Optimization of Snow-Related Parameters in Noah Land Surface Model (v3.4.1) Using Micro-Genetic Algorithm (v1.7a)

Sujeong Lim[1,2], Hyeon-Ju Gim[3], Ebony Lee[1,2,4], Seungyeon Lee[1,2,4], Won Young Lee[1,2], Yong Hee Lee[5], Claudio Cassardo[6], and Seon Ki Park[1,2,4]

[1]Center for Climate/Environment Change Prediction Research, Ewha Womans University, Seoul, 03760, Republic of Korea
[2]Severe Storm Research Center, Ewha Womans University, Seoul, 03760, Republic of Korea
[3]Korea Institute of Atmospheric Prediction System (KIAPS), Seoul, 07071, Republic of Korea
[4]Department of Climate and Energy System Engineering, Ewha Womans University, Seoul, 03760, Republic of Korea
[5]High Impact Weather Research Department, National Institute of Meteorological Sciences, Gangneung, 25457, Republic of Korea
[6]Department of Physics and NatRisk Centre, University of Torino, Torino, 10125, Italy

Correspondence: Seon Ki Park (spark@ewha.ac.kr)

Abstract. The snowfall prediction is important in winter and early spring because snowy conditions generate enormous economic damages. However, there is a lack of previous studies dealing with snow prediction, especially using land surface models (LSMs). Numerical weather prediction models directly interpret the snowfall events, whereas the LSMs evaluate the snow cover fraction, snow albedo, and snow depth through interaction with atmospheric conditions. When the initially-developed empirical parameters are local or inadequate, we need to optimize the parameter sets for a certain region. In this study, we seek for the optimal parameter values in the snow-related processes — snow cover fraction, snow albedo, and snow depth — of the Noah LSM, for South Korea, using the micro-genetic algorithm and the in-situ surface observations and remotely-sensed satellite data. Snow data from surface observation stations representing five land cover types — deciduous broadleaf forest, mixed forest, woody savanna, cropland, and urban and built-up lands — are used to optimize five snow-related parameters that calculate the snow cover fraction, maximum snow albedo of fresh snow, and the fresh snow density associated with the snow depth. Another parameter, reflecting the dependence of snow cover fraction on the land cover types, is also optimized. Optimization of these six snow-related parameters has led to improvement in the root-mean squared errors by 17.0 %, 6.2 %, and 3.3 % on snow depth, snow albedo, and snow cover fraction, respectively. In terms of the mean bias, the underestimation problems of snow depth and overestimation problems of snow albedo have been alleviated through optimization of parameters calculating the fresh snow by about 44.2 % and 31.0 %, respectively.

1 Introduction

Land surface models (LSMs) act as the lower boundary conditions for regional numerical weather prediction (NWP) and climate models, to which they provide the surface fluxes (Ek et al., 2003). However, LSMs include inevitable uncertainties due to insufficient knowledge of surface layer processes and characteristics; for instance, unreasonable representation of the spatio-temporal surface heterogeneity and the inaccuracy of the parameters based on empirical relations contribute to the uncertainties

in LSMs. In particular, uncertainties in the snow-related processes of LSMs are appreciable and exert significant impacts on the performance of regional climate models to which the LSMs are coupled (e.g., Zhao and Li, 2015; Suzuki and Zupanski, 2018; Günther et al., 2019; Kim and Park, 2019; Xu et al., 2019; Jiang et al., 2020).

Intense snowfall events often occur in the Korean Peninsular during winter and early spring. In South Korea (SK), heavy snowfalls are the third most serious source of natural disasters, following typhoons and heavy rainfalls (Kim et al., 2018) with severe economic consequences. Most of the previous studies focused on classification of snowfall (Cheong et al., 2006 (In Korean with English abstract)), investigation of synoptic characteristics (Jung et al., 2012), and comparisons of different LSM options in the coupled atmosphere-land surface prediction system (Wang and Sun, 2018; Kim and Park, 2019). Being coupled to the atmospheric models, the LSMs play an important role to predict the snowfall in NWP because they calculate the fractional snow cover, snow albedo, and snow depth through interactions with the atmosphere. For example, the choice of land surface scheme is crucial to simulate the spatial distributions of snowfall in the land surface-coupled NWP models (e.g., Wang and Sun, 2018; Kim and Park, 2019). In other words, the numerical snowfall forecast is strongly affected by the performance of the coupled LSM; thus, improvement in the snow-process parameterizations of the off-line LSMs can bring about better performance in NWP models.

Uncertainties in parameterized physical processes have been observed and quantified in various numerical models (e.g., Mallet and Sportisse, 2006; Gubler et al., 2012; Shutts and Pallarès, 2014; Folberth et al., 2019; Li et al., 2020; Olafsson and Bao, 2020; Pathak et al., 2020; Souza et al., 2020). Such uncertainties can be reduced by estimating optimal parameter values in the subgrid-scale parameterization schemes (e.g., Annan and Hargreaves, 2004; Lee et al., 2006; Neelin et al., 2010; Yu et al., 2013; Zhang et al., 2015; Kotsuki et al., 2018; Li et al., 2018; Chinta and Balaji, 2020). Here, the parameter is a constant that makes up the equations, which is usually fixed during the simulation and differs from the variable representing the time-varying state of the model. Because empirical parameters are commonly derived from the observations or theoretical calculations, their estimated values are strongly dependent on the local characteristics of the region and period where the observations are made. Thus, *parameter estimation* that fits the model outputs to the observations is essentially required to obtain an adequate parameter (Duan et al., 2017). It may be done using a *trial and error* approach by manual, but the *optimization algorithm* helps to replace enormous experiments by automatically minimizing the difference between model and observations (Duan et al., 2006). For example, a global optimization tool, called the micro-genetic algorithm (micro-GA), has been effectively used for estimating the optimal parameter values in NWP model (e.g., Yu et al., 2013).

Most snow processes in the LSMs are parameterized based on the observations in specific local regions, and hence they may not represent adequately the situation in SK and be the source of uncertainties for numerical snow prediction over SK. We aim at obtaining the optimal parameter values of the snow-related processes — snow cover, snow albedo, and snow depth — in a LSM using the micro-GA, which causes a better LSM performance over SK. This study represents the first attempt to develop a coupled system of micro-GA and Noah LSM for parameter estimation of the snow processes. Section 2 describes the methodology, including the snow processes of the LSM and the micro-GA optimization tool. Section 3 explains experiment design. Results, discussion and conclusions are provided in sections 4, 5 and 6, respectively.

2 Methodology

2.1 Snow-related processes in Noah Land Surface Model

In this study, we employ the Noah Land Surface Model (Noah LSM; Chen et al., 1996; Koren et al., 1999; Ek et al., 2003) to simulate the single-site land surface processes (Mitchell, 2005), including the surface energy and water flux, and to verify energy and water budgets in the near-surface atmospheric layer by simulating the soil moisture and temperature, and the snowpack. Noah LSM is a stand-alone and one-dimensional column model, developed through multi-institutional cooperation. In the soil, to simulate soil moisture and soil temperature, we selected four layers with depths of 10, 30, 60, and 100 cm, respectively, from top to bottom, for a total depth of 2 m. The model also evaluates various other variables, including skin temperature, snow depth, snow water equivalent, snow density, canopy water content, etc. (Mitchell, 2005). The energy and water fluxes are calculated through the surface energy and water balance equations, respectively. Due to its adequate complexity and computational efficiency (Mitchell et al., 2004), the Noah LSM has been coupled to the operational NWP model of the Korea Meteorological Administration (KMA), named the Korean Integrated Model (KIM; Hong et al., 2018) — see Koo et al. (2017) for the details of the coupled KIM-Noah LSM system.

The current Noah LSM (version 3.4.1) uses a single-layer representation to the snow processes considering a bulk snow-soil canopy layer (Sultana et al., 2014). If air temperature is less than 0 °C, the resulting precipitation is considered snow. The fractional snow cover is determined as a function of snow water equivalent (SWE) using a generalized snow depletion curve. Snow albedo is calculated based on the fractional snow cover and the maximum snow albedo (Ek et al., 2003). Snow depth is represented by SWE and the bulk snow density (Jonas et al., 2009). The equations in Noah LSM describe the heat exchanges at the snow-atmosphere and snow-soil interfaces as well as snow accumulation, sublimation, and melting (Suzuki and Zupanski, 2018). In the followings, we describe the details of the snow parameters to be optimized.

2.1.1 Fractional snow cover (FSC)

The FSC (σ_s) is important for the accumulation and ablation processes (Livneh et al., 2010). As a function of SWE (W_s) extracted by the atmospheric input values (Livneh et al., 2010), σ_s varies nonlinearly as in Eq. (1), following the empirical snow depletion curves of Anderson (1973):

$$\sigma_s = 1 - e^{-P_s W} + W e^{-P_s}. \tag{1}$$

Here, P_s is the distribution shape parameter and $W = W_s/W_{max}$, where W_{max} is the threshold of W_s above which σ_s is 100%. Note that, from Eq. (1), σ_s is a function of P_s and W_{max} — these two parameters are to be optimized.

It is noteworthy that P_s has a positive correlation with snow cover (Fig. 1(a)). For example, σ_s increases as P_s increases, resulting in relatively slow snow melting. In Eq. (1), the value of P_s usually ranges between 2 and 4 (e.g., Anderson, 1973; Koren et al., 1999), and its default value in Noah LSM is 2.6. We seek the optimal value of P_s, which lies between 2 and 4 and is suited to SK.

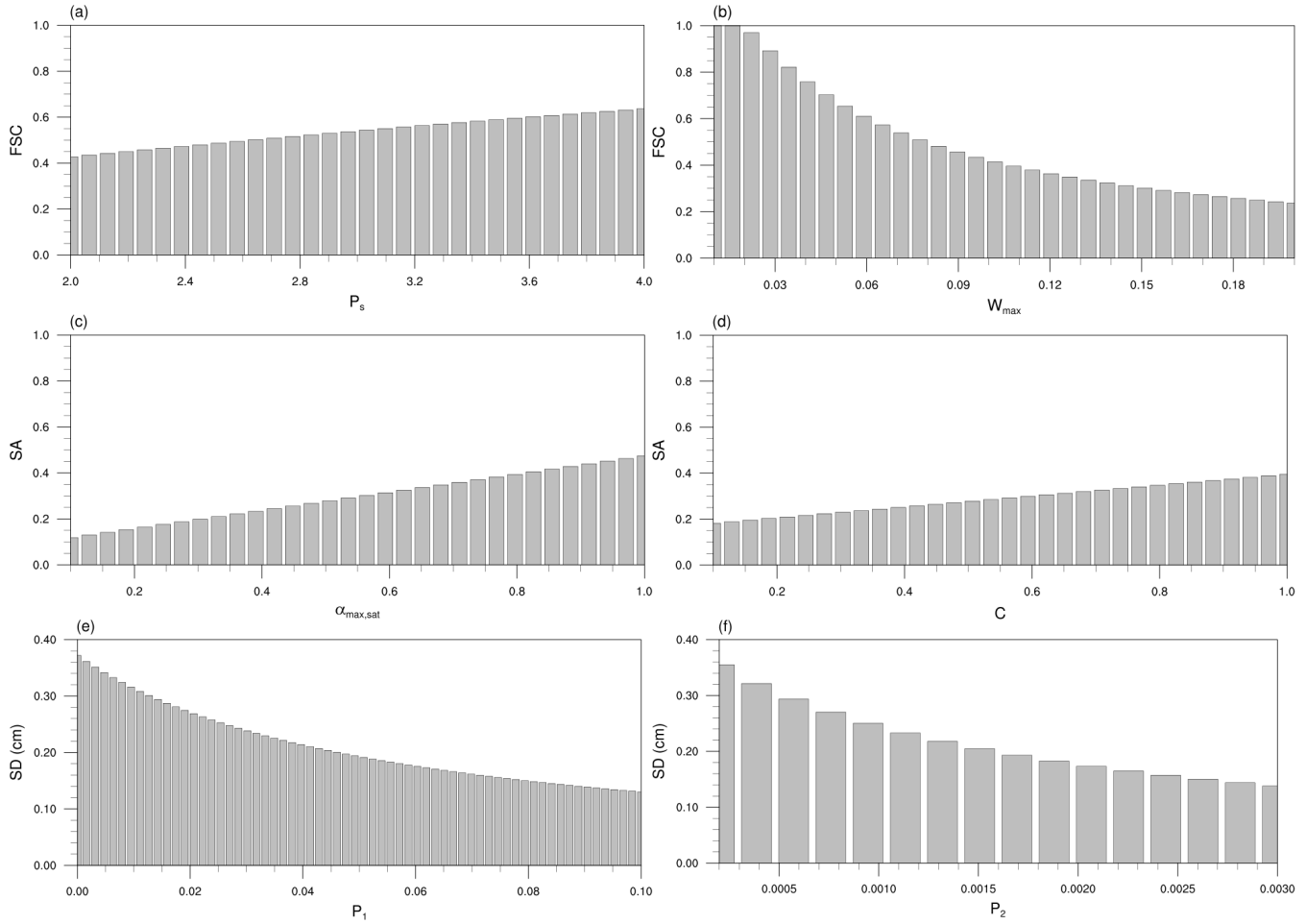

Figure 1. Responses of the snow variables to the variations in the snow-related parameters for given ranges: (a, b) Responses of FSC, for W_s = 0.02, to variations in P_s (with W_{max} = 0.08) and in W_{max} (with P_s = 2.6), respectively; (c, d) Responses of SA, for $\alpha_{max,sat}$ = 0.2 and t = 10 days, to variations in $\alpha_{max,CofE}$ (with C = 0.5) and in C (with α_{cofE} = 0.85), respectively; and (e, f) Responses of SD (in cm), for W_s = 0.02 and T_{air} = -5 °C), to variations in P_1 (with P_2 = 0.0017) and in P_2 (with P_1 = 0.05 g cm^{-3}).

The SWE threshold, W_{max}, has a negative correlation with snow cover, as shown in Eq. (1) and it is more sensitive compared to P_s within a given parameter's range (Fig. 1(b)). In Noah LSM, the values of W_{max} are prespecified in a table (VEGPARM.TBL), varying with the land cover types (LCTs). W_{max} has the largest value over forest, reflecting the irregular geometry of forest cover (Wang et al., 2010). Previous studies suggest the uncertainty range in the values of W_{max}; for instance, Livneh et al. (2010) used 0.04 m for forest and 0.02 m for non-forest, respectively, whereas Wang et al. (2010) used 0.2 m for tall vegetation and 0.01 m for short vegetation. The default values in Noah LSM are 0.08 m for forest and 0.04 m for non-forest. We estimate the optimal W_{max} values, suited to SK, in the range between 0.01 m and 2 m.

2.1.2 Snow albedo (SA)

SA is defined as the fraction of incident radiation reflected by the snowpack and is crucial for evaluating surface-energy balance, particularly during snow melting (Warren and Wiscombe, 1980; Warren, 1982); however, accurate representation of SA is difficult due to numerous complexities (Livneh et al., 2010).

Surface albedo generally increases over snow, but it may react differently over a shallow snowpack: when accumulation starts by snowfall or diminution occurs by snow melt, patchy areas can be generated and corresponding model grid boxes may not be covered by snow (Ek et al., 2003). The Noah-LSM reflects this patchiness effect by calculating surface albedo (α) as a composite of snow-covered surface albedo (α_s) and snow-free surface albedo (α_0) as

$$\alpha = \alpha_0 + \sigma_s(\alpha_s - \alpha_0). \tag{2}$$

Note that SA is generally highest over the fresh snow and decays thereafter, and the decay rate depends on the seasonal snow phase — faster during the ablation phase and slower during the accumulation phase. By reflecting this fact, α_s is evaluated as a function of the fresh SA (α_{max}), the number of days after the last snowfall (t), and the albedo-decay rates (A and B) as

$$\alpha_s = \alpha_{max} A^{t^B}, \tag{3}$$

where the default values of empirical parameters A and B are 0.94 and 0.58, respectively, during the accumulation phase and 0.82 and 0.46, respectively, during the ablation. However, the current Noah LSM activates only the accumulation phase in Eq. (3), and both A and B are excluded from our optimization.

Spatial variation in SA is taken into consideration in α_{max}, by incorporating the satellite-based maximum SA ($\alpha_{max,sat}$) from Robinson and Kukla (1985) and by imposing adjustment to a maximum SA ($\alpha_{max,CofE}$) from USACE (1956) (see also Livneh et al., 2010), as

$$\alpha_{max} = \alpha_{max,sat} + C(\alpha_{max,CofE} - \alpha_{max,sat}), \tag{4}$$

where C is a proportionality coefficient. We optimize two empirical parameters that show positive relation to SA — $\alpha_{max,CofE}$ and C, whose default values are 0.85 and 0.5, respectively (Fig. 1(c)-(d)): SA shows similar sensitivities to both parameters within the same range but is a bit more sensitive $\alpha_{max,CofE}$. Some other values have been used in previous studies (e.g., Livneh et al., 2010), such as 0.6 to 0.95 for $\alpha_{max,CofE}$ and 1.0 for C. For the parameter estimation in this study, we set the ranges from 0.1 to 1.0 for both parameters.

2.1.3 Snow depth (SD)

In Noah LSM, SD is evaluated as the ratio of SWE (W_s) to snow density (μ_s), i.e., W_s/μ_s (Gotleib, 1980; Koren et al., 1999). While SWE is determined by precipitation in the model, snow density is determined by several other parameters such as the compression and melting of snow (Koren et al., 1999). Fresh snow density ($\mu_{s,fresh}$) depends on air temperature (T_{air}), i.e., 2 m temperature (Gotleib, 1980) as

$$\mu_{s,fresh} = P_1 + P_2(T_{air} + 15)^{1.5}, \tag{5}$$

where $P_1 = 0.05$ g cm^{-3} and $P_2 = 0.0017$ are the default values of the coefficients. If T_{air} is less than -15 °C, $\mu_{s,fresh}$ is set to 0.05 g cm^{-3}; otherwise, $\mu_{s,fresh}$ tends to increase as T_{air} increases. As the empirical parameters P_1 and P_2 are directly associated with $\mu_{s,fresh}$, we seek optimal values of these parameters. Because snow density is inversely proportional to SD, both P_1 and P_2 have negative correlations with the SD (Fig. 1(e)-(f)), where SD shows similar sensitivities to both parameters.

2.2 Optimization tool: micro-genetic algorithm

The genetic algorithm (GA) is a global optimization algorithm developed by John Holland in the 1970s (e.g., Holland, 1973, 1975) and is based on Darwinian principles of natural selection (Golberg, 1989). It uses reproduction selection, crossover and mutation to operate a set of potential solutions, i.e., *population* or *individual*, which are expressed by a string, called a *chromosome*: its binary form is called a *gene* (Koren et al., 1999; Rudnaya and Santosa, 2000). The *reproduction* operator first selects good solutions or eliminates bad solutions based on the fitness value; then, the *crossover* operator exchanges the genetic information between the solutions using the single-point or uniform types. The *mutation* operator modifies the value of each gene of the chromosomes by replacing it with the opposite value, e.g., 0 with 1, which prevents premature convergence. When a new generation is created, the above processes are repeated until the convergence condition or the prescribed number of iterations is satisfied.

Micro-GA is an advanced and simplified GA with smaller generation sizes, thus requiring less computational time than the conventional GA (Krishnakumar, 1990; Wang et al., 2010). It has been used in meteorology for optimal parameter estimation (e.g., Yu et al., 2013) or scheme-based optimization (e.g., Hong et al., 2014, 2015; Park and Park, 2021; Yoon et al., 2021). Its main difference from the conventional GA is the population size; for example, micro-GA uses 5 individuals while the conventional GA uses more than 30 individuals. Note that the conventional GA with a small population quickly converges to non-optimal solutions due to insufficient information; however, micro-GA solves this problem by using *elitism*, which assigns the best individual among the 5 individuals based on the fitness evaluation and carries it to the next generation — this guarantees to preserve the good solutions during the generations. Furthermore, micro-GA does not take mutation to achieve diversity; instead, it uses the *re-initialization* which starts with a new individual whenever the diversity is lost.

2.2.1 Coupling micro-GA with Noah LSM and parallelization

Figure 2 describes the process of parameter optimization in the micro-GA-Noah LSM coupled system: 1) Micro-GA initializes the snow parameter combinations represented by the binary encoding through the random samples of the individual; 2) Micro-GA controls Noah LSM by editing the parameter-related files, such as GENPARM.TBL, VEGPARM.TBL, and the Fortran code (module_sf_noahlsm.F) and prepares the forcing data for each station; 3) As recommended in Carroll (1996), the 5 individuals configured with the different snow parameters execute the ensemble runs of Noah LSM in parallel; 4) The performance of each Noah LSM is evaluated in comparison with the observation through a given fitness function; 5) Micro-GA selects the highest fitness comparing a number of individuals through the tournament selection; 6) New combinations for the next generation are produced through the crossover using the selected ones in the previous step; 7) When the convergence

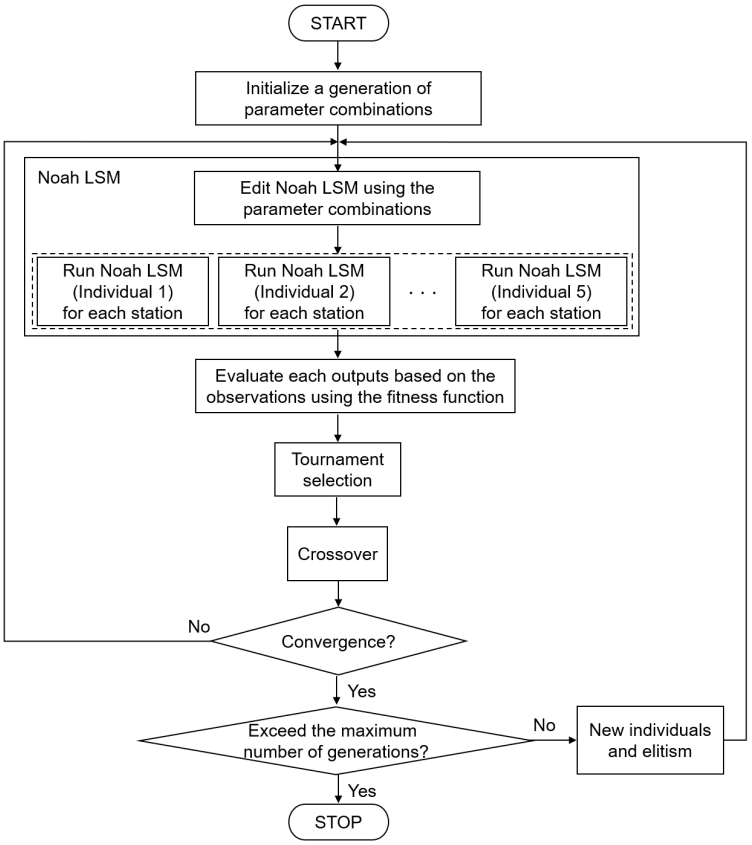

Figure 2. A flow chart of parameter optimization from the micro-GA-Noah LSM coupled system. The dashed box depicts the parallel system for Noah LSM, running for each individual.

is satisfied, the other 4 individuals except the best individual marked by elitism are randomly regenerated; and 8) Micro-GA repeats these processes until the prescribed-entire iteration converges into a global maximum of the fitness function.

Although micro-GA is computationally more efficient than the conventional GA, it still demands substantial computing time because each individual serially executes the model. Therefore, we have developed a parallel processing system in the micro-GA-Noah LSM coupled system. Instead of sequentially performing each individual and calculating the fitness within a generation, we run the model simultaneously for all populations to obtain the fitness and select the best individual when all tasks are finished (see the dashed box in Fig. 2). This new parallel system linearly reduces the execution time, which is proportional to the number of individuals. In addition, since the coupling system was created in a shell script, it is possible to assign multiple cores for model execution for various stations. The new parallel processing system, created by reflecting these two main points, improves the computation time — making it different from the non-parallel processing of a coupled system, e.g., the micro-GA–Noah-MP system (see Hong et al., 2014).

2.2.2 Fitness function

The fitness function is a performance index to evaluate how well potential solutions fit the objective. In the GA optimization, the fitness function should be carefully defined because it is used for all generations and individuals. Generally, the root-mean-square error (RMSE) is a widely used indicator for evaluating the performance of a model (e.g., Yan et al., 2019). Since our aim is to improve the snowfall prediction, we simultaneously evaluate all related snow variables — FSC, SA, and SD. We have first calculated the RMSE for each snow variable as

$$RMSE(\mathbf{x}) = \sqrt{\frac{\sum_{i=1}^{N} (\hat{\mathbf{x}}_i - \mathbf{x}_i)^2}{N}}, \tag{6}$$

where $\mathbf{x}$ is a vector representing the three snow variables and N is the total number of observation time. Here, $\hat{\mathbf{x}}$ is the predicted values in the Noah LSM while $\mathbf{x}$ is the observed values. The number of observations is dependent on the observational types: the Automated Synoptic Observing System (ASOS) produces hourly data for SD while the MODerate resolution Imaging Spectroradiometer (MODIS), a sensor onboard the polar orbiting satellite Terra, produces daily data for FSC and SA. To calculate the RMSE between the model solutions and observations, the Noah LSM simulations are made over the observation locations. For SD, the RMSE is directly obtained on the same grid point. As the MODIS data have a coarser resolution, we use the observation point nearest the ASOS location (see the details in section 2.3).

We have then obtained the improvement ratio, $r(\mathbf{x})$, by comparing the RMSEs from the model runs with non-optimized parameters (say, CNTL) and optimized parameters (say, OPTM), respectively, as

$$r(\mathbf{x}) = \frac{RMSE(\mathbf{x})_{CNTL} - RMSE(\mathbf{x})_{OPTM}}{RMSE(\mathbf{x})_{CNTL}}. \tag{7}$$

Lastly, we have averaged all the improvement ratios for the snow variables to define the fitness function, $f(\mathbf{x})$, as

$$f(\mathbf{x}) = \sum_{j=1}^{M} \frac{r(\mathbf{x})_j q_j}{M} \tag{8}$$

where M is the number of stations and q is a quality control flag (QCF) — either 0 or 1. The QCF is employed to secure a sufficient number of snow observations. It is set to 0 (i.e., the fitness function is not accumulated) for the following cases: 1) snow events are not simulated after optimization; and 2) the number of snow observations is less than 2. Furthermore, when the performance gets deteriorated after optimization, we give a penalty by doubling Eq. (7) to prevent degradation of the optimization.

We finally define the normalized fitness function, $f_n(\mathbf{x})$, as

$$f_n(\mathbf{x}) = \frac{f(FSC) + f(SA) + f(SD)}{3}, \tag{9}$$

whose values lie in the range $[-1, 1]$. Thus, the micro-GA finds the maximum fitness based on Eq. (9).

2.3 Data

The land surface processes were forced by six meteorological fields from ASOS (https://data.kma.go.kr): wind speed (m s^{-1}), wind direction (degrees), temperature (K), relative humidity (%), surface pressure (hPa), and precipitation rate (kg $m^{-2}s^{-1}$). When missing data exist in less than 72 hours, linear interpolation was performed except for precipitation. Stations with the missing rate greater than 1 %, during the entire experimental period, have been excluded. For the initial and boundary conditions, downward shortwave/longwave radiation (W m^{-2}), precipitation rate (kg $m^{-2}s^{-1}$), soil temperature (K), soil moisture (m^3 m^{-3}), and surface temperature (K) have been obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) — the fifth generation ECMWF reanalysis-Land (ERA5L) hourly data (Muñoz-Sabater, 2019) — having a spatial resolution of 9 km and four soil layers with depths of 7, 21, 72, and 189 cm, respectively, from top to bottom for a total depth of 2.89 m. We have used the data at the ERA5L grid nearest point to the ASOS station.

The snow observations (i.e., SD, FSC, and SA) are used for the model verification and the fitness function calculation. For SD, the hourly model outputs are evaluated using the hourly ASOS data. To confirm the snow season, we have excluded the SD observations lower than 0.1 cm. For FSC and SA, we have no ASOS observations over SK; thus, we have used the MODIS/Terra Snow Cover Daily L3 Global 500 m SIN Grid radiance data (Hall and Riggs, 2016). They are generated from the MODIS/Terra Snow Cover 5-Min L2 Swath 500 m data (Hall et al., 2006) by selecting the best observation based on a scoring algorithm when they are closest to nadir with maximum coverage of the cell (Hall and Riggs, 2007). In particular, FSC is generated by the Normalized Difference Snow Index (NDSI). The MODIS snow data at the points nearest to the ASOS locations were extracted and used for verification of the model-generated FSC and SA. Being a polar orbiting satellite, MODIS contains only one observation per day; thus, we have extracted the model output for verification at 02 UTC when the satellite (Terra) passes over SK. For the calculations, we have converted the percent values of FSC and SA to the decimal values; then, we have excluded observational data with values below 0.05 (i.e., 5%) for both FSC and SA.

For the optimization experiment, we have selected some stations that represent different land covers in SK, aiming at having a representative combination of snow-related parameters over SK. We have defined a representative set of LCTs within a 2.5 km radius from the ASOS observations, excluding the water body. The LCTs have been taken from the MODIS (onboard Terra and Aqua) Land Cover Type Yearly Climate Modeling Grid (CMG) Version 6 (Friedl and Sulla-Menashe, 2015), in which maps are provided from the land cover classification schemes of the International Geosphere-Biosphere Programme (IGBP), the University of Maryland (UMD) and the Leaf Area Index (LAI), all at a 0.05 degree spatial resolution in geographic latitude/longitude projection (see Sulla-Menashe and Friedl, 2018), for the entire globe from 2001 to 2019. Finally, we have compiled a set of five representative stations for each different LCT — deciduous broadleaf forest (DBF), mixed forest (MF), woody savanna (WS), cropland (CL), and urban and built-up lands (UB) — as shown in Table 1.

3 Experimental design

We have designed the following two GA optimization experiments: 1) OPT_5 that optimizes five snow parameters (P_s, $\alpha_{max,CofE}$, C, P_1, and P_2); and 2) OPT_W that optimizes W_{max}. Among the six parameters, W_{max} is the only parame-

Table 1. Five representative LCTs over SK, following the IGBP classification — DBF, MF, WS, CL, and UB. For each LCT, five selected stations are shown with the station name (abbreviation in parenthesis), location in latitude (°N) and longitude (°E), ratio of LCT in 2.5 km buffer (%), soil type, and missing ratio (%). The experiment OPT_5 employs only the stations highlighted in bold while the other experiments use all the stations.

IGBP LCT	Station Name	Latitude	Longitude	Ratio of LCT in 2.5 km Buffer	Soil Type	Missing Ratio
DBF	**Ulleungdo (UL)**	**37.481**	**130.899**	**82.7**	**Silt Loam**	**0.15**
	Taebaek (TB)	37.170	128.989	67.0	Loam	0.15
	Inje (IJ)	38.060	128.167	62.7	Sandy Loam	0.07
	Chupungnyeong (CP)	36.220	127.995	56.8	Silt Loam	0.04
	Youngwol (YW)	**37.181**	**128.457**	**42.6**	**Clay**	**0.09**
MF	Bongwha (BW)	36.944	128.914	38.7	Loam	0.11
	Hapcheon (HP)	35.565	128.170	32.1	Loam	0.51
	Hongcheon (HC)	37.683	127.880	26.3	Silty Clay Loam	0.05
	Miryang (MY)	**35.491**	**128.744**	**22.5**	**Sandy Loam**	**0.16**
	Gumi (GM)	**36.131**	**128.321**	**24.1**	**Sandy Loam**	**0.05**
WS	Imsil (IS)	35.612	127.286	53.1	Sandy Loam	0.12
	Andong (AD)	36.573	128.707	43.9	Loamy Sand	0.04
	Boeun (BE)	36.488	127.734	41.2	Sandy Loam	0.07
	Uljin(UJ)	**36.992**	**129.413**	**39.2**	**Loam**	**0.19**
	Bukgangneong (NG)	**37.805**	**128.855**	**37.5**	**Sandy Loam**	**0.04**
CL	Buan(BA)	35.730	126.717	87.8	Loam	0.03
	Icheon(IN)	37.264	127.484	74.6	Sandy Loam	0.16
	Haenam(HN)	34.554	126.569	63.7	Sandy Loam	0.29
	Boryeong (BR)	**36.327**	**126.557**	**53.8**	**Silty Clay Loam**	**0.14**
	Jeongeup (JE)	**35.563**	**126.839**	**51.7**	**Silt Loam**	**0.28**
UB	Gwangju(GJ)	35.173	126.892	94.6	Loam	0.03
	Seoul (SL)	**37.571**	**126.966**	**90.8**	**Loam**	**0.08**
	Daejeon (DJ)	36.372	127.372	72.2	Sandy Loam	0.03
	Suwon(SW)	**37.257**	**126.983**	**71.4**	**Sandy Loam**	**0.10**
	Incheon (IC)	37.478	126.625	70.1	Loam	0.07

ter that depends on the LCTs; thus, we conducted OPT_5 and OPT_W separately. Note that SK is represented by five different LCTs considering the sufficient days of snowfall and ASOS observation (see Table 1). Because OPT_5 optimizes with more

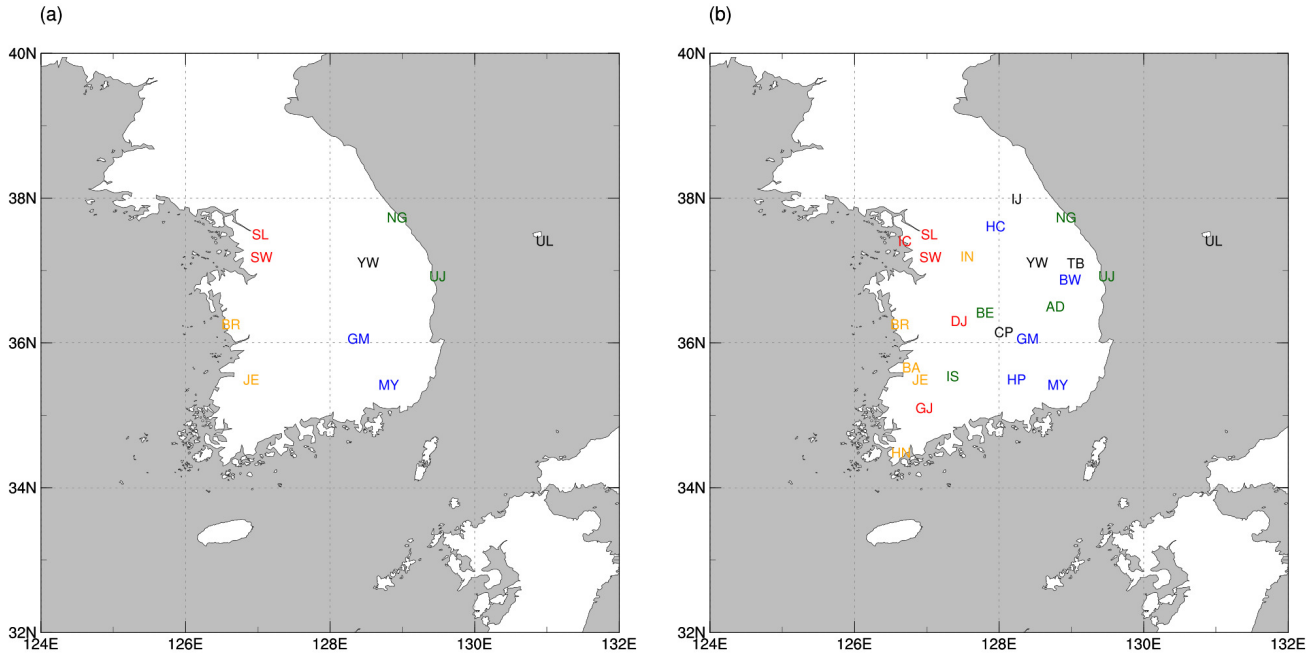

Figure 3. Stations used for the experiments (a) OPT_5 and (b) OPT_W, CNTL, VRF_5 and VRF_6. Different colors in the station acronyms represent different LCT: DBF (black), MF (blue), WS (green), CL (yellow), and UB (red). See Table 1 for the acronyms of stations and LCTs.

parameters and generations, we have selected 10 stations (i.e., 2 stations per LCT) based on snowfall amount to reduce the computation time. To investigate the performance of snow prediction through optimized snow parameters, we have designed the following three verification experiments for the 25 observation stations: 1) CNTL using non-optimized (i.e., default) parameters; 2) VRF_5 using the five optimized parameters obtained from OPT_5; and 3) VRF_6 using the six optimized parameters obtained from both OPT_5 and OPT_W (see Fig. 3(b)).

For the micro-GA optimization, we have pre-specified the following input parameters: 1) the population size, i.e., a collection of individuals; 2) the number of parameters to be used for optimization; 3) the number of chromosomes expressing an arbitrary solution; 4) the maximum number of generations to iterate the optimization; 5) the type of crossover operator that creates a new structure of chromosomes through the exchange of the chromosome; and 6) the elitism to decide whether the most suitable individual would be preserved for next generation. The micro-GA-Noah LSM coupled system has been repeatedly performed to find a parameter combination within the specified generations. We utilized the uniform crossover in which each gene is selected randomly from one of the parent chromosomes.

Table 2 describes the input parameters for micro-GA used in this study. We follow the options known as the best performance in micro-GA; it is done with a population size of 5 and a uniform crossover (i.e., crossover operator = 1.0) with elitism (Carroll, 1996; Yu et al., 2013; Yoon et al., 2021). The uniform crossover makes all populations perform a crossover at every generation

Table 2. The input parameters for micro-GA in experiments OPT_5 and OPT_W.

Input Parameter	OPT_5	OPT_W
Population size	5	5
Crossover operator	1.0	1.0
Elitism	on	on
Number of parameters	5	1
Number of chromosomes	30	5
Maximum value of generations	200	100

to acquire the diversity (Lee et al., 2005). The number of parameters to be optimized is 5 for OPT_5 and 1 for OPT_W. The number of chromosomes determines the number of cases expressed in a binary format. For example, the selected parameters — P_s, $\alpha_{max,CofE}$, C, P_1, P_2, and W_{max} — use different chromosomes, i.e., 5, 5, 5, 6, 4, and 5, respectively; thus, the total number of chromosomes is 30 for OPT_5 and 5 for OPT_W. The maximum value of generations at the end of optimization is generally set to 100 (Yu et al., 2013; Yoon et al., 2021; Zhu et al., 2019), whereas we increased generations up to 200 in OPT_5 due to larger number of parameters to be optimized.

In this study, we have conducted the optimization experiments from 0000 UTC 1 May 2009 to 2300 UTC 30 April 2018. During this 9 years period, the number of snow observations was continuously secured. Data from the first 5 months (May–Oct in 2009) were utilized for model initialization and spin-up, thus they were not considered for the verification. Cross validation has been conducted using the 1 year data from 0000 UTC 1 May 2018 to 2300 UTC 30 April 2019. Since they showed similar aspects, we only discuss the results of optimization periods having sufficient samples.

4 Results

4.1 Spin-up analysis

Numerical prediction models generally require spin-up to reach a statistical equilibrium state where the initial conditions under a forcing are adjusted to the model's own physics/dynamics and numerics (Bonekamp et al., 2018). Without sufficient spin-up, the LSMs can generate severe bias of initial conditions (Cosgrove et al., 2003). Prior to the optimization experiments, we have conducted a spin-up experiment in one of the stations, Seoul, to check the appropriate spin-up time. It was carried out in two ways: 1) using a spin-up period recursive in 9 years (e.g., Jun et al., 2020); and 2) using a spin-up period that was not included in the analysis.

First, the Noah LSM has been repeatedly executed using the atmospheric forcing for 9 years. This recursive simulation has been conducted from 1 May 2009 to 30 April 2018 to see whether the model was able to reach an equilibrium by setting the repetition loop as 0, 300, 600, and 1000. Our results indicated no significant differences; thus, we concluded that repetition was

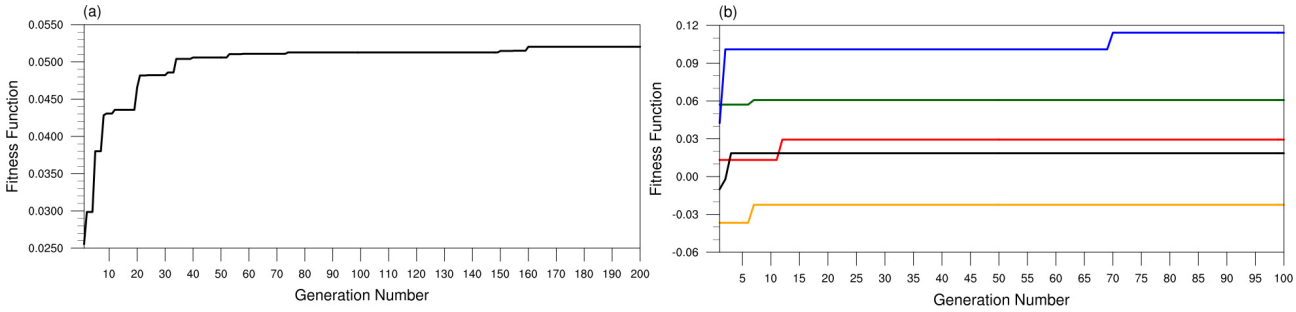

Figure 4. The fitness function for generations during the optimization of (a) five snow parameters optimization (OPT_5), and (b) W_{max} optimization (OPT_W) for DBF (black), MF (blue), WS (green), CL (yellow), and UB (red) LCTs.

not required. Second, we have performed sensitivity tests to identify the spin-up period due to changes in the initial conditions by adding biases (± 0.1 m^3 m^{-3} for soil moisture and ± 3 K for soil temperature) to the ERA5L data. As a result, we found that the adequate spin-up periods were about 3 months and 1 year for soil moisture and soil temperature, respectively; however, the snow variables were insensitive to the initial condition changes, thus requiring no spin-up period. Although the spin-up is not necessary for this study that focuses on the snow processes, we have performed the optimization experiments starting from May when snow is absent.

4.2 Optimal estimation of snow parameters

To optimize snow parameters specialized in SK, we have employed the micro-GA-Noah LSM coupled system using the observations over SK. Figure 4(a) shows the evolution of the fitness function for OPT_5 in a total of 200 generations, as well as Fig. 4(b) for OPT_W in a total of 100 generations. Since the OPT_W optimizes solely W_{max} parameter, it has smaller generations. In OPT_5, the fitness function converges at 160^{th} generation, while the fitness function of OPT_W quickly converges in all LCTs (Fig. 4(b)). The convergence occurs at 3^{rd} generation for DBF, 70^{th} generation for MF, 7^{th} generation for both WS and CL, and 12^{th} generation for UB.

As a result, we have obtained the optimized six snow parameters over SK (Table 3). The OPT_5 simultaneously generates the optimized five snow parameters (P_s, $\alpha_{max,CofE}$, C, P_1, and P_2) associated with the FSC, SA, and SD while the OPT_W, depending on the LCTs, generates the optimized W_{max} associated with the FSC. The first snow parameter, P_s, is optimized from its standard value of 2.6 to 2.7097, which results in an increase of the FSC. The second snow parameter, W_{max}, is optimized depending on each LCT. In detail, the W_{max} in DBF and WS increases from 0.08 to 0.1632 and from 0.03 to 0.0406, respectively. They lead to a decrease of the FSC due to a negative correlation. On the other hand, the W_{max} in MF and UB decreases from 0.08 to 0.0529 and from 0.04 to 0.0284, respectively, thus increasing the FSC. The optimized CL shows a similar value from 0.04 to 0.0406, which means that the current value was proper to SK. The third snow parameter related to the SA, $\alpha_{max,CofE}$, decreases from 0.85 to 0.7387, inducing a decrease of SA. The fourth snow parameter, C, also shows a

Table 3. Summary of optimized snow parameters related to snow variables. Minimum (Min), Default, Maximum (Max) are the ranges used in the optimization process. Default is the empirical value used in the Noah LSM.

Snow Variable	Snow Parameter	LCTs	Min/Default/Max	Optimized Value
FSC	P_s	–	2.0/2.6/4.0	2.7097
	W_{max}	DBF	0.01/0.08/2.00	0.1632
		MF	0.01/0.08/2.00	0.0529
		WS	0.01/0.03/2.00	0.0406
		CL	0.01/0.04/2.00	0.0406
		UB	0.01/0.04/2.00	0.0284
SA	$\alpha_{max,CofE}$	–	0.10/0.85/1.00	0.7387
	C	–	0.1/0.5/1.0	0.5355
SD	P_1	–	0.00/0.05/0.10	0.0698
	P_2	–	0.0002/0.0017/0.003	0.0002

similar value from 0.5 to 0.5355, thus this value was proper to SK. The fifth snow parameter, P_1, increases from 0.05 to 0.0698, resulting in a decrease of SD. The last snow parameter, P_2, reduces from 0.0017 to 0.0002, leading to an increase of SD.

We have investigated the mean bias (MB) using the box plot expressing the quartile and the distribution of extreme values: it explains how much the bias of the CNTL is reduced in optimization experiments by comparing the model with the observations. Before optimization, the CNTL showed under-estimated FSC and SD and over-estimated SA (-0.133, -4.39 cm, and 0.0408, respectively; see Fig. 5). However, the bias patterns in FSC and SA vary on each station owing to the lower spatial and temporal resolution of satellite observation. On the other hand, the SD shows an under-estimation at all stations; the increase in the SD due to fresh snow was under-estimated, and snow melting was proceeding faster than the observation.

The performance has been evaluated using the improvement ratio, which indicates how much the RMSE, MB, and coefficient of determination (R^2) of experiments using optimized parameters (i.e., VRF_5 and VRF_6) is improved compared to CNTL, as shown in Eq. (7) (Table 4). In the VRF_5, new parameter values — P_s, $\alpha_{max,CofE}$, C, P_1, and P_2 — optimized by the micro-GA result in an improvement of RMSE for FSC, SA and SD, such as 0.7 %, 5.4 % and 13.7 %, respectively (Table 4). However, the RMSE of FSC relatively weakly improved by about 0.7 % because the other parameter, W_{max}, is not yet optimized. In terms of MB, we anticipate that the increase of P_s overcomes the under-estimated FSC. But the VRF_5 strengthens the under-estimation of FSC from -0.133 to -0.145, thus it deteriorates the MB by about 9.1 % (Table 4 and Fig. 5(a)). Regarding the SA, the optimized $\alpha_{max,CofE}$ decreases the SA to solve the over-estimation in CNTL. The other parameter C has optimized to its default value, 0.5355: which means that this was an appropriate constant for SK snowfall prediction. Therefore, the MB of SA is improved by 26.9 % by reducing the SA from 0.0408 to 0.0298 (Table 4 and Fig. 5(b)). Next, SD shows the greatest RMSE improvement of 13.7 % (Table 4). In fact, the Noah LSM suffers from a negative bias for SWE, especially in early spring

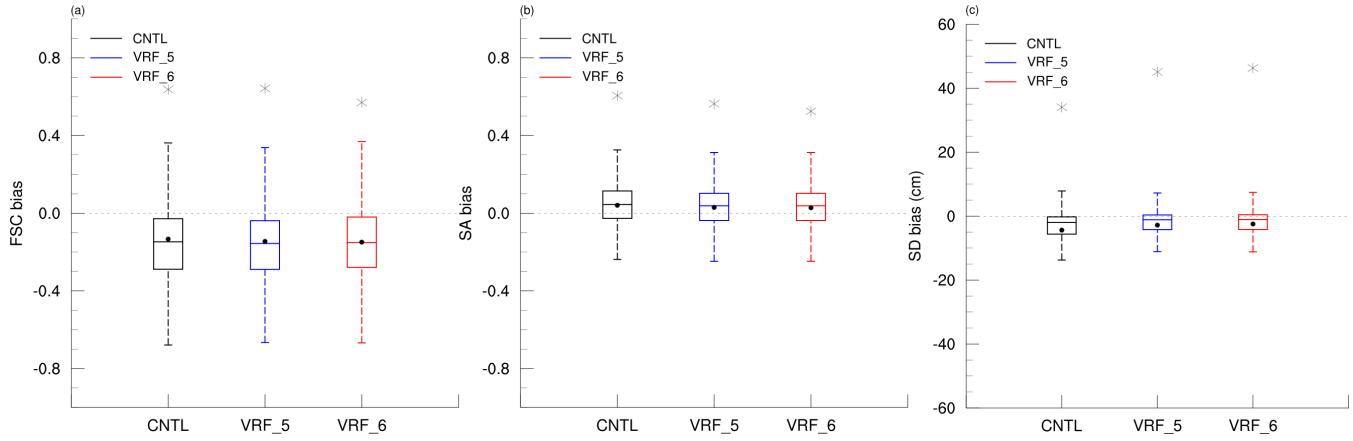

Figure 5. Box plots of (a) FSC bias, (b) SA bias, and (c) SD bias (cm) for CNTL, VRF_5 and VRF_6. The maximum differences are indicated with the black star symbol (e.g., 0.637 (CNTL), 0.643 (VRF_5), 0.570 (VRF_6) for FSC, 0.605 (CNTL), 0.563 (VRF_5), and 0.525 (VRF_6) for SA, and 34.1 cm (CNTL), 45.1 cm (VRF_5), and 46.3 cm (VRF_6) for SD). Each mean of snow variables is indicated as a black circle (e.g., -0.133 (CNTL), -0.145 (OPT_5), and -0.149 (VRF_6) for FSC, 0.0408 (CNTL), 0.0298 (VRF_5), and 0.0281 (VRF_6) for SA, and -4.39 cm (CNTL), -2.81 cm (VRF_5), and -2.45 cm (VRF_6) for SD).

(Sheffield et al., 2003; Ek et al., 2003; Pan et al., 2003; Mitchell et al., 2004; Jin and Miller, 2007; Livneh et al., 2010). Because SD is proportional to SWE, the under-estimation can be exhibited due to negative bias of SWE. However, the optimized P_1 leads to a decrease in SD, thus it intensifies the under-estimation for SD. On the other hand, the optimized P_2 increases the SD as follows: when the air temperature is warmer than the $-15\ ^{\circ}C$, the fresh snow density slowly increases, which quickly induces an increase of SD following Eq. (5). Therefore, the optimization of P_2 solves the under-estimated SD by about 35.9 % due to increased SD from -4.39 cm to -2.81 cm within most of the temperature ranges (Table 4 and Fig. 5(c)). We also investigated R^2, which measures the proportion of variation for a dependent variable that can be explained by an independent variable. Although the R^2 values are low in FSC and SA, the difference between CNTL and verification experiment (e.g., VRF_5) has 95% statistical significance, as evaluated with a two-tailed t-test. After optimization, the R^2 values in VRF_5 improve by 3.3 % and 1.5 % for FSC and SD, respectively. However, these changes are insignificant compared to the other statistics such as RMSE and MB.

To supplement insufficient improvement in the FSC, we have additionally optimized the W_{max} in function of LCT (OPT_W) based on the five parameters optimization results from OPT_5. Here, we have only used the FSC to define the fitness function, they not considering SA and SD. Therefore, the fitness function is defined using Eq. (8) where the $\boldsymbol{x}$ is only the FSC, so the normalized process with Eq. (9) is not needed. As a result, the OPT_W further improves the RMSE of FSC compared to previous optimization results in the DBF, MF, WS, and UB by 4.6 %, 11.9 %, 7.7 %, and 5.5 %, respectively, while weakly decreases by 0.1 % in CL. To solve the under-estimated FSC that occurred at all stations in VRF_5, we anticipate OPT_W decreases the W_{max}, which leads to an increase of FSC. Consequently, the OPT_W generates a decreased W_{max} in the MF and UB and other LCTs (e.g., DBF, WS and CL) generate increased W_{max}.

Table 4. The RMSE, MB, R^2 of snow variables and improvement ratio (%) in parentheses from CNTL to VRF_5, and VRF_6 over the 25 representative stations. The difference between CNTL and verification experiments (i.e., VRF_5 and VRF_6) has 95% statistical significance, as evaluated with a two-tailed t-test.

EXP	CNTL			VRF_5			VRF_6		
Snow Variable	FSC	SA	SD	FSC	SA	SD	FSC	SA	SD
RMSE	0.249	0.132	9.094	0.247 (0.7 %)	0.125 (5.4 %)	7.847 (13.7 %)	0.124 (3.3 %)	0.125 (6.2 %)	7.547 (17.0 %)
MB	-0.133	0.0408	-4.39	-0.145 (-9.1 %)	0.0298 (26.9 %)	-2.81 (35.9 %)	-0.149 (-11.9 %)	0.0281 (31.0 %)	-2.45 (44.2 %)
R^2	0.257	0.281	0.808	0.265 (3.3 %)	0.276 (-1.7 %)	0.821 (1.5 %)	0.277 (8.0 %)	0.274 (-2.2 %)	0.834 (3.0 %)

Finally, all six parameters related to the snow variables have been verified in VRF_6 having the same 25 stations used in the CNTL. When the optimized five parameters are used except the W_{max} (VRF_5), SA and SD are improved, and FSC shows a weakly improvement in RMSE performance. However, when the optimized W_{max} depending on the LCTs from the OPT_W is used (VRF_6), the FSC appears in a larger positive impact with other variables. As a result, an improvement of RMSE for the FSC, SA, and SD is 3.3, 6.2, and 17.0 %, respectively (Table 4). However, the MB for the FSC strengthens from 9.1 % to 11.9 % in VRF_6 (Table 4) due to larger negative bias especially in the DBF. On the other hand, SA and SD reduce the MB against the CNTL and enhance the improvement ratio from 26.9 % to 31.0 % and from 35.9 % to 44.2 %, respectively. Like the RMSE, the R^2 of FSC and SD also improved in VRF_5 and VRF_6. The SA worsened in VRF_5 was a bit more severe in VRF_6. However, they are still small impacts compared to RMSE and MB.

To understand more details of the improvements due to the optimization, we analyzed the scatter plots that compare the observations and the model results in Figure 6 and listed their RMSE and R^2 in Table 5. Since the observation patterns are different for different stations, we selected the representative station for each LCT. For FSC, it is relatively hard to recognize the explicit bias patterns, as shown in Fig. 6 (left panels); however, compared to CNTL, the RMSE decreased in VRF_5 and further reduced in VRF_6 (see Table 5). The VRF_6 revealed the largest R^2 values over most LCTs, except WS (station NG) and CL (station BR). In particular, VRF_6 produced the highest FSC over MF (station GM) (see Fig. 6(d)) with the smallest RMSE and the largest R^2, which significantly alleviated the underestimation problem. For SA, its overestimation in CNTL has been prominently reduced in both VRF_5 and VRF_6 — see Fig. 6 (middle panels). For instance, SA decreased over DBF (station UL) in both VRF_5 and VRF_6, with a larger decrease VRF_6 (Fig. 6(b)). The performance statistics of both VRF_5 and VRF_6 demonstrated improvements over most LCTs except UB (station SL) (see Table 5). For SD, the parameter optimization brought about remarkable improvement compared to FSC and SA — see Fig. 6 (right panels). Note that SD is optimized using the hourly in-situ observations (i.e., larger amount of data) while both FSC and SA are optimized using the daily satellite observations. For example, VRF_6 with DBF produced notably large SD values (Fig. 6(c)) with the lowest RMSE and the highest R^2 (Table 5), diminishing the underestimation problem in CNTL. It is hard to say which verification experiment gives the best results (i.e., VRF_5 versus VRF_6), but the performance with optimized parameters is usually better than CNTL in terms of RMSE (e.g., for most LCTs such as DBF, MF, WS, UB) and R^2 (e.g., for LCTs including DBF, MF, and CL). Overall, both VRF_5 and VRF_6 produced snow variables that are closer to observations than CNTL for most LCTs (i.e., stations), and VRF_6 generally showed the lowest RMSE and the highest R^2 in all the snow variables.

Figure 7 compares the time series of snow variables between the observations and the model simulations — CNTL and VRF_6 — for DBF represented by UL. The CNTL shows positive or negative biases in FSC, positive bias (overestimation) in SA, and negative bias (underestimation) in SD: these biases are all reduced down in VRF_6. The bias patterns in Fig. 7 are consistent with those in Fig. 6.

Lastly, we have investigated how the optimized snow parameters can affect the other variables in LSM. Figure 8 depicts the time series of the differences of LSM variables (soil temperature, sensible heat flux, and soil moisture) between VRF_6 and CNTL (i.e., VRF_6 minus CNTL) following the changes in SD. Although the LSM variables here are not directly optimized, they respond to the optimized snow parameters through associated physical processes. Note that the underestimation of SD

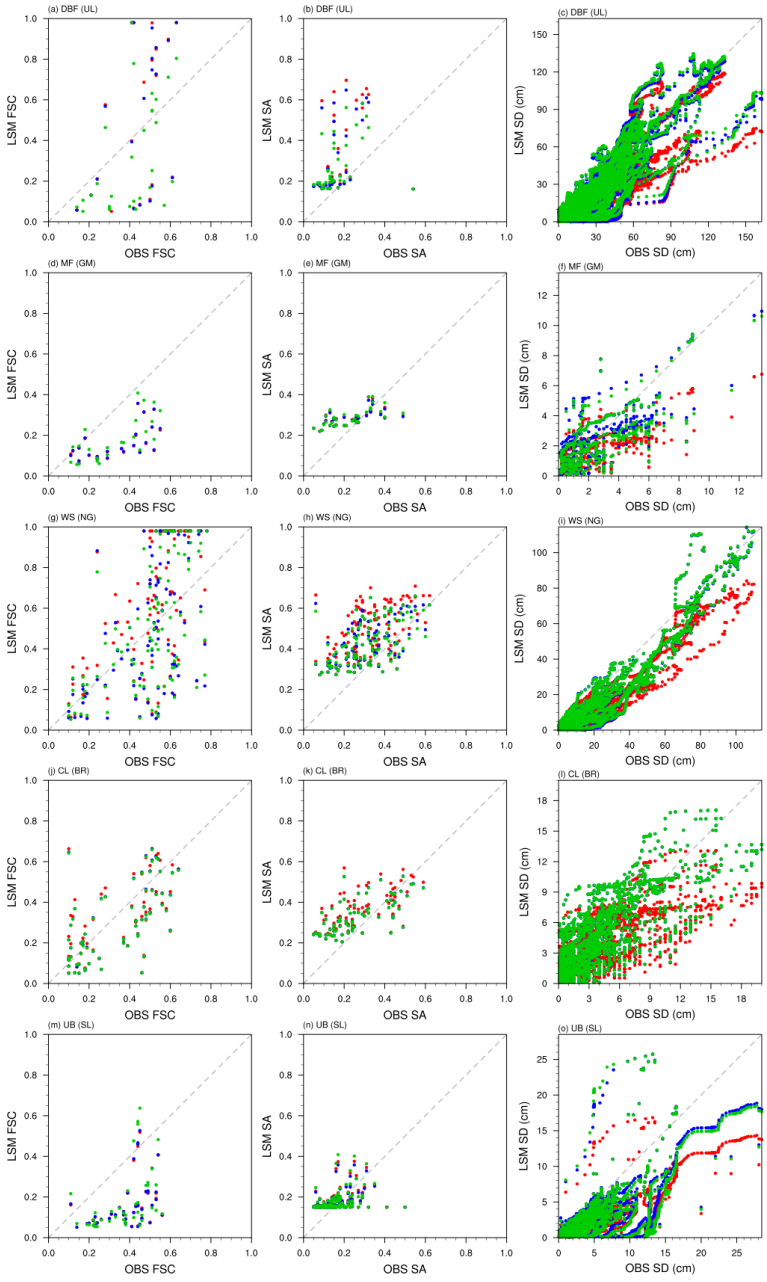

Figure 6. Scatter plots of observations (OBS) and model results (LSM) for snow variables FSC (left panels), SA (middle panels), and SD (in cm; right panels) from the verification experiments — CNTL (red dots), VRF_5 (blue dots), and VRF_6 (green dots), which are evaluated over different LCTs; (a–c) DBF represented by the station UL, (d–f) MF by GM, (g–i) WS by NG, (j–l) CL by BR, and (m–o) UB by SL.

in CNTL has been alleviated in VRF_6 by using the optimized snow parameters (see Fig. 7(c) and Fig. 8(a)). Next, soil temperature in the first soil layer (7 cm) increases as SD increases after optimization, which consequently increases sensible

Table 5. Statistics of model performance using non-optimized parameters (CNTL) and optimized parameters (VRF_5 and VRF_6) over different LCTs represented by different stations — DBF represented by UL, MF by GM, WS by NG, CL by BR, and UB by SL. The RMSEs and R^2 values are shown for three snow variables — FSC, SA, and SD.

Statistics		RMSE			R^2		
LCT	Snow Variable	CNTL	VRF_5	VRF_6	CNTL	VRF_5	VRF_6
DBF (UL)	FSC	0.328	0.327	0.252	0.248	0.215	0.256
	SA	0.218	0.197	0.159	0.157	0.157	0.176
	SD	15.763	13.640	12.616	0.764	0.781	0.796
MF (GM)	FSC	0.208	0.206	0.178	0.388	0.408	0.520
	SA	0.105	0.103	0.103	0.411	0.421	0.460
	SD	1.789	1.526	1.542	0.435	0.502	0.493
WS (NG)	FSC	0.279	0.269	0.249	0.354	0.333	0.341
	SA	0.196	0.160	0.156	0.314	0.328	0.324
	SD	9.836	8.231	8.009	0.895	0.887	0.888
CL (BR)	FSC	0.163	0.160	0.160	0.363	0.385	0.384
	SA	0.132	0.122	0.122	0.443	0.457	0.456
	SD	2.542	2.583	2.590	0.478	0.540	0.539
UB (SL)	FSC	0.255	0.252	0.242	0.184	0.195	0.195
	SA	0.071	0.070	0.073	0.150	0.148	0.124
	SD	4.790	4.286	4.699	0.484	0.449	0.385

heat flux. The residual of surface energy balance is close to zero, implying that the surface energy balance is well conserved even after optimization. Soil moisture depends on snow melt, following the trend of increased snowfall in the previous winter. Extreme fluctuations sometimes appear in the time series analyses due to nonlinear effects, but we can understand the overall tendency according to the increased SD in the land surface.

5 Discussion

Generally, the Noah LSM tends to simulate less snow amount during the peak winter and earlier snow melting and consequently overestimates SA (Saha et al., 2017). Our experiment with no optimization (CNTL) reveals underestimation of SD and FSC and overestimation of SA compared to the in-situ or satellite observations. We developed a coupled system of micro-GA and Noah LSM to reduce the uncertainties in parameterized snow processes through optimization of parameter values. This parameter estimation is an effort to further improve the model performance by reducing uncertainty in pre-existing parameterization schemes by optimizing the parameter values inside the schemes based on the observational data that reflect local characteristics to improve snow simulation. If the employed parameterization scheme has less uncertainty, improvement by parameter estimation on that scheme may not be significant; if the scheme has large uncertainty in parameter values, parameter

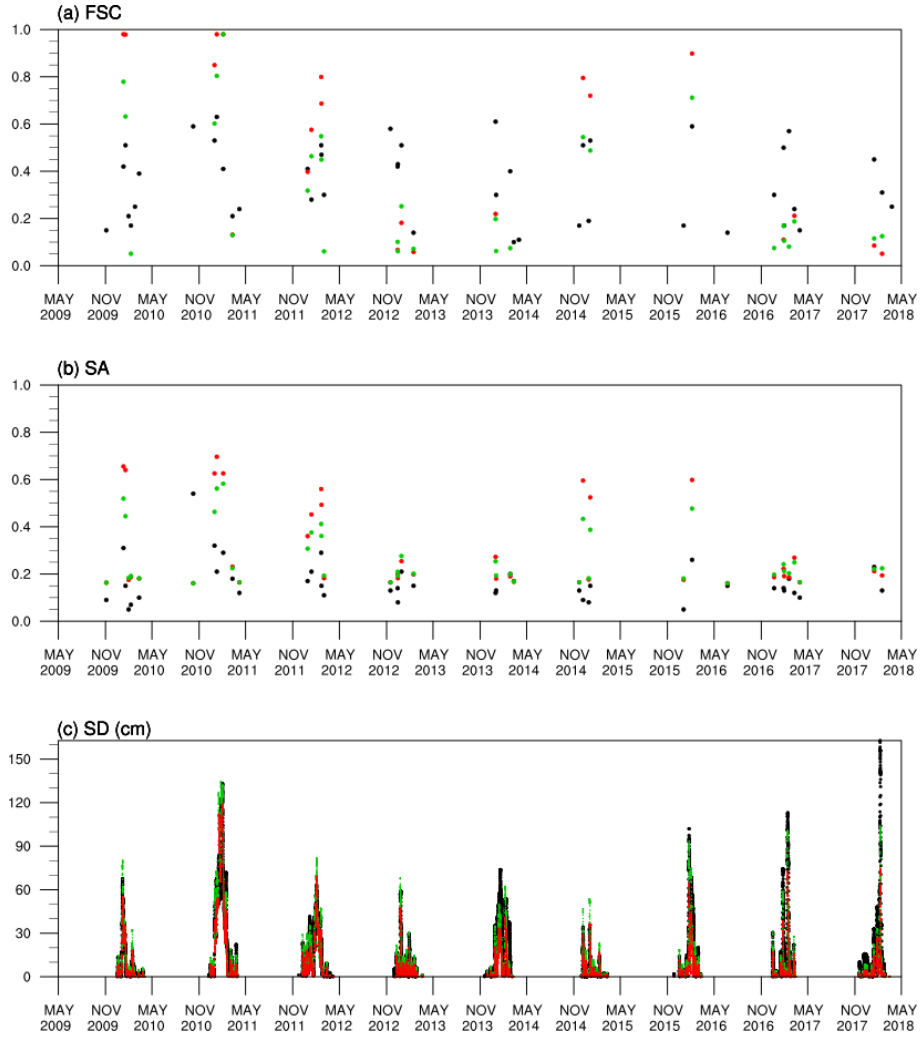

Figure 7. Time series of the snow variables for DBF from May 2009 to April 2018: (a) FSC, (b) SA, and (c) SD (in cm). Observations are in black dots and model results are in red dots for CNTL and in green dots for VRF_6.

estimation may bring about prominent improvement in the scheme's performance. Our results showed improvement in all snow variables in terms of RMSE by 3.3 %, 6.2 %, and 17.0 % for FSC, SA, and SD, respectively. Furthermore, SD increased after optimization, which lead to increases in both soil temperature and sensible heat flux due to insulating response; soil moisture also increased due to increased SD in previous years. This implies that the optimized snow parameters not only let the model solutions close to the observations but also act in a physically consistent manner. In case of some worsen statistics such as MB or R^2 in VRF_6, the insufficient stations used for optimization or a coarse resolution in satellite observation can limit to improve the snow variables. As the further study, the online Noah LSM can help to include more observation stations by

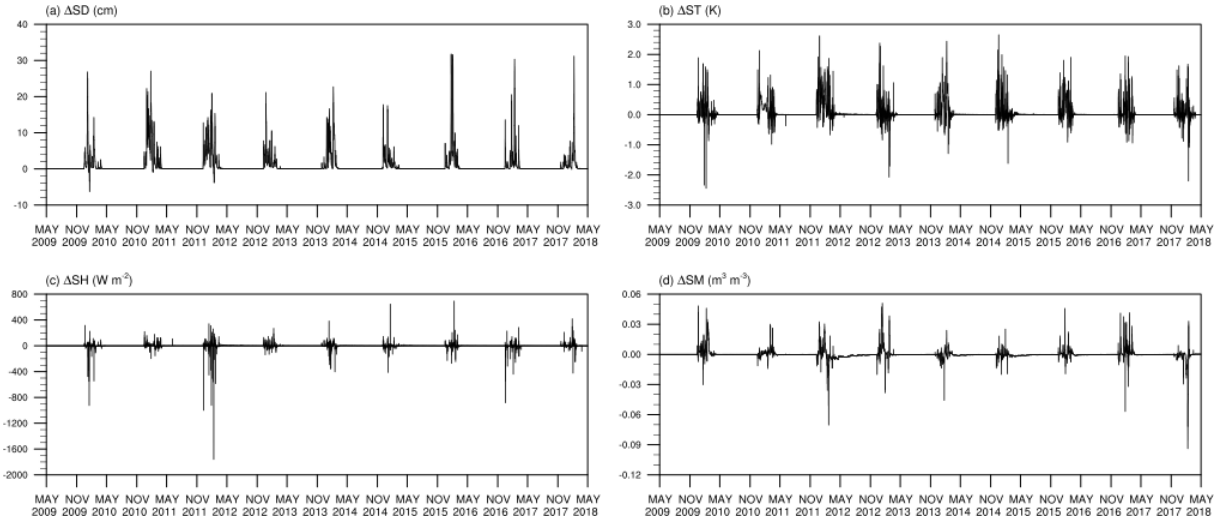

Figure 8. Time series of difference between CNTL to VRF_6 for the UL in DBF during the May 2009 to April 2018: (a) SD (cm), (b) soil temperature at the top soil layer (7 cm) (ST; K), (c) sensible heat flux (SH; W m^{-2}), (d) soil moisture at the top soil layer (7 cm) (SM; m^3 m^{-3}).

covering the all grid points over SK. Moreover, we can optimize other parameters that indirectly affects to snow processes not only direct parameters used in this study.

The coupling system of micro-GA and Noah LSM automatically estimates the optimal snow-related parameters by objectively comparing observations and model solutions through the fitness function. Instead of trial-and-error procedures, it has an advantage to reduce a substantial amount of computational time. The original micro-GA reduces the computational time using the elitism and re-initialization methods in the small number of individuals. We have developed a parallel system on the coupled system to further improve the computational efficiency in this study; it enables us to simultaneously execute multiple individuals in one generation and multiple Noah LSM runs in one individual.

Based on the encouraging optimization results in the off-line Noah LSM, we plan to optimize the Noah LSM in a coupled land-atmosphere prediction system. The online Noah LSM can produce a spatial distribution of model variables over the land surface, which allows a two-dimensional assessment of model performance. We anticipate the optimized snow parameters can lead to positive effects on the atmospheric variables through the changes of heat fluxes as well as snow variables in Noah LSM. As a result, we can identify how optimal parameters are appreciated in SK in terms of both horizontal and vertical distributions. In addition, our coupled system of micro-GA and Noah LSM can be utilized to optimize other parameters in Nosh LSM.

6 Conclusions

Six parameters included in the snow processes in Noah Land Surface Model (Noah LSM) have been optimized by using a micro-genetic algorithm (micro-GA) during the period 2009-2018 in South Korea (SK). The first parameter — is the distri-

bution shape parameter that participates in the snow cover fraction calculation and shows a positive correlation with the snow cover fraction: the optimized value is expected to increase the snow cover fraction, but it is not sufficient to alleviate its under-estimation problems. The second parameter — is snow water equivalent threshold value that implies 100 % snow cover and also is used in the snow cover fraction calculation depends on the land cover type: its optimized value improves the snow cover fraction in terms of RMSE and mean bias over some stations. The third parameter — is the maximum snow albedo coefficient: its decreased optimized value improves the RMSE reducing the over-estimation of snow albedo. The fourth parameter — is the coefficient in the maximum albedo of fresh snow, and its optimized value was similar to the default one. The other two parameters — are related to the fresh snow density used for the snow depth calculation. In particular the sixth parameter — is the coefficient depends on the air temperature is the most effective. The optimized reduced value produces the biggest improvement in the RMSE for the snow depth and it remarkably reducing the under-estimation of the snow depth. The reason for the largest improvement of snow depth is due to the higher spatial and temporal resolutions of observations. Although satellite observations have a limitation, their role in the optimization was also effective. The best combinations of snow parameters optimized for SK can be used to improve the snowfall prediction. The micro-GA has proved helpful to optimize the parameter by maximizing the prescribed fitness function, without need of time-consuming multiple trial-and-error sensitivity tests. Also, this coupling system of micro-GA and Noah LSM also can be used in other areas to optimize remained uncertain parameters in the Noah LSM. The next step of this study could be to identify how these optimized values interact with the numerical weather prediction model by indirectly change the surface fluxes.

Code availability. The current version of the Noah LSM is available from the website: https://ral.ucar.edu/solutions/products/unified-noah-lsm (last access: 4 April 2022). The current version of the GA is available from the website: https://cuaerospace.com/products-services/genetic-algorithm/ga-drive-free-download (last access: 4 April 2022). The exact version of Noah LSM and GA used in this study are archived at: https://doi.org/10.5281/zenodo.6873384 (Lim et al., 2021). It also contains the forcing data and output files of Noah LSM and micro-GA-Noah LSM coupled system and the scripts to plot the same figures as in this manuscript.

Data availability. The 1-hourly forcing data for Noah LSM are obtained from Open MET Data Portal, which is available at https://data.kma.go.kr (last access: 4 April 2022) and ERA5-Land, which is available at https://cds.climate.copernicus.eu (last access: 4 April 2022). The snow depth is also obtained from Open MET Data Portal. The daily fractional snow cover and snow albedo from MODIS/Terra Snow Cover Daily L3 Global 500 m SIN Grid, Version 61, is available at https://nsidc.org/data/MOD10A1 (last access: 4 April 2022).

Author contributions. SL, SKP, HJG, WL, YHL, and CC contributed to conceptualization. SL, SKP, and CC designed the experiments and SL carried them out with the investigation. SL, HJG, and EL developed the model code and EL and SYL contributed to the validation. SL prepared the manuscript with contributions from all co-authors.

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

Acknowledgements. This work is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1A6A1A08025520) and Development of Numerical Weather Prediction and Data Application Techniques (NTIS-1365003222) funded by the Korea Meteorological Administration. It is partly supported by the NRF grant funded by the Korea government (MSIT) (NRF-2021R1A2C1095535).

References

Anderson, E. A.: National Weather Service River Forecast System: Snow Accumulation and Ablation Model, vol. 17, US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, 1973.

Annan, J. D. and Hargreaves, J. C.: Efficient parameter estimation for a highly chaotic system, Tellus A: Dyn. Meteorol. Oceanogr., 56, 520–526, 2004.

Bonekamp, P. N. J., Collier, E., and Immerzeel, W. W.: The impact of spatial resolution, land use, and spinup time on resolving spatial precipitation patterns in the Himalayas, J. Hydrometeorol., 19, 1565–1581, 2018.

Carroll, D. L.: Genetic algorithms and optimizing chemical oxygen-iodine lasers, Developments in theoretical and applied mechanics, 18, 411–424, 1996.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., Duan, Q. Y., Ek, M., and Betts, A.: Modeling of land surface evaporation by four schemes and comparison with FIFE observations, J. Geophys. Res.-Atmos., 101, 7251–7268, 1996.

Cheong, S.-H., Byun, K.-Y., and Lee, T.-Y.: Classification of snowfalls over the Korean Peninsula based on developing mechanism, Atmosphere, 16, 33–48, 2006 (In Korean with English abstract).

Chinta, S. and Balaji, C.: Calibration of WRF model parameters using multiobjective adaptive surrogate model-based optimization to improve the prediction of the Indian summer monsoon, Clim. Dynam., 55, 631–650, 2020.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., and Tarpley, J. D.: Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), J. Geophys. Res.-Atmos., 108, 8845, https://doi.org/10.1029/2002JD003316, 2003.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H., Gusev, Y., Habets, F., Hall, A., Hay, L., et al.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, J. Hydrol., 320, 3–17, 2006.

Duan, Q., Di, Z., Quan, J., Wang, C., Gong, W., Gan, Y., Ye, A., Miao, C., Miao, S., Liang, X., et al.: Automatic model calibration: A new way to improve numerical weather forecasting, Bull. Am. Meteorol. Soc., 98, 959–970, 2017.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, J. Geophys. Res.-Atmos., 108, https://doi.org/doi:10.1029/2002JD003296, 2003.

Folberth, C., Elliott, J., Müller, C., Balkovič, J., Chryssanthacopoulos, J., Izaurralde, R. C., Jones, C. D., Khabarov, N., Liu, W., Reddy, A., Schmid, E., Skalský, R., Yang, H., Arneth, A., Ciais, P., Deryng, D., Lawrence, P. J., Olin, S., Pugh, T. A. M., Ruane, A. C., and Wang, X.: Parameterization-induced uncertainties and impacts of crop management harmonization in a global gridded crop model ensemble, PLoS One, 14, e0221 862, https://doi.org/10.1371/journal.pone.0221862, 2019.

Friedl, M. and Sulla-Menashe, D.: MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05 Deg CMG V006. NASA EOSDIS Land Processes DAAC, accessed on 21-MAY-2020, 2015.

Golberg, D. E.: Genetic Algorithms in Search, Optimization, and Machine Learning, Addion-Wesley, Reading, MA, 1989.

Gotleib, L.: A general runoff model for snowcovered and glacierized basins, in: Nord. Hydrol. Conf, vol. 6, pp. 172–177, 1980.

Gubler, S., Gruber, S., and Purves, R. S.: Uncertainties of parameterized surface downward clear-sky shortwave and all-sky longwave radiation, Atmos. Chem. Phys., 12, 5077–5098, 2012.

Günther, D., Marke, T., Essery, R., and Strasser, U.: Uncertainties in snowpack simulations—Assessing the impact of model structure, parameter choice, and forcing data error on point-scale energy balance snow model performance, Water Resour. Res., 55, 2779–2800, 2019.

Hall, D. K. and Riggs, G. A.: Accuracy assessment of the MODIS snow products, Hydrol. Process., 21, 1534–1547, 2007.

Hall, D. K. and Riggs, G. A.: MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 6. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA, https://doi.org/10.5067/MODIS/MOD10A1.006, accessed on 21-MAY-2020, 2016.

Hall, D. K., Riggs, G. A., and Salomonson, V. V.: MODIS/Terra Snow Cover 5-Min L2 Swath 500m, Version 5. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA, https://doi.org/10.5067/ACYTYZB9BEOS, 2006.

Holland, J. H.: Genetic algorithms and the optimal allocation of trials, SIAM J. Comput., 2, 88–105, 1973.

Holland, J. H.: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, University of Michigan Press, Ann Arbor, MI, 1975.

Hong, S., Yu, X., Park, S. K., Choi, Y.-S., and Myoung, B.: Assessing optimal set of implemented physical parameterization schemes in a multi-physics land surface model using genetic algorithm, Geosci. Model Dev., 7, 2517–2529, 2014.

Hong, S., Park, S. K., and Yu, X.: Scheme-based optimization of land surface model using a micro-genetic algorithm: Assessment of its performance and usability for regional applications, SOLA, 11, 129–133, 2015.

Hong, S.-Y., Kwon, Y. C., Kim, T.-H., Kim, J.-E. E., Choi, S.-J., Kwon, I.-H., Kim, J., Lee, E.-H., Park, R.-S., and Kim, D.-I.: The Korean Integrated Model (KIM) system for global weather forecasting, Asia-Pac. J. Atmos. Sci., 54, 267–292, 2018.

Jiang, Y., Chen, F., Gao, Y., He, C., Barlage, M., and Huang, W.: Assessment of uncertainty sources in snow cover simulation in the Tibetan Plateau, J. Geophys. Res.-Atmos., 125, e2020JD032 674, https://doi.org/https://doi.org/10.1029/2020JD032674, 2020.

Jin, J. and Miller, N. L.: Analysis of the impact of snow on daily weather variability in mountainous regions using MM5, J. Hydrometeorol., 8, 245–258, 2007.

Jonas, T., Marty, C., and Magnusson, J.: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps, J. Hydrol., 378, 161–167, 2009.

Jun, S., Park, J.-H., Boo, K.-O., and Kang, H.-S.: Analyzing off-line Noah land surface model spin-up behavior for initialization of global numerical weather prediction model, Journal of Korea Water Resources Association, 53, 181–191, (In Korean with English abstract), 2020.

Jung, S.-H., Im, E.-S., and Han, S.-O.: The effect of topography and sea surface temperature on heavy snowfall in the Yeongdong region: A case study with high resolution WRF simulation, Asia-Pac. J. Atmos. Sci., 48, 259–273, 2012.

Kim, D.-E. and Park, S. K.: Uncertainty in predicting the Eurasian snow: Intercomparison of land surface models coupled to a regional climate model, The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-15, 2019.

Kim, G., Joo, H., and Kim, H.: The study for damage effect factors of heavy snowfall disasters: Focused on heavy snowfall disasters during the period of 2005 to 2014, Journal of the Korea Academia-Industrial cooperation Society, 19, 125–136, (In Korean with English abstract), 2018.

Koo, M.-S., Baek, S., Seol, K.-H., and Cho, K.: Advances in land modeling of KIAPS based on the Noah land surface model, Asia-Pac. J. Atmos. Sci., 53, 361–373, 2017.

Koren, V., Schaake, J., Mitchell, K., Duan, Q.-Y., Chen, F., and Baker, J. M.: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, J. Geophys. Res.-Atmos., 104, 19 569–19 585, 1999.

Kotsuki, S., Terasaki, K., Yashiro, H., Tomita, H., Satoh, M., and Miyoshi, T.: Online model parameter estimation With ensemble data assimilation in the real global atmosphere: A case With the nonhydrostatic icosahedral atmospheric model (NICAM) and the global satellite mapping of precipitation data, J. Geophys. Res.-Atmos., 123, 7375–7392, 2018.

Krishnakumar, K.: Micro-genetic algorithms for stationary and non-stationary function optimization, in: Intelligent Control and Adaptive Systems, vol. 1196, pp. 289–296, International Society for Optics and Photonics, 1990.

Lee, J., Kim, S.-M., Park, H.-S., and Woo, B.-H.: Optimum design of cold-formed steel channel beams using micro Genetic Algorithm, Eng. Struct., 27, 17–24, 2005.

Lee, Y. H., Park, S. K., and Chang, D.-E.: Parameter estimation using the genetic algorithm and its impact on quantitative precipitation forecast, Ann. Geophys., 24, 3185–3189, 2006.

Li, J., Duan, Q., Wang, Y.-P., Gong, W., Gan, Y., and Wang, C.: Parameter optimization for carbon and water fluxes in two global land surface models based on surrogate modelling, Int. J. Climatol., 38, e1016–e1031, https://doi.org/https://doi.org/10.1002/joc.5428, 2018.

Li, J., Chen, F., Lu, X., Gong, W., Zhang, G., and Gan, Y.: Quantifying contributions of uncertainties in physical parameterization schemes and model parameters to overall errors in Noah-MP dynamic vegetation modeling, J. Adv. Model. Earth Sy., 12, e2019MS001 914, https://doi.org/10.1029/2019MS001914, 2020.

Lim, S., Gim, H.-J., Lee, E., Lee, S.-Y., Lee, W. Y., Lee, Y. H., Cassardo, C., and Park, S. K.: Code and Data: Optimization of Snow-Related Parameters in Noah Land Surface Model (v3.4.1) Using Micro-Genetic Algorithm (v1.7a), Zenodo [code and data set], https://doi.org/10.5281/zenodo.6873384, 2021.

Livneh, B., Xia, Y., Mitchell, K. E., Ek, M. B., and Lettenmaier, D. P.: Noah LSM snow model diagnostics and enhancements, J. Hydrometeorol., 11, 721–738, 2010.

Mallet, V. and Sportisse, B.: Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, J. Geophys. Res.-Atmos., 111, D01 302, https://doi.org/10.1029/2005JD006149, 2006.

Mitchell, K. E.: The community Noah land-surface model (LSM): User's guide public release version 2.7.1, NCEP/EMC Doc., 26 pp., 2005.

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A.: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, J. Geophys. Res.-Atmos., 109, https://doi.org/10.1029/2003jd003823, 2004.

Muñoz-Sabater, J.: ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.e2161bac, accessed on 20-APR-2021, 2019.

Neelin, J. D., Bracco, A., Luo, H., McWilliams, J. C., and Meyerson, J. E.: Considerations for parameter optimization and sensitivity in climate models, Proceedings of the National Academy of Sciences, 107, 21 349–21 354, 2010.

Olafsson, H. and Bao, J.-W., eds.: Uncertainties in Numerical Weather Prediction, Elsevier, 364 pp., 2020.

Pan, M., Sheffield, J., Wood, E. F., Mitchell, K. E., Houser, P. R., Schaake, J. C., Robock, A., Lohmann, D., Cosgrove, B., Duan, Q., et al.: Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent, J. Geophys. Res.-Atmos., 108, https://doi.org/https://doi.org/10.1029/2003JD003994, 2003.

Park, S. and Park, S. K.: A micro-genetic algorithm (GA v1.7.1a) for combinatorial optimization of physics parameterizations in the Weather Research and Forecasting model (v4.0.3) for quantitative precipitation forecast in Korea, Geosci. Model Dev., 14, 6241–6255, 2021.

Pathak, R., Sahany, S., and Mishra, S. K.: Uncertainty quantification based cloud parameterization sensitivity analysis in the NCAR community atmosphere model, Sci. Rep., 10, 17 499, https://doi.org/10.1038/s41598-020-74441-x, 2020.

Robinson, D. A. and Kukla, G.: Maximum surface albedo of seasonally snow-covered lands in the Northern Hemisphere, J. Appl. Meteorol. Clim., 24, 402–411, 1985.

Rudnaya, S. and Santosa, F.: Application of a micro-genetic algorithm in optimal design of a diffractive optical element, in: System Modelling and Optimization: Methods, Theory and Applications. CSMO 1999, IFIPAICT, vol. 46, edited by Powell, M. J. D. and Scholtes, S., pp. 251–267, Springer, Boston, MA, USA, https://doi.org/10.1007/978-0-387-35514-6_12, 2000.

Saha, S. K., Sujith, K., Pokhrel, S., Chaudhari, H. S., and Hazra, A.: Effects of multilayer snow scheme on the simulation of snow: O ffline N oah and coupled with NCEP CFS v2, J. Adv. Model. Earth Sy., 9, 271–290, 2017.

Sheffield, J., Pan, M., Wood, E. F., Mitchell, K. E., Houser, P. R., Schaake, J. C., Robock, A., Lohmann, D., Cosgrove, B., Duan, Q., et al.: Snow process modeling in the North American Land Data Assimilation System (NLDAS): 1. Evaluation of model-simulated snow cover extent, J. Geophys. Res.-Atmos., 108, 8849, https://doi.org/10.1029/2002JD003274, 2003.

Shutts, G. and Pallarès, A. C.: Assessing parametrization uncertainty associated with horizontal resolution in numerical weather prediction models, Philos. T. R. Soc. A., 372, 20130 284, https://doi.org/10.1098/rsta.2013.0284, 2014.

Souza, A. N., Wagner, G. L., Ramadhan, A., Allen, B., Churavy, V., Schloss, J., Campin, J., Hill, C., Edelman, A., Marshall, J., Flierl, G., and Ferrari, R.: Uncertainty quantification of ocean parameterizations: Application to the K-profile-parameterization for penetrative convection, J. Adv. Model. Earth Sy., 12, e2020MS002 108, https://doi.org/10.1029/2020MS002108, 2020.

Sulla-Menashe, D. and Friedl, M. A.: User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product, USGS, Reston, VA, USA, 18 pp., 2018.

Sultana, R., Hsu, K.-L., Li, J., and Sorooshian, S.: Evaluating the Utah Energy Balance (UEB) snow model in the Noah land-surface model, Hydrol. Earth Syst. Sc., 18, 3553–3570, 2014.

Suzuki, K. and Zupanski, M.: Uncertainty in solid precipitation and snow depth prediction for Siberia using the Noah and Noah-MP land surface models, Front. Earth Sci., 12, 672–682, 2018.

USACE: Snow Hydrology: Summary Report of the Snow Investigations, Tech. rep., US Army Corps of Engineers, North Pacific Division, Portland, Orgeon, USA, 1956.

Wang, Q., Fang, H., and Zou, X.-K.: Application of Micro-GA for optimal cost base isolation design of bridges subject to transient earthquake loads, Struct. Multidiscip. O., 41, 765–777, 2010.

Wang, S. and Sun, B.: The impacts of different land surface parameterization schemes on Northeast China snowfall simulation, Meteorol. Atmos. Phys., 130, 583–590, 2018.

Warren, S. G.: Optical properties of snow, Rev. Geophys., 20, 67–89, 1982.

Warren, S. G. and Wiscombe, W. J.: A model for the spectral albedo of snow. II: Snow containing atmospheric aerosols, J. Atmos. Sci., 37, 2734–2745, 1980.

Xu, Y., Jones, A., and Rhoades, A.: A quantitative method to decompose SWE differences between regional climate models and reanalysis datasets, Sci. Rep., 9, 16 520, https://doi.org/10.1038/s41598-019-52880-5, 2019.

Yan, J., Xu, Z., Yu, Y., Xu, H., and Gao, K.: Application of a hybrid optimized BP network model to estimate water quality parameters of Beihai Lake in Beijing, Appl. Sci., 9, 1863, https://doi.org/10.3390/app9091863, 2019.

Yoon, J. W., Lim, S., and Park, S. K.: Combinational optimization of the WRF physical parameterization schemes to improve numerical sea breeze prediction using micro-genetic algorithm, Appl. Sci., 11, 11 221, 2021.

Yu, X., Park, S. K., Lee, Y. H., and Choi, Y. S.: Quantitative precipitation forecast of a tropical cyclone through optimal parameter estimation in a convective parameterization, SOLA, 9, 36–39, 2013.

Zhang, X., Zhang, S., Liu, Z., Wu, X., and Han, G.: Parameter optimization in an intermediate coupled climate model with biased physics, J. Climate, 28, 1227–1247, 2015.

Zhao, W. and Li, A.: A review on land surface processes modelling over complex terrain, Adv. Meteorol., 2015, Article ID 607 187, https://doi.org/https://doi.org/10.1155/2015/607181, 2015.

Zhu, J., Shu, J., and Yu, X.: Improvement of typhoon rainfall prediction based on optimization of the Kain-Fritsch convection parameterization scheme using a micro-genetic algorithm, Front. Earth Sci, 13, 721–732, 2019.