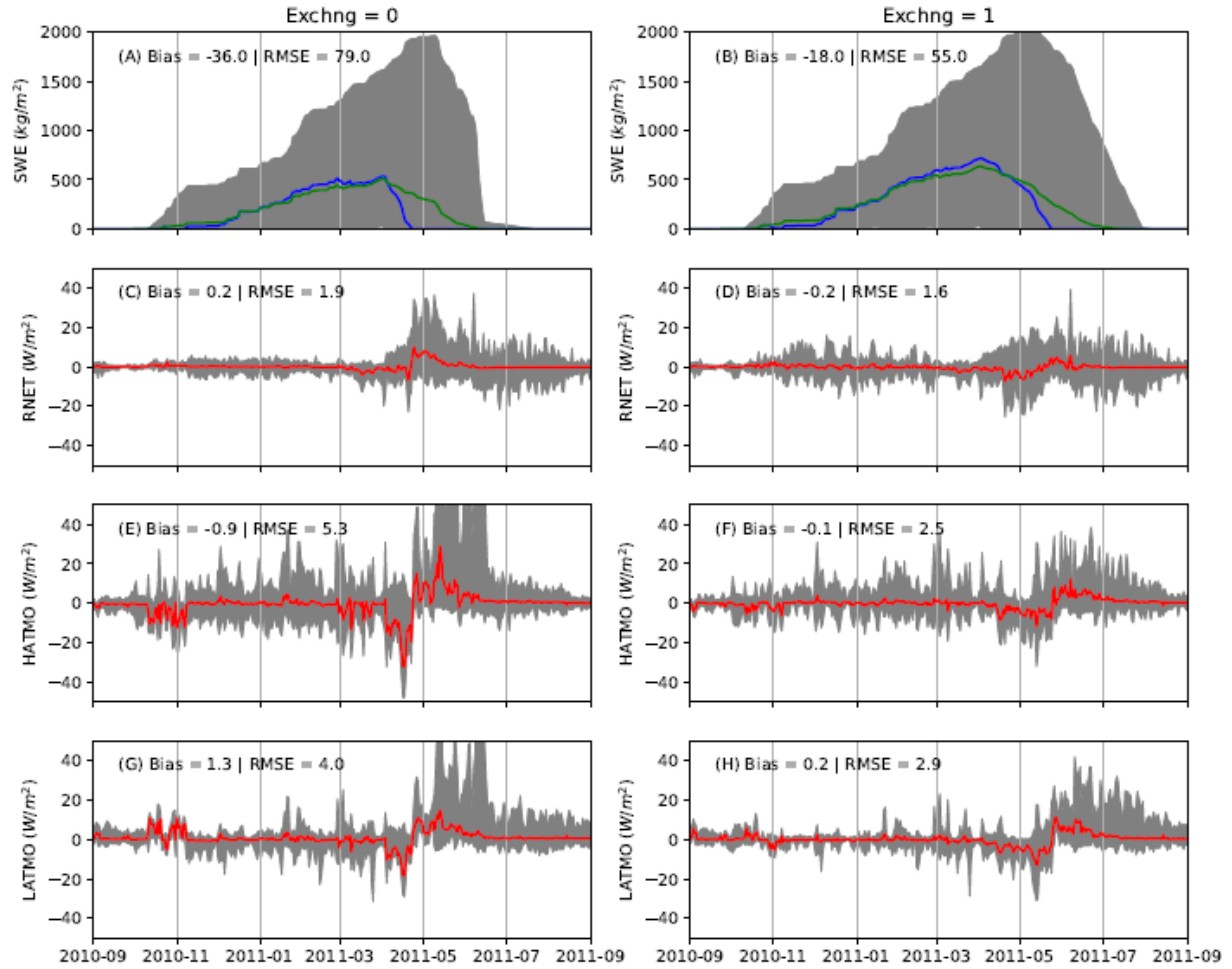


Figure 14. SWE for the 1 km (gray shaded area), average of the 1 km (green line), and the 50 km resolution (blue line) for winter 2010/2011 (panel a and b). Scale-error (i.e. difference between the 50 and 1 km resolution) for net radiation (panel c and d), sensible heat fluxes (panel e and f), and latent heat fluxes (panel g and h) for all 1 km grid cells (gray shaded area), and the averaged 1 km results (red line). For the right panels, Option 1 was used or all processes. For the left panels, we used Option 1 for all processes except the turbulent heat exchanges, which was set to Option 0. The figure shows results from the grid cell with highest RMSE for SWE between the 50 and upscaled 1 km resolution runs (see southwestern part of Norway in Figure 12).

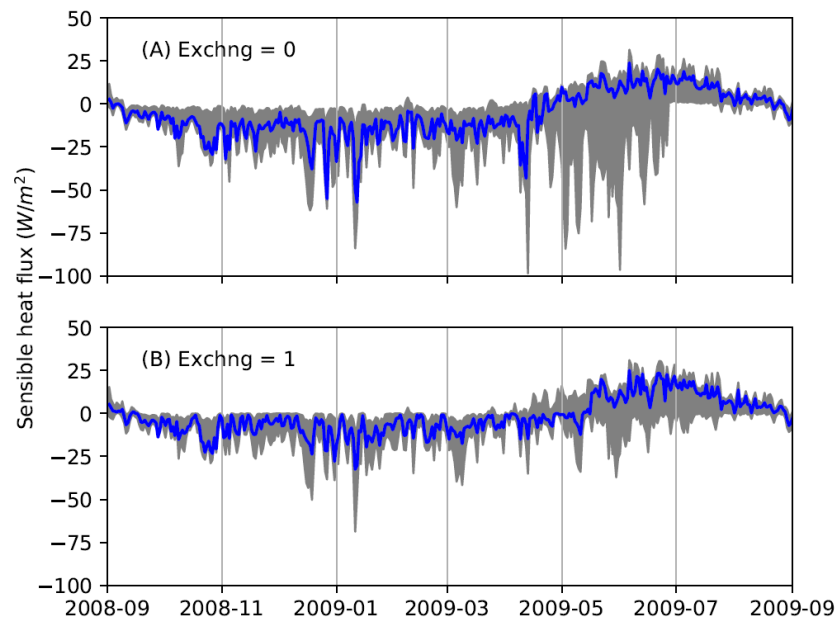


Figure 15. The upper panel shows the results from the configuration using Option 1 for all process apart from turbulent heat exchanges, which was set to Option 0. The lower panel shows the results from the configuration where Option 1 was used for all processes. The gray shaded area shows the results from 1 km simulations contained within the 50 km resolution grid cell depicted by a blue line. The figure shows results from the grid cell with highest RMSE for SWE between the 50 and upscaled 1 km resolution runs (see southwestern part of Norway in Figure 12).