Article

Evaluation of Snow and Streamflows Using Noah-MP and WRF-Hydro Models in Aroostook River Basin, Maine

Engela Sthapit [1,*], Tarendra Lakhankar [1], Mimi Hughes [2], Reza Khanbilvardi [1], Robert Cifelli [2], Kelly Mahoney [2], William Ryan Currier [2], Francesca Viterbo [2,3] and Arezoo Rafieeinasab [4]

1 NOAA Center for Earth System Sciences and Remote Sensing Technologies, The City College of New York, 160 Convent Avenue, New York, NY 10031, USA; tlakhankar@ccny.cuny.edu (T.L.); khanbilvardi@ccny.cuny.edu (R.K.)
2 NOAA Earth System Research Laboratory, Physical Sciences Laboratory, Boulder, CO 80305, USA; mimi.hughes@noaa.gov (M.H.); rob.cifelli@noaa.gov (R.C.); kelly.mahoney@noaa.gov (K.M.); william.r.currier@noaa.gov (W.R.C.); fraviterbo88@gmail.com (F.V.)
3 Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80309, USA
4 National Center for Atmospheric Research, Research Applications Laboratory, Boulder, CO 80301, USA; arezoo@ucar.edu
* Correspondence: esthapi00@citymail.cuny.edu

Abstract: Snow influences land–atmosphere interactions in snow-dominated areas, and snow melt contributes to basin streamflows. However, estimating snowpack properties such as the snow depth (SD) and snow water equivalent (SWE) from land surface model simulations remains a challenge. This is, in part, due to uncertainties in the atmospheric forcing variables, which propagate into hydrological model predictions. This study implements the Weather Research and Forecasting (WRF)-Hydro framework with the Noah-Multiparameterization (Noah-MP) land surface model in the NOAA's National Water Model (NWM) version 2.0 configuration to estimate snow in a single column and subsequently the streamflow across the Aroostook River's sub-basins in Maine for water years (WY) 2014–2016. This study evaluates how differences between two atmospheric forcing datasets, the North American Land Data Assimilation version 2 (NLDAS-2) and in situ (Station), translate into differences in the simulation of snow. NLDAS-2 was used as the meteorological forcing in the retrospective NWM 2.0 simulations. The results from the single-column study showed that differences in the simulated SWE and SD were linked to differences in the 2 m air temperature (T2m), which influenced the precipitation partitioning of rain and snow, as parameterized in Noah-MP. The negative mean bias of −0.7 K (during the accumulation period) in T2m for NLDAS-2, compared to the Station forcing, was a major factor that contributed to the positive mean bias of +52 mm on average in the peak SWE in the NLDAS-2-forced Noah-MP simulation during the study period. The higher T2m values at the Station led to higher sensible heat fluxes towards the snowpack, which led to a higher amount of net energy at the snow's surface and melt events during the accumulation season in Station-forced Noah-MP simulations. The results from the retrospective NWM version 2.0's simulation in the basin showed that the streamflow estimates were closer to the United States Geological Survey gage observations at the two larger sub-basins (NSE = 0.9), which were mostly forested, compared to the two smaller sub-basins (NSE ≥ 0.4), which had more agricultural land-use. This study also showed that the spring snowmelt timing was captured quite well by the timing of the decline in the simulated SWE and SD, providing an early indication of melt in most sub-basins. The simulated fractional snow cover area (fSCA) however provided less information about the changes in snow or onset of snowmelt as it was mostly binary (full snow cover in winter), which differed from the more realistic fSCA values shown by the Moderate Resolution Imaging Spectroradiometer.

Keywords: National Water Model; land surface model; meteorological forcing; snow water equivalent; snow depth; fractional snow cover area

1. Introduction

Snow is an important land surface characteristic during the winter due to its impact on land–atmosphere interactions. The snowpack alters the surface energy and mass balances by influencing surface heat fluxes, ground temperature, runoff, and soil moisture [1,2]. Snow on land increases surface albedo [3,4] and insulates the soil from the atmosphere [5], whereas snowmelt is important for water resource management, as it contributes to the timing and magnitude of runoff [6–8]. The calculations for energy on the ground are performed by land surface models (LSMs), which have known challenges in calculating surface fluxes over snow-covered surfaces due to poor simulations of snow water equivalents and their evolution over time [9].

The accurate prediction of hydrological processes is a challenging task for numerical weather and water prediction models. This is due to uncertainties in both meteorological forcing variables and finding the most appropriate parameterizations to accurately represent small-scale land and atmospheric processes [2]. Errors in precipitation, temperature, and humidity, which determines the partitioning of precipitation to rain or snow, affects snow accumulation and the simulated runoff; errors in downward shortwave and long-wave radiation influence the snowpack temperature and corresponding melt rates; and errors in wind speed, humidity, and snow-surface temperatures worsen the simulation of snow sublimation [10,11].

In August 2016, the National Oceanic and Atmospheric Administration (NOAA) implemented the National Water Model (NWM) as an operational hydrological model to simulate and forecast streamflows throughout the contiguous United States (CONUS). The core architecture of NWM—the Weather Research and Forecasting Hydrologic model (WRF-Hydro), supported by the National Center for Atmospheric Research (NCAR)—is a widely used hydrological model. Both NWM and WRF-Hydro have undergone continuous improvements, and numerous studies have been conducted focusing on streamflows and model states, including snow, in various regions of the US and other parts of the world [12–15].

Snowmelt-dominated basins are vulnerable to temperature increases, as they can cause a decrease in peak snow water equivalents and alter the flow timing towards earlier river flows [16,17]. In an analysis of snowpack change from 1982 to 2016 over the conterminous United States (US), the snow season was significantly shortened due to the earlier ending of the snow season over the western US and its later arrival over the eastern US [18]. Numerous studies have investigated snow storage's impact on snowmelt and streamflow forecasts in snow-dominated regions of the mountainous western US [8,17,19,20] but studies that explored streamflow predictability in the wet northeastern regions of the US are limited. One 40-year study of water budgets in the northeastern US, using Noah-MP and WRF-Hydro with the NWM configuration, documented increases in temperatures and changes in precipitation patterns, which resulted in more water availability during the winter and less during the spring as the snow accumulation dwindled during the winter [20]. It also highlighted that more substantial changes, including high and low flows, are expected to occur in river reaches closer to head waters or smaller basins compared to larger basins as the larger basins are able to dampen the signals, whereas the regional models are often too coarse to capture small-scale watershed processes [21,22].

Operational seasonal streamflow forecasts from snowmelt are mostly based on empirical relationships from historical monitoring data and statistical models [23]. Since these forecasts are only possible where the historical data exist, physics-based numerical models were developed to enable streamflow forecasts at un-gaged locations. The move towards physical models most recently culminated in the implementation of the high-resolution, operational NWM on a continental scale [24].

Point-based field observations and snow data from in situ stations can be used for calibrating numerical models, but the network of those measurements is usually sparse. Remotely sensed snow is valuable in providing continuous coverage across space and time. During the past few decades, many snow products have been developed to measure SWE or snow cover areas. Remotely sensed snow cover observations can help improve operational

snowmelt and streamflow forecasting, especially in remote areas [24]. Although the remote sensing of snow faces challenges from retrieval errors and is provided at coarse resolutions in most cases [25], the use of it will increase, as it provides opportunities for adaptive physical model predictions by representing the physical conditions at all locations [26]. In remote sensing, Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover products have been used for many hydrological and modeling applications [27]. In this paper, we used the cloud-gap-filled fractional snow cover area (fSCA) product derived from Moderate Resolution Imaging Spectroradiometer (MODIS) on the TERRA satellite [28].

This study implements the WRF-Hydro framework with the Noah-Multiparameteriz-ation (Noah-MP) LSM [29] in the NWM version 2.0 configuration (parametrizations and calibration) to estimate snow in a single column as well as across the basin, along with streamflow estimates in the basin. The overall goal was to evaluate the Noah-MP and WRF-Hydro's ability to simulate snow and runoff processes both at a single-column and a basin scale. A single-column experiment helps to understand how the uncertainties in the forcing data would affect snow in the model simulations. In the single column, the lateral water transfer mechanisms are deactivated; as such, only processes occurring in a vertical column remain, giving more clarity to the relationship between atmospheric forcing and snow on the ground. Ultimately, comparing the snow-specific performance of Noah-MP with different forcing datasets provides insight into the physical reasons underlying LSM behavior, which could help in the development of future versions of the NWM or other process-based numerical prediction models.

Specifically, the objectives of this study were to (1) use a single-column experiment to explore the differences between the snow depth (SD) and snow water equivalent (SWE) from the Noah-MP model simulated with two forcing datasets—North American Land Data Assimilation System-2 (NLDAS-2-Noah-MP simulation) and in-situ Station (Station-Noah-MP simulation)—that provided the meteorological measurements. NLDAS-2 is used as a meteorological forcing for the NWM 2.0 retrospective dataset, (2) to compare the simulated snow with the snow observed at an in situ location in Caribou, Maine, (3) to evaluate the performance of WRF-Hydro in a retrospective NWM version 2.0 simulation (forced with NLDAS-2 meteorology) to predict the streamflow by comparing it to observed flows at four USGS gage locations in the Aroostook River sub-basins, and (4) to evaluate snowmelt timings and peak flows in relation to the changes in SWE, SD, and fSCA from the retrospective NWM version 2.0 simulation and MODIS's MOD10A1F remotely sensed snow cover dataset.

We assess the capability of WRF-Hydro with Noah-MP LSM to simulate hourly basin-scale runoff in the Aroostook River sub-basins. We do not attempt to perform calibrations of any parameters; instead, our focus was to evaluate the model's performance in NWM version 2.0 formulation. We demonstrate the model's performance using default param-eters and provide information about its biases, which could lead to future numerical model improvements.

The following Section 2 describes the study area, followed by Section 3 for methods and data, which describes the datasets, NWM 2.0 parameterization, and methods for the single-column experiment and basin-level study. Next, the results are presented in Section 4, including analysis of snow and forcing in the NLDAS2-Noah-MP and Station-Noah-MP simulations in a single-column study, along with the comparison of the observed values. It also includes the basin-level study that evaluates the observed streamflow with those simulated from WRF-Hydro in relation to the changes in snow variables at four stream gage locations within the basin. In Sections 5 and 6, we conclude with discussions on the significant findings of this study, insights from other studies, and limitations of this work and future work ideas.

2. Study Area

The study area, ARB, Hydrologic Unit Code (HUC) 8-01010004, part of the northeast-ern United States, lies between 48–48.5° N and 67–70° W and has an approximate area

of 4922 sq. km. (Figure 1a). In situ meteorological observations were obtained from the CUNY-Snow Analysis and Field Experiment (CUNY-SAFE) site and National Weather Services (NWS) Forecast Station in Caribou, ME. The CUNY-SAFE site is located at 46.86691° N and 68.01330° W, and the NWS site is located at 46.87049° N and 68.01723° W (Figure 1b). These sites are referred to hereafter as the stations. The NWS station is located in an open area, with thee Caribou Municipal Airport to the north and west, and a forested area and residential areas further to the south and east. The CUNY-SAFE station is located south of the NWS station, with mostly open areas to the north and west and residential areas to the east and south (Figure 1b).
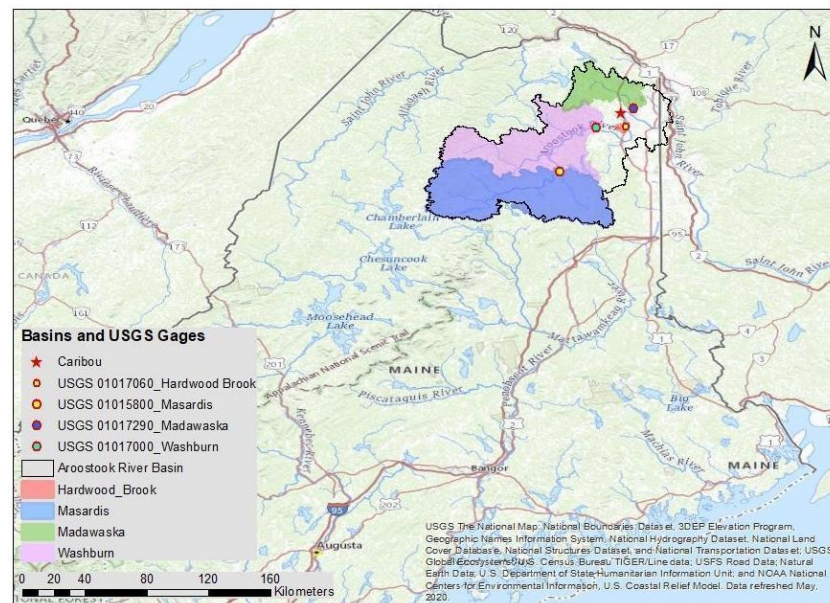
(a)

(b)

Figure 1. (**a**). Locations of the basin, sub-basins, and gaging stations used in this study. (**b**). Location of the CUNY-SAFE and NWS weather stations.

There are four USGS streamflow gages located at the outlet of four sub-basins within the Aroostook River Basin (ARB): USGS 01017060 Hardwood Brook below Glidden Brook

near Caribou, USGS 01017290 Little Madawaska River at Caribou, USGS 01015800 Aroostook River near Masardis, and USGS 01017000 Aroostook River at Washburn, referred hereafter as Hardwood, Madawaska, Masardis, and Washburn stations, respectively. The area of each watershed, as well as the range in elevation and slope of each sub-basin, are shown in Table 1. The basin land cover is mostly forested with some cropland, grassland, and urban areas. The fractional area for major land-cover classes in each sub-basin (Figure 2) is as determined from the United States Geological Survey (USGS) North American Landcover Dataset (NLCD) [30].

Table 1. Geographic characteristics of selected sub-basins.

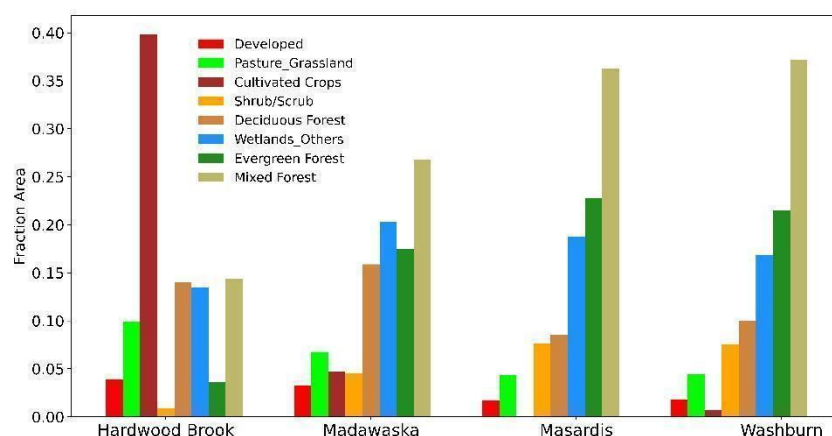| Index	Hardwood Brook	Madawaska	Masardis	Washburn
Area (km^2)	15	604	2338	4303
Elevation (m)	146–214	123–360	161–746	133–746
Percent Slope	0–35	0–73	0–189	0–189

Figure 2. Major Land-cover classes for sub-basins determined by USGS NLCD.

3. Methods and Data

3.1. National Water Model, WRF-Hydro, and Noah-MP Model

The NOAA's NWM is a hydrologic modeling framework that simulates observed and forecast streamflows over the CONUS. It provides complementary hydrological guidance at approximately 3600 official NWS locations across the CONUS and produces guidance at approximately 2.7 million streamflow locations that do not have traditional river forecasts. The core of NWM is the WRF-Hydro modeling system, which allows the coupling of hydrological model components to an atmospheric model such as WRF, but also functions in an un-coupled mode [31]. WRF-Hydro's architecture has Noah-MP as a land surface model option, which allows the users to select multiple physics options that enable specific parameterizations to solve the mass and energy balance in a one-dimensional vertical column [29,31,32]. WRF-Hydro has surface/sub-surface flow and channel routing options that provide lateral water transport for water leaving the one-dimensional column setting.

For this study, WRF-Hydro v5.1.1 [31] in the un-coupled mode was used. The simulation had the same domain physics options as the version 2.0 of NWM, with the resolution of 1 km Noah-MP LSM and 250 m overland routing. The 'namelist.hrldas' parameterization options in Noah-MP used for this study are as shown in Table 2, which can be found at [33]. In the basin study, runoff was simulated through parameterization options in the 'hydro.namelist' [33] such as surface and subsurface flows, and the Muskingum–Cunge channel routing down the National Hydrography Dataset (NHDPlusV2) river network. For the single-column study, however, all the flows and routing options were deactivated since the streamflow was not simulated.

Table 2. Name list for physics options selected in Noah-MP.

Physics	Model Selected Option
Dynamic Vegetation Option	4—Table LAI
Canopy Stomatal Resistance Option	1—Ball-Berry
Soil Moisture Factor for Stomatal Resistance	1—Noah
Runoff and Groundwater	3—Original surface and sub-surface runoff (free drainage)
Surface Layer Drag Coefficient	1—M-O
Frozen Soil Permeability	1—no iteration, Niu and Yang, 2006, JHM
Supercooled Liquid Water	1—linear effects, more permeable, Niu and Yang, 2006, JHM
Radiative Transfer	3—two-stream applied to vegetated fraction (gap = 1-FVEG)
Snow Surface Albedo	1—BATS
Precipitation Partitioning	1—Jordan 1991
Lower Boundary Condition of Soil Temperature	2—Noah
Snow/Soil Temperature Time Scheme	3—semi-implicit with FSNO for TS
Surface Resistance	4—separate non-snow and snow
Glacier Treatment	2—Noah

3.2. Meteorological Forcing

NWM V2.0 retrospective analysis uses NLDAS-2 for meteorological forcing. NLDAS-2 is a retrospective forcing dataset that integrates a large quantity of observation-based and model reanalysis data to drive offline (not coupled to atmosphere) LSMs [34,35]. NLDAS-2 operates at 1/8th degree grid spacing over central North America, with an hourly temporal resolution, and comprises data from Jan 1979 to present. Land-surface forcing files for NLDAS-2 are derived from the analysis fields of the National Center for Environmental Prediction's (NCEP) North American Regional Reanalysis (NARR). The NLDAS-2 forcing dataset used here—NLDAS-2 Primary Forcing Data L4 Hourly 0.125 × 0.125-degree V002 (NLDAS_FORA0125_H)—is available from the National Aeronautics and Space Administration (NASA) Goddard Earth Sciences Data and Information Services Center (GES DISC). Forcing from NLDAS-2 was re-gridded and downscaled from the native 0.125-degree resolution to the 1-km Noah-MP land-surface modeling domain to drive the model. The re-gridded and downscaled NLDAS-2 files for NWM 2.0 simulation were provided by NCAR (accessed in March 2020).

The in-situ measurements for incoming/outgoing shortwave (SW) radiation, incoming/outgoing longwave (LW) radiation, and albedo were measured using a Campbell Scientific NR01 net radiation sensor at the CUNY-SAFE station ('Daily CREST station meteo data files') [36]. A Campbell Scientific GMON gamma ray sensor was used to measure SWE ('Daily CREST station gamma-ray SWE data files') [36]. The hourly 2 m air temperature (T2m), total precipitation, relative humidity (RH%), and wind speed were obtained from the National Weather Service Forecast Office Caribou, ME, which uses Automated Surface Observing Systems (ASOS). ASOS uses heated tipping bucket precipitation gauges with a 12-inch diameter collector funnel, pivoting dual chamber buckets, and surrounding wind shields to measure the liquid and frozen precipitation. It has an accuracy of 0.02 inches or 4% of the hourly total (whichever is greater) [37]. They can also be accessed from the NOAA NESDIS website ('Daily Caribou NWS synoptic data files') [36]. The ground-based measurements are subject to observational biases.

3.3. Single-Column Experiment

The snow parameters, SWE and SD, were analyzed for two open-loop (no assimilation) retrospective model simulations: (1) 6-year, hourly simulation forced by NLDAS-2, downscaled and re-gridded on the 1 km Noah-MP land-surface grid, (2) 6-year, hourly simulation forced by in situ station (Station) forcing variables. The snow variable outputs

were analyzed against in situ observations from the CREST-SAFE and NWS station in Caribou, Maine. As the hydro options such as the terrain routing and channel/groundwater flows were all deactivated in the single-column study, the simulations are only from the LSM (i.e., Noah-MP). Simulations of SWE and SD from the Noah-MP model with atmospheric forcing from NLDAS-2 and the in-situ station (referred to as NLDAS2-Noah-MP and Station-Noah-MP model simulation hereafter) were generated for years 2013 through 2019. The year 2013 was not included in the analysis to allow for LSM spin-up; the analysis therefore focused on 2014–2019. For days where station forcing variables were missing, Station-Noah-MP was simulated using NLDAS-2 forcing. For water year 2019 (WY2019), where the SWE and SD station measurements were available, observed SWE and SD values were compared with the values from the two single-column experiments—NLDAS2-Noah-MP and Station-Noah-MP. Similar comparisons were made for the forcing variables—incoming LW radiation, incoming SW radiation, T2m, precipitation, and wind speed, for the entire 2014–2019 time period—to quantify differences in the NLDAS-2 values versus the Station values.

The significant differences were tested with a paired-wise Student's t-test at the 0.05 (alpha) significance level [38] for hourly and average monthly values; hereafter, we define 'significantly' differently, as meaning the differences that satisfied this criterion. The tests were conducted for each water year (1 October–30 September) and by the snow accumulation (December–March) and melt (April–May) periods in the water year.

Our single-column experiment does have limitations due to difference in scales—differences in point-based ground measurements and 1 km NLDAS-2 forcing—and the resulting simulation outputs will be partially due to the differences in scales.

3.4. Comparing Simulated with Observed Streamflows

We compared simulated hourly streamflow results with those observed at the four USGS streamflow gage locations with standard goodness-of-fit measures such as Nash-Sutcliffe Efficiency (NSE), Percent Bias (PBIAS), Root Mean Square Error Standard Deviation Ratio (RSR), and Kling-Gupta Efficiency (KGE). Models with NSE > 0.5, PBIAS ± 25%, and RSR < 0.70 were considered to represent satisfactory streamflow simulations [39].

NSE is a normalized statistic that determines the relative magnitude of the residual variance 'noise' compared to the measured data variance 'information' (Equation (1)). NSE ranges from $-\infty$ to 1; values closer to 1 indicate better model predictions.

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n} \left(Y_i^{obs} - Y_i^{sim}\right)^2}{\sum_{i=1}^{n} \left(Y_i^{obs} - Y_{mean}^{obs}\right)^2} \tag{1}$$

PBIAS measures the average tendency of the simulated data to be larger or smaller than the observed counterparts; it is the deviation of the data being evaluated, expressed as a percentage (Equation (2)). The optimal value for PBIAS is 0; positive values indicate model underestimation (negative bias), and negative values indicate model overestimation (positive bias).

$$\text{PBIAS} = \frac{\sum_{i=1}^{n} \left(Y_i^{obs} - Y_i^{sim}\right)}{\sum_{i=1}^{n} \left(Y_i^{obs}\right)} \times 100\% \tag{2}$$

RSR is the standardized value of the Root Mean Square Error (RMSE) using the standard deviation of the observation data (Equation (3)), where RMSE is the standard deviation of the residual of the prediction errors, or a measure of the spread of the residuals. RSR varies from 0 to a large positive value; values closer to 0 indicate a more accurate model or less RMSE or residual variation.

$$\text{RSR} = \frac{\text{RMSE}}{STDEV_{obs}} = \frac{\sqrt{\sum_{i=1}^{n} \left(Y_i^{obs} - Y_i^{sim}\right)^2}}{\sqrt{\sum_{i=1}^{n} \left(Y_i^{obs} - Y_{mean}^{obs}\right)^2}} \tag{3}$$

KGE compares the modeled and observed data based on the correlation, average, and variability (Equation (4)). The values range from $-\infty$ to 1, with a more accurate model being closer to 1.

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}$$
$$r = \frac{cov_{so}}{\sigma_s \sigma_o}, \ \alpha = \frac{\sigma_s}{\sigma_o}, \ \beta = \frac{\mu_s}{\mu_o} \tag{4}$$

where cov_{so} is the covariance between the simulated and observed values for calculating r, which is the correlation coefficient; σ_s and σ_o are standard deviation values of the simulated and observed values, respectively; and μ_s and μ_o are the mean simulated and observed values, respectively.

3.5. Fractional Snowcover Area from MODIS

In this study, we use the MOD10A1F cloud-gap filled (CGF) snow cover dataset from the MODIS/TERRA Snow Cover Daily L3 Global 500m SIN Grid (MOD10A1) product [28]. In MOD10A1F, the MOD10A1 grid cells obscured by cloud cover are filled by retaining previous days' clear sky values. Snow cover is detected using the Normalized Difference Snow Index (NDSI)—a spectral band ratio between visible wavelengths (0.4–0.7 μm) and the short infrared region (1–4 μm). Snow-covered land typically has high reflectance in visible wavelengths and low reflectance in the short infrared wavelengths, and NDSI reveals the magnitude of this difference. The NDSI snow algorithm for the above product uses Terra MODIS band 4 (*b4*, center at 0.555 μm) and band 6 (*b6*, center at 1.640 μm) [40,41], calculated as the following ratio:

$$\text{NDSI} = \frac{b4 - b6}{b4 + b6} \tag{5}$$

This study uses the following regression relationship developed for fractional snow cover area (fSCA) using TERRA MODIS NDSI observations [38]:

$$\text{fSCA} = -0.01 + 1.45 \times \text{NDSI} \tag{6}$$

MOD10A1 does suffer from reductions in accuracy in forests due to obstruction by the canopy [42].

4. Results

4.1. Single-Column Experiment

4.1.1. Comparison between the Simulations

Simulated SWE and SD from NLDAS2-Noah-MP were significantly higher than Station-Noah-MP for all water years (Figure 3a,b), with a positive mean bias of +52 mm and +200 mm in NLDAS2-Noah-MP (at peak SWE), in SWE, and SD, respectively (Table 3). In the paragraphs that follow, these differences were investigated in relation to the differences in forcing and other simulation outputs.
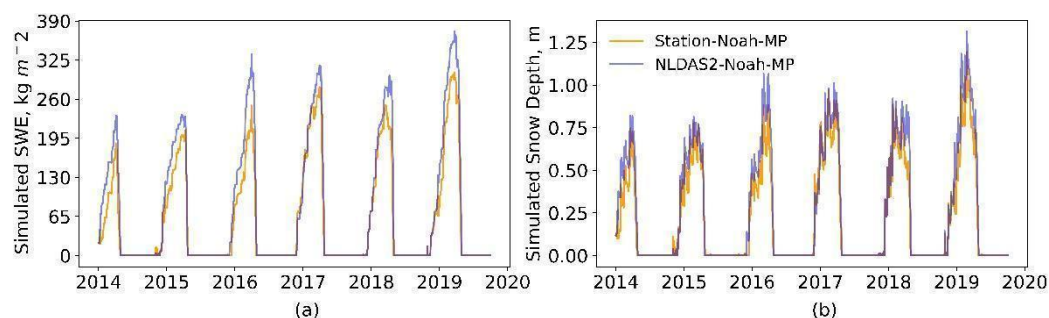

Figure 3. Simulated (**a**) Snow Water Equivalent; (**b**) Snow Depth, from Noah-MP Model versus Station-Noah-MP from WY 2014–2019.

Table 3. Mean Bias in NLDAS-2 and NLDAS2-Noah-MP compared to the Station and Station-Noah-MP (WY 2014–2019).

Period	Forcing					Simulation Outputs					
	In. SW (W/m²)	In. LW (W/m²)	T2m (K)	Wind (m/s)	RH (%)	Abs. SW (W/m²)	Net LW (W/m²)	Trad (K)	Alb-edo	SWE (mm)	SD (m)
Accumu-lation	+23.9	−7	−0.7	+0.3	+21	+5.6	+5.4	−0.2	+0.005	+52	+0.2
Melt	+35.2	−9.4	−0.4	+0.2	+12	+59.4	+5.3	−0.1	+0.03		

Note: Mean bias is the mean of the differences in accumulation or melt period combined for WY 2014–2019. Mean of SWE and SD differences at peak SWE date (dates shown in Table 4) in WY 2014–2019.

Table 4. Snow and Liquid Fractions of the Total Precipitation during the accumulation period at Peak SWE for water years 2014–2019.

		NLDAS-2							Station					
WY	Peak SWE (mm-dd hh)	Sim. Peak SWE (mm)	Total Snow (mm)	Total Rain (mm)	Total PRCP (mm)	% Snow	% Rain	Peak SWE (mm-dd hh)	Sim. Peak SWE (mm)	Total Snow (mm)	Total Rain (mm)	Total PRCP (mm)	% Snow	% Rain
2014	04-06 09	233	216	13.2	229	94.2	5.8	04-06 07	187	202	52.4	254	79.4	20.6
2015	03-23 09	234	232	14.0	246	94.3	5.7	04-11 00	208	274	40.4	315	87.2	12.8
2016	03-29 14	335	327	44.2	371	88	12	03-29 14	250	286	67	353	81	19
2017	03-29 03	316	292	13.5	306	95.6	4.4	03-29 02	280	286	38	325	88	12
2018	04-07 21	300	336	29.4	370	92	8	03-15 20	251	287	38	323	88.2	11.8
2019	03-23 12	373	316	18	334	94.6	5.4	03-22 23	305	282	59	342	82.7	17.3

The T2m in NLDAS-2 was significantly lower than the Station for all water years (Figure 4a), with a negative mean bias of −0.7 K and −0.4 K in NLDAS-2 during the accumulation and melt periods, respectively (Table 3). Total precipitation was higher in NLDAS-2 (compared to the Station) for all water years except WY 2015 (Figure 4b). However, NLDAS-2 total precipitation from the onset of the snow season to peak SWE was lower in WYs 2014, 2015, 2017, and 2019, and higher in WYs 2016 and 2018 (Table 4).
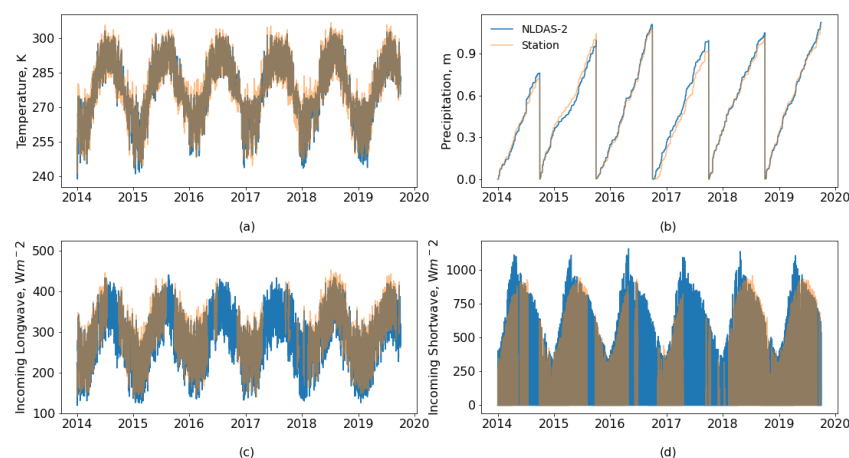

Figure 4. Comparison of (**a**) 2m Air Temperature (T2m); (**b**) Cumulative Precipitation; (**c**) Incoming Longwave Radiation; (**d**) Incoming Shortwave Radiation, forcing from NLDAS-2 versus those observed at the Station from WY2014–WY2019 (Note: days where Station values were missing are left blank; NLDAS-2 forcing was used for the Station-Noah-MP simulation for the days that the Station forcing was missing).

The negative bias in T2m in NLDAS-2 was associated with higher SWE and SD accumulations in the NLDAS2-Noah-MP model. Depending on the air temperature, precipitation could fall as either rain or snow. In Noah-MP, the precipitation partitioning scheme depends on T2m. Per Jordan's 1991 parametrization in Noah-MP, precipitation is:

Snow only, when T2m ≤ 273.66 K
Rain only, when T2m ≥ 275.66 K
Mix of rain and snow, when T2m is between 273.66–275.66 K with,

snow fraction = 1 − (− 54.632 + 0.2 × T2m) at T2m ≤ 275.16 K
snow fraction = 0.6 at T2m > 275.16 K

The analysis of the total snow and rainfall at peak SWE (Table 4) revealed that NLDAS-2 generally accumulated more snow in all years besides WY2015. In 2015, NLDAS-2 had less total precipitation but had a higher snow fraction. The Station had a higher liquid fraction of the total precipitation for all water years, whereas NLDAS-2 experienced a larger fraction of snow than at the station due to its cold bias. The difference in peak SWE was between 26–85 mm, 52 mm on average, for WY2014–2019. The peak SWE was between 23 March–7 April and 15 March–11 April in NLDAS2-Noah-MP and Station-Noah-MP, respectively. Total precipitation was higher at the Station for all water years except WY2016 and 2018.

In addition to snowfall, peak SWE is influenced by accumulation from refrozen rain and ablation through sublimation, melt, or evaporation. Snow loss can happen from sublimation or melt depending on the net amount of energy on the snow's surface (Equation (7)) and other environmental factors—relative humidity, wind, and surface pressure. Ablation processes during the accumulation season could be responsible for the lower peak SWE than total snow in the Station simulation. Interestingly, in all water years, the total snowfall was less than the peak SWE in the NLDAS-2 simulation (besides WY 2018), and the total snowfall was also less than the peak SWE in the Station simulation in WY 2019 (Table 4). This is likely explained by rain falling on the snowpack and refreezing within the snowpack. The refreezing of rain is possible during low-intensity rain over a long duration in deep and cold snowpack. The liquid water percolation and holding capacity of snowpack allows the snowpack to retain the rainwater long enough for it to refreeze as the temperature drops below freezing [43].

Snow ablation depends on the net total available energy at the snow's surface, which is defined as:

$$\text{Total Energy} = Q_{SW} + Q_{LW} + Q_L + Q_H + Q_G + Q_{RAIN} \tag{7}$$

where Q_{SW} is absorbed SW radiation, Q_{LW} is net LW, Q_G is conductive ground heat flux, Q_{RAIN} is sensible heat supplied by rainfall, and Q_H and Q_L are the turbulent heat exchange by sensible and latent heat fluxes, respectively. Q_G is the heat flux from the ground, and Q_{RAIN} is the heat added to the snowpack from the rain itself. Both Q_G and Q_{RAIN} were not considered in the calculation of total energy. Positive amounts of total net energy at the snow's surface will raise the snow's surface temperature, while negative values will cool the snow's surface temperature. Once the snow's surface temperature reaches an upper limit of 273.15 K (0 °C), any excess energy will result in snowmelt.

Solar radiation, strong winds, dry air, and low air pressure are the precursors for sublimation. The loss in SWE due to potential sublimation is:

$$\Delta\text{SWE}_{Sublimation} = \frac{LH}{\rho \times Ls} \tag{8}$$

where *LH* is the available latent heat, *Ls* (2,838,000 j/kg) is the latent heat of sublimation, and ρ (917 kg m^3) is the density of ice.

SWE differences in the two simulations were due to the differences in snowfall as well as the differences in available energy for ablation. For instance, in WY2014, the 46 mm (233 − 187) difference in peak SWE could be explained by a 14 mm (216 − 202) difference in snowfall, while the remaining 32 mm difference (Table 4) would be related to the differences in the sublimation and melt. In order to better understand the interplay between snow accumulation and ablation during the accumulation period in Noah-MP, we further analyzed two one-week time periods in 24–30 November and 19–25 December in 2018 (Figures 5 and 6) where deviations in snow accumulations were observed between the two simulations.
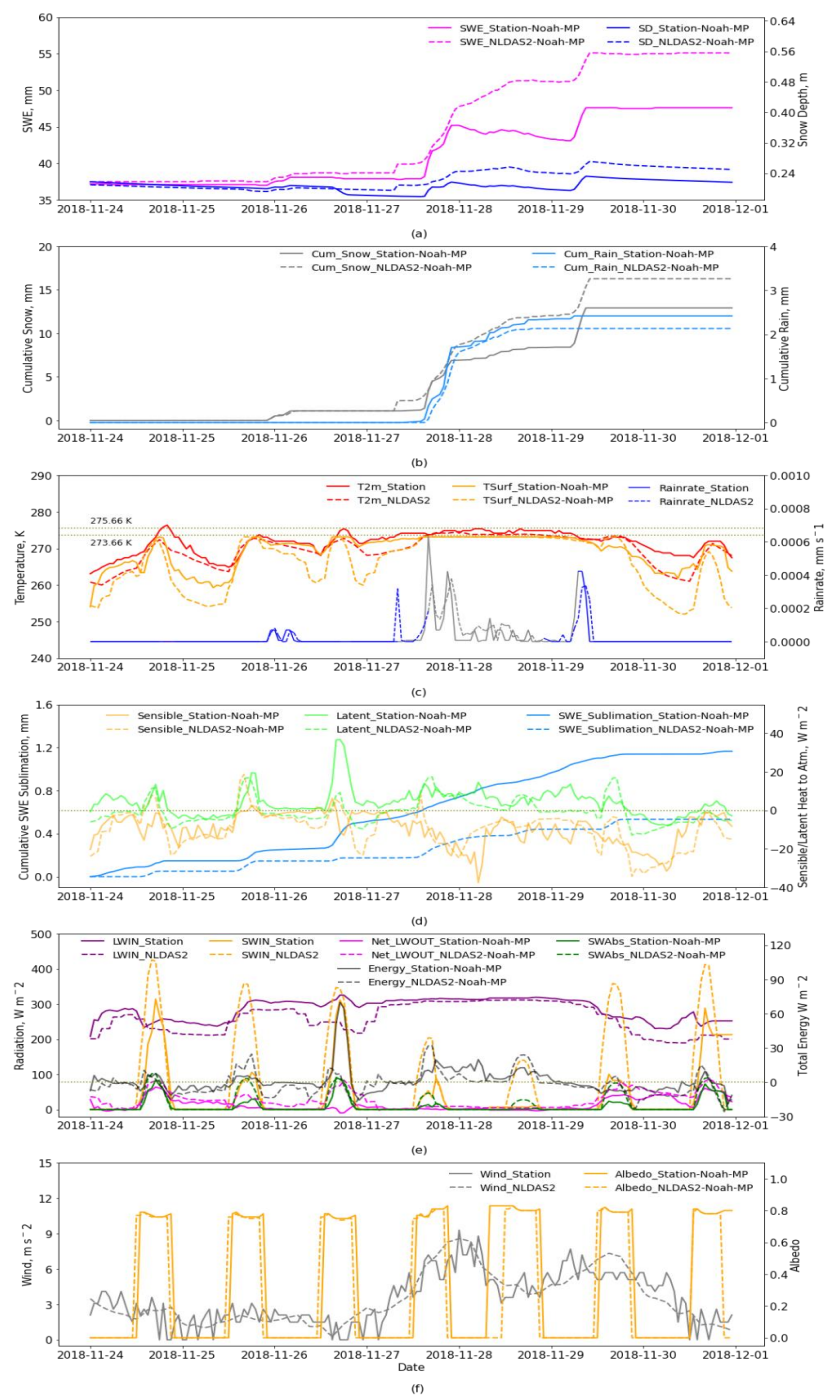
Figure 5. (**a**) Snow Water Equivalent and Snow Depth; (**b**) Cumulative Snow and Rain; (**c**) 2 m air Temperature (T2m), Surface Radiative Temperature (TRAD), and Rain-rate (gray color indicating mixed precipitation and blue color indicating snowfall); (**d**) Sensible and Latent Heat to the Atmosphere and Cumulative Snow/Ice Sublimation; (**e**) Incoming Longwave and Shortwave Radiation, Absorbed Shortwave Radiation, Total Net Outgoing Longwave Radiation, and Total Energy; and (**f**) Surface Albedo and Wind Speed, from 24−30 November 2018.
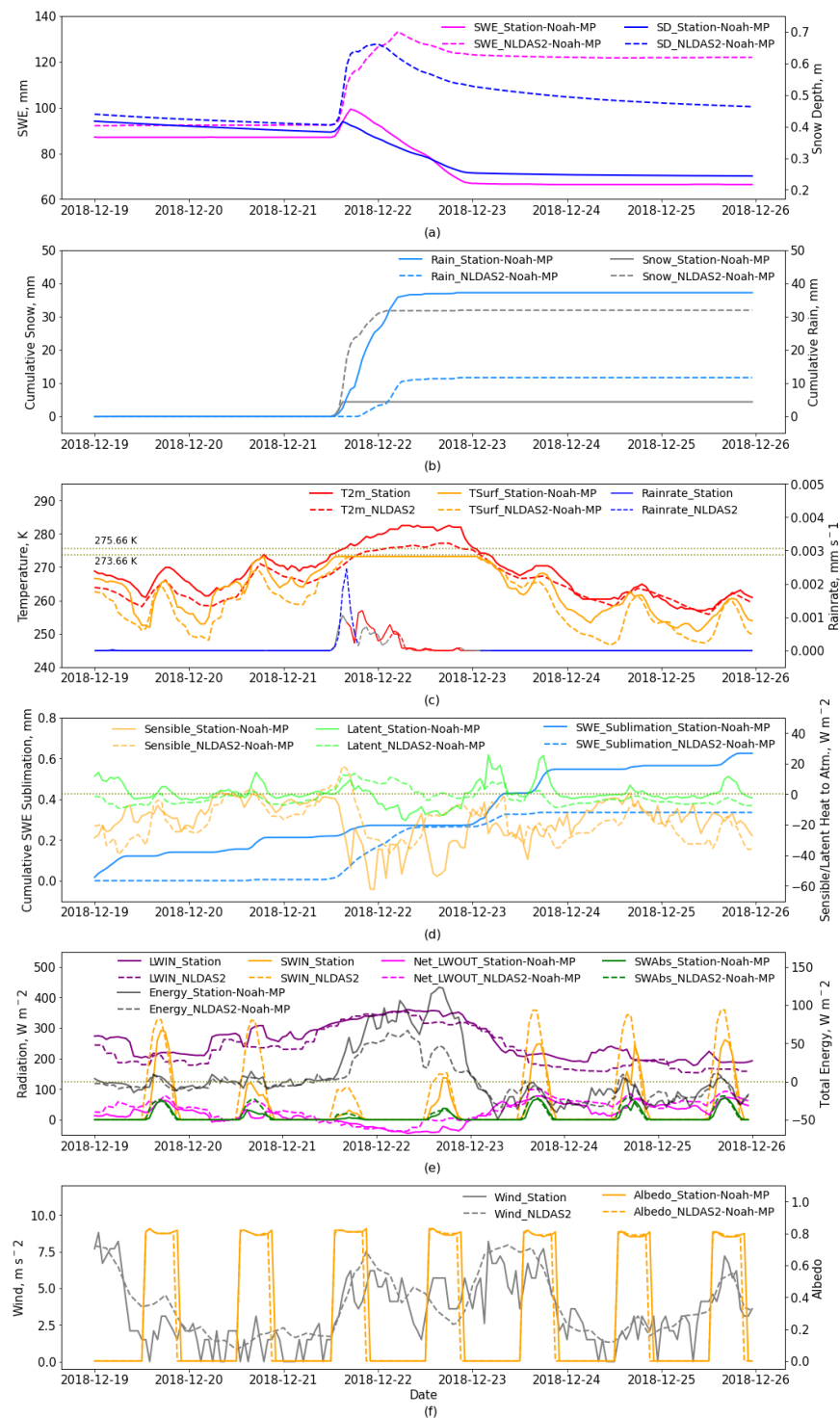
Figure 6. (**a**) Snow Water Equivalent and Snow Depth; (**b**) Cumulative Snow and Rain; (**c**) 2 m air Temperature, Surface Radiative Temperature (TRAD), and Rain-rate (the blue color indicates snowfall, gray color indicates mixed precipitation, and red color indicates rain); (**d**) Sensible and Latent Heat to the Atmosphere and Cumulative Snow/Ice Sublimation; (**e**) Incoming Longwave and Shortwave Radiation, Absorbed Shortwave Radiation, Total Net Outgoing Longwave Radiation, and Total Energy; and (**f**) Surface Albedo and Wind Speed, from 19−25 December 2018.

The precipitation event on 27 November brought rain and snow that increased SWE and SD in both simulations, with a greater increase in NLDAS2-Noah-MP compared to the Station-Noah-MP simulation (Figure 5a,c). Right before 28 November, SWE started to

decrease in the Station-Noah-MP simulation but continued to accumulate in the NLDAS2-Noah-MP simulation (Figure 5a). Greater snow accumulation in NLDAS2-Noah-MP was due to the lower T2m and the higher snow fraction of the mixed precipitation during 28 November (Figure 5b,c). Higher amounts of net energy at the snow's surface in Station-Noah-MP was responsible for the decrease in SWE during 28 November. In the Station-Noah-MP simulation, there was a higher sensible heat flux directed towards the snowpack, which was the predominant reason for the higher net energy. Net LW (also the incoming LW) did not have any major contributions to the differences in the net energy as they were similar in the two simulations. Differences in the absorbed SW (also the incoming SW) had some contribution to the net energy at the snow's surface during the day, shown by the higher values in NLDAS2-Noah-MP and the corresponding increases in the net energy, but this did not have a major effect on the snow (or its loss) in the simulation. Differences in the absorbed SW and net LW were also indicated in Table 3. While there were significant differences in the incoming LW and SW radiations for the water years (Figure 4c,d), with positive bias in SW and negative bias in LW in NLDAS-2 during the accumulation and melt periods (Table 3), radiation biases counterbalanced each other in energy budgets. Higher sensible heat fluxes for the November event were the primary reason for snow melt on November 28. Higher sensible heat fluxes resulted primarily from a higher T2m value, as both simulations had a snow-surface temperature of 273.15 K, and both NLDAS-2 and the Station had similar wind speeds. Furthermore, we note during this period, sublimation amounts were negligible (<0.4 mm).

Similarly, during the precipitation event on 21 December (Figure 6a), there was a higher amount of snow accumulation in NLDAS-Noah-MP compared to Station-Noah-MP. Just prior to 22 December, there was a prominent deviation in the SWE and SD between the two simulations. SWE started to decrease in the Station-Noah-MP simulation but continued to accumulate in NLDAS2-Noah-MP (Figure 6a). Greater snow accumulation in NLDAS2-Noah-MP was because of the higher snow fraction due to the lower T2m during 22 December (Figure 6b,c). Station-Noah-MP had a larger rain fraction as reflected from the color red in the rain rate line plot (Figure 6b,c) as T2m at the Station rose above 275.66 K. The reduction in SWE and SD in Station-Noah-MP was mainly due to snow melt from the available net energy at the snow's surface, which was considerably higher in Station-Noah-MP compared to NLDAS2-Noah-MP (Figure 6e). The higher net energy in Station-Noah-MP is due to the corresponding higher sensible and latent heat fluxes directed towards the snow's surface. Similar to the November case, higher sensible heat fluxes resulted primarily from a higher T2m value, as both simulations had a snow-surface temperature at 273.15 K, and both NLDAS-2 and the Station had similar wind speeds. In the Station-Noah-MP simulation, there was an increase in latent heat on the snow's surface, which was from condensation of moisture at the snow's surface (Figure 6d). Interestingly, while there was condensation in the Station-Noah-MP simulation, there was a sublimation event in the NLDAS-Noah-MP simulation, but the amount of sublimation was negligible (~0.2 mm) on 22 December (Figure 6d,f).

4.1.2. Comparison between Simulated and Observed SWE

For WY2019, when the measurements of observed SWE were available, the measured SWE was higher than those simulated with both the NLDAS2-Noah-MP and Station-Noah-MP models (Figure 7). However, there was a general match in the timing of accumulation and melt events in the simulations and observations.
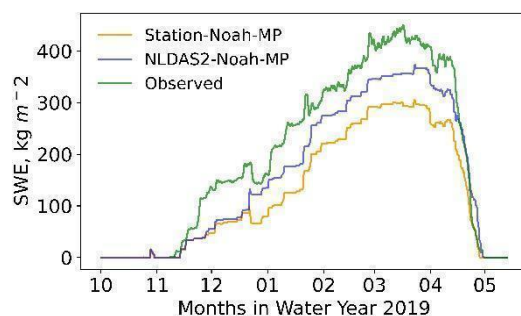
Figure 7. Simulated Snow Water Equivalent from Noah-MP Model and Station-Noah-MP versus the Observed values from WY2019.

Several possible reasons exist for the SWE differences between the two simulations and the observations, including: the inaccuracy of the forcing in NLDAS-2, instrumentation errors at the Station, inadequate representation of snow processes in the model, and wind-driven snow in the observations. The largest difference in SWE was between the observed and Station-Noah-MP. This was unexpected since the same Station forcing that was used to simulate the Station-Noah-MP influenced the Observed SWE. This could be due to observational uncertainties such as precipitation gauges under the catch lowering SWE in Station-Noah-MP, or the gamma sensor overestimating SWE in the observations, or the accumulation of wind-blown snow under the gamma sensor. However, since the WY2019's cumulative precipitation at the Station was higher than in NLDAS-2—with an 8 mm difference in total precipitation and a 68 mm difference in peak SWE (Table 4)—the differences in rain versus snow perhaps play a larger role here (Figures 5 and 6). Testing the possible reasons in further detail were outside the scope of this study and were not further investigated. Instead, we analyzed an instance where the simulations and observations deviated from one another, in order to understand the factors contributing to the differences.

In mid-November WY2019, there was a visible deviation in observed SWE from the NLDAS2-Noah-MP and Station-Noah-MP simulations (Figure 7). The increase in observed SWE and the resulting differences carried on through the rest of the accumulation period. We investigated this period to further understand the possible causes for the separation. There was a prominent increase in observed SWE in 23–24 November, although there was no accompanying precipitation event, whereas the simulated SWE remained steady (Figure 8a,b). T2m was below 273.66 K during the time in all cases, but the wind speed increased by approximately 9 m/s at the Station (Figure 8c). The results suggest that the wind-blown re-deposition of snow might have caused the SWE measurement to increase at the Station during this period, which is probable given the location of the study site. The site is open to the north and west and more enclosed to the east and south (Figure 1b). The site could be prone to wind-blown snow deposition, as the east and south areas could have acted as a windbreak to the wind blowing from the northerly and easterly directions. The wind-driven deposition of snow, which is a process not accounted for by Noah-MP, could have a prominent effect at a point location, such as in this single-column study. Other studies have also suggested the possibility of the heterogeneous distribution of snow due to wind-induced erosion and transport that is missed by land surface models [44,45], especially at a sub-grid scale (<1 km) in non-forested areas where wind drift often dominates the local snow distribution pattern [46].
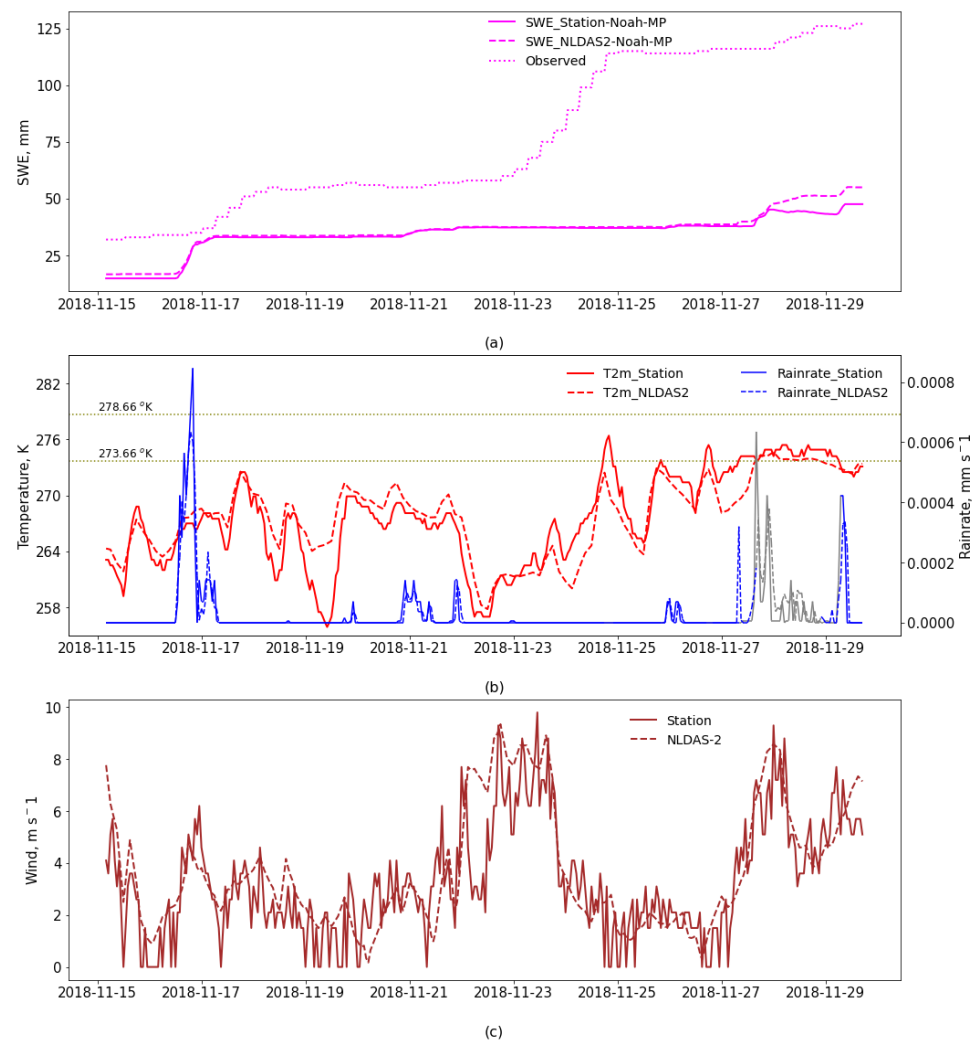
Figure 8. (**a**) Snow Water Equivalent; (**b**) T2m and Rain-rate (blue color indicates snowfall, gray color indicates mixed precipitation, and red color indicates rain (not in this plot)); (**c**) Wind Speed, in 15–29 November 2018.

4.2. Streamflow Prediction

Comparisons of the simulated runoff from the WRF-Hydro model with the observed streamflow at the four USGS gage locations in WY 2014–2019 suggested that the model simulated runoff considerably well at two out of the four USGS gage locations in the study area (Figure 9). The model's accuracy compared to the observed values are shown in Table 5. At the Washburn and Masardis stations, the simulated streamflow represents the observed values well, as shown by NSE > 0.5, RSR < 0.70, and KGE closer to 1. Note that the Masardis sub-basin is part of the Washburn sub-basin. At the Hardwood and Madawaska stations, the indices NSE < 0.5, RSR > 0.7, and KGE further from 1 indicate a poorer representation of the observed streamflow by the simulated runoff. Although PBIAS was within $\pm$ 25% for all stations, Madawaska station had the highest positive PBIAS, indicating some model underestimation, and Hardwood station had a negative PBIAS, indicating model overestimation; the model slightly overestimated runoff at Masardis and Washburn stations. Although the NSE and RSR for Madawaska was slightly better than Hardwood, the higher KGE in Hardwood Brook indicates a better simulation. The higher KGE came from better r-value (correlation with the observed values) and alpha value (ratio of the standard deviation between the simulated to observed closer to 1).
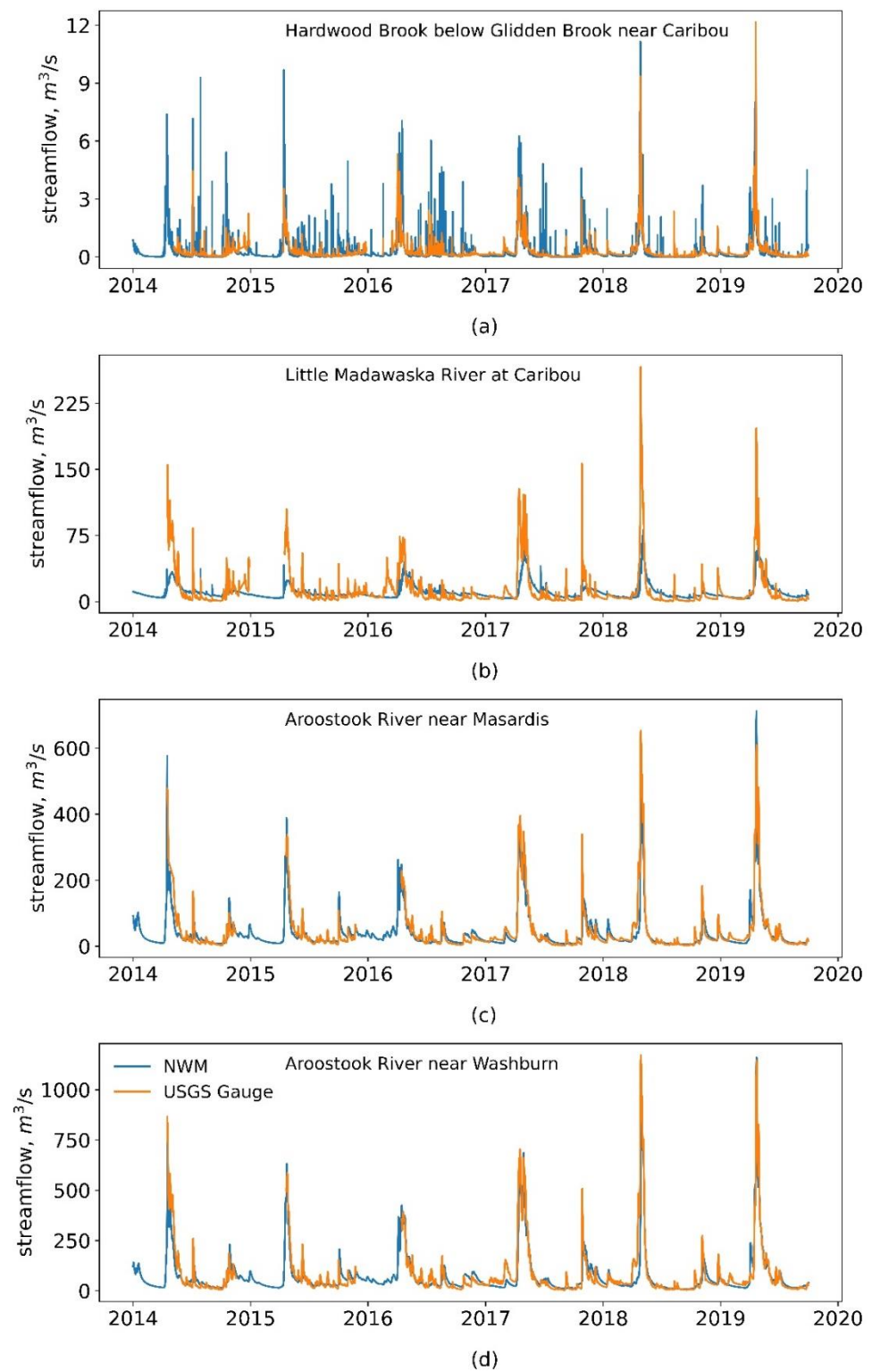
Figure 9. Comparison of the runoff from NWM V2.0 and the streamflow from USGS gages at (**a**) Harwood Brook; (**b**) Madawaska; (**c**) Masardis; (**d**) Washburn stations from WY2014-WY2019.

Table 5. NSE, PBIAS, RSR, and KGE for simulated runoff compared with the observed streamflow at the four USGS gage locations: Hardwood Brook, Madawaska, Masardis, and Washburn stations.

Index	Hardwood Brook	Madawaska	Masardis	Washburn
NSE	0.40	0.42	0.90	0.90
PBIAS%	−9.02	10.21	6.80	3.11
RSR	0.77	0.75	0.30	0.30
KGE	0.63	0.33	0.86	0.87
r	0.81	0.74	0.96	0.96
Alpha	1.3	0.39	0.89	0.89
Beta	1.1	0.9	0.93	0.97

It was interesting that the simulated hydrographs at Hardwood Brook and Little Madawaska showed differing response compared to its observed hydrographs. Simulated hydrographs at Hardwood Brook showed flow peaks at possibly every rain event, whereas they were muted at Little Madawaska. More precipitation was likely being intercepted by the canopy and soil at Hardwood Brook than what the model simulates, whereas there was more runoff at Little Madawaska. Although the reasons for such varied responses need further investigation, this emphasizes the difficulties of numerical modelling in accurately representing the land and soil properties and precipitation estimates in smaller basins covered by fewer model grids.

While the volume was over- or underestimated, the model seems to have captured the timing of the peaks during the spring snowmelt rather well. The Hardwood and Madawaska sub-basins, where the model was least accurate, are smaller in area and have a higher percentage of cultivated, urban, and pasture land-uses compared to the corresponding sub-basins of Masardis and Washburn stations, which are bigger in area and have a higher percentage of forested areas. The basin and land-use likely play an important role in influencing the runoff and streamflow at the basin outlet.

We also examined simulated estimates of the fractional snow cover (fSCA), SWE, and SD in the four sub-basins to further contextualize the streamflow results for snow accumulation and melt seasons from 2015–2019 (Figure 10). The evolution of the simulated SWE and SD estimates were visible for all sub-basins and years, with a gradual increase during the accumulation period and a relatively faster decline in the melt period, whereas simulated fSCA rose quickly to a maximum value early on during the accumulation period and declined sharply during the melting period. Simulated fSCA showed little variation during the accumulation period—this is because once the ground is fully covered by snow, fSCA in Noah-MP reaches a maximum value after about 5 cm of snow depth [47]. Parameterization in Noah-MP does not produce temporal variations in grid-scale snow cover fractions and the depletion occurs too rapidly—fSCA is nearly always classified as no snow or fully covered [48]. The binary pattern of fSCA in Noah-MP is a product of the snow depletion equation included in Noah-MP by Niu et al. [47] as a hyperbolic tangent function of SD and SWE, as shown in Equation (8):

$$\text{fSCA} = tanh\left(\frac{h_{sno}}{2.5 z_{0,g}\frac{(\rho_{sno}}{\rho_{new})}m}\right) \tag{9}$$

where h_{sno} is snow depth in meters, $z_{0,g}$ is the ground roughness length (equal to 0.01 m), ρ_{sno} is the snow density in kg m^{-3}, ρ_{new} is thee fresh snow density (equal to 100 kg m^{-3}), and m is a dimensionless melting factor.
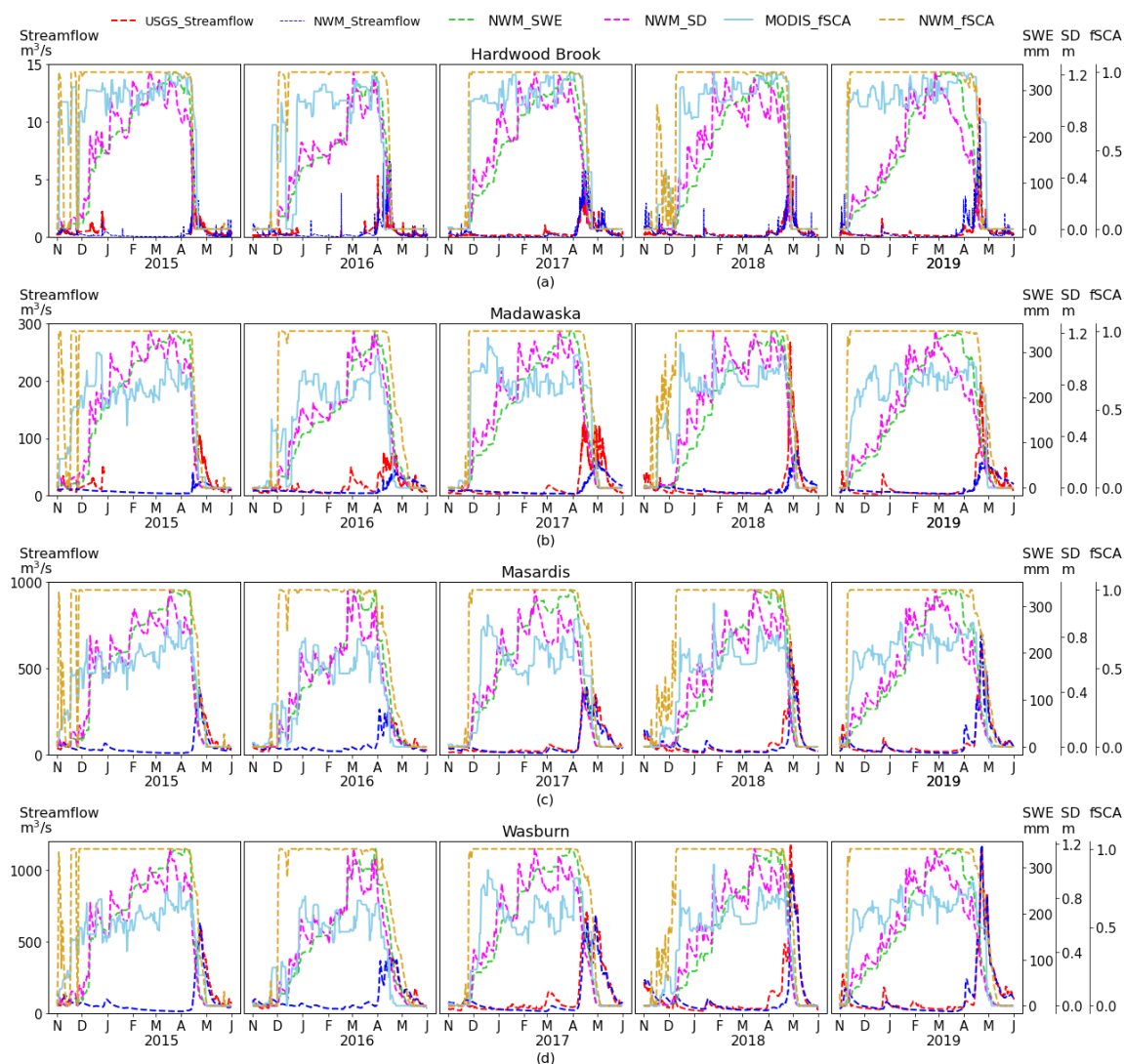
Figure 10. Runoff, fractional snow cover, snow water equivalent, and snow depth from NWM V2.0 estimates and streamflow from USGS gages at (**a**) Harwood Brook; (**b**) Madawaska; (**c**) Masardis; (**d**) Washburn stations from 2014–2019 snow accumulation and melt seasons.

For all water years in all four sub-basins, the decline in the simulated SD appeared to start prior to the decline in the simulated SWE, indicating further snow compaction during the melting period as the SWE would keep rising before it declines as the snow fully melts. Although increases in the streamflow was accompanied by simultaneous declines in fSCA, SD, and SWE, SWE and SD were earlier indicators of melting. The variability in SWE and SD was able to capture earlier runoff peaks during the spring melt, followed by changes in fSCA. In most cases, the decline in SWE and SD from their peaks matched the timing of the streamflow peaks during the spring snowmelt.

5. Discussion

The precipitation partitioning scheme in Noah-MP, based on the T2m, made a substantial difference in the simulated patterns of snow accumulation and melt in the single-column study. The negative bias in NLDAS-2 T2m compared to those observed at the Station increased the snow fraction, leading to a higher snow accumulation in the NLDAS2-Noah-MP model. From the analysis of snow and rain fractions at the peak SWE, NLDAS-2 had the higher snow fraction and Station had the higher rain fraction. The total precipitation itself was higher at the Station except for WYs 2016 and 2018 (Table 4). Although the

precipitation bias was not the dominant factor contributing to the differences in SWE and SD in this study, the biases, if present along with the biases of T2m, would have a major impact on snow versus rain and thus in the simulated SWE and SD. A study of NWM reanalysis snow outputs for observed SWE in the western US showed that the errors in precipitation inputs were more important than the cooler temperature bias, as the NWM generally underpredicted the SWE [20]. This shows that the contribution of inputs in the NWM simulation might have regional differences. Previous work has shown that wet-bulb temperature precipitation partitioning schemes to partition precipitation, or directly using outputs from a numerical weather models' microphysical scheme (e.g., WRF, HRRR), can improve the snowfall predictions [48,49]. The wet-bulb temperature, which incorporates the atmospheric humidity, better reflects the hydrometeor's temperature, which is often cooler than the air temperature due to evaporation of the hydrometeor as it falls. Microphysical schemes parameterize the formation, growth, and fall out of precipitation from clouds, but require running computationally expensive atmospheric models.

The analysis of two diverging events during WY 2019 also revealed that differences in snow accumulation between the two simulations were due to the precipitation phase differences resulting from the differences in the T2m. Additionally, the differences in snow accumulation were due to the differences in the snow melt resulting from the differences in the net energy at the snow's surface. The net energy differences were mainly due to the differences in the sensible heat flux, as the differences in the absorbed SW and net LW between the two simulations were small. The differences in the T2m resulted in different amounts of sensible heat in the simulations, which translated into more melt in the Station-Noah-MP simulation because during the diverging events, wind speeds were similar and both had a snow-surface temperature near the melting/freezing point (273.15 K).

We note that while Noah-MP's simulation of fSCA was too high relative to observed fSCA, as also noted in other studies [20,50], more realistically simulating fSCA might degrade the simulation of SWE since the surface albedo, rather than the snow albedo, is used to calculate the energy balance at the snow's surface. Per Noah-MP's parameterization:

$$\text{Surface Albedo} = \text{Soil Albedo} \, (1 - \text{fSCA}) + \text{Snow Albedo} \, (\text{fSCA}) \tag{10}$$

where Snow Albedo = f f(snow age, zenith angle)

Noah-MP's energy budget calculation would be more accurate if Noah-MP calculated an energy balance for snow-covered and non-snow-covered surfaces separately and then averaged the resulting fluxes of energy and mass within the grid cell, rather than its current method of using a grid cell average albedo (Equation (9)). We note that the timing of the accumulation and melt was similar between the NLDAS2-Noah-MP and Station-Noah-MP (Figures 3 and 7). As a result, the snow age—and therefore the snow albedo—was also similar in the two simulations, which caused similar surface albedos in the accumulation season for two simulations (Table 3, Figures 6e and 7e).

In the basin study, we suspect that the basin sizes and land-uses played a key role in the simulation of runoff and streamflow in this study. The retrospective NWM V2.0 was able to simulate streamflow with a greater accuracy at the Masardis and Washburn sub-basins, which had a higher fraction of forested areas. The streamflow simulation was less accurate in the Harwood and Madawaska sub-basins, which were both smaller in areas with higher percentages of agricultural land-use. The accuracy of the vegetation types and soil properties were not further investigated for our study basins.

Smaller basins, due to shorter hydrologic response times and fewer grid cells, allow less room for error in simulating runoff. So, the accurate model inputs for atmospheric forcing and/or basin characteristics and the proper representation of physical processes occurring in the basin becomes especially critical, as biases in the small number of nearby grid cells tend to be similar in magnitude or direction, whereas in larger basins, to a certain extent, the errors and biases acting in opposite directions have a chance to cancel out. This is partially due to the averaging effect in larger areas covered by more simulation grid cells and the longer time it takes for an error to propagate downstream to the basin outlet.

Spring snowmelt timing in the Masardis and Washburn basin simulations were captured well, with an NSE of 0.9 and KGE of 0.86–0.87, and coincided with the changes in SWE and SD. The declines in SWE and SD in the simulation gave an early indication of the spring snowmelt in the simulations in most cases. The simulated fSCA was less useful after peak snow cover was reached. On the other hand, the remotely sensed data from MODIS exhibited more temporal variability in fSCA, which could be useful information about the fSCA on the ground.

6. Conclusions

As snowpack accumulation and melting become more variable in a changing climate, the need to consider changes in snowpack variability for adaptive planning becomes crucial [51]. As in situ measurements of fSCA, SWE, and SD are sparse, and satellite remote sensing has its own challenges in terms of retrievals and/or resolution, accurate numerical snow simulations can be valuable to assess water availability from seasonal snowmelts. An outstanding challenge for numerical water predictions is understanding and reducing uncertainties in the meteorological forcing data and finding appropriate parameterizations to accurately represent the physical processes. We suspect that several factors collectively contributed to deficiencies in accurately representing surface energy budgets and, consequently, the snow accumulation in this simulation study. Specifically, the biases in the forcing data such as the underestimation of T2m (-0.7 K in the accumulation period) in NLDAS-2 allowed the Noah-MP precipitation partitioning scheme to fraction more snow than rain, leading to greater amounts of snow (+52 mm peak SWE on average) in NLDAS2-Noah-MP. On the other hand, the higher T2m at the Station led to higher amounts of net energy at the snow's surface, which led to more melting and less snow accumulation in Station-Noah-MP. We also noted that Noah-MP's energy budget calculation would be more accurate if it calculated an energy balance for snow-covered surfaces separately from the non-snow-covered surfaces and would improve the fSCA by achieving more realistic values as shown by MODIS. Correcting the forcing biases, especially in the T2m and precipitation, along with separate calculation of the energy balance in snow-covered areas, would help in the improvement of snow simulations in future developments of NWM or other numerical water prediction models.

Streamflow prediction has been one of the most important goals in hydrology. Factors contributing to the changes in streamflow are buffered in larger areas by a multitude of competing and self-canceling effects in larger watersheds, thus calling for greater attention to the careful study of smaller watersheds, as they would more likely bear the impacts of extreme events [20]. For example, the factors contributing to streamflow such as precipitation distribution and timing, soil conditions, or vegetation type would vary differently over a large area, and therefore could average out and have less immediate impact on the streamflow. It will therefore be important for process-based numerical models or future versions of NWM to focus on smaller basins where smaller-scale processes are critical to accurately simulating streamflow, especially in the basins with disturbed soil such as in agricultural lands.

Given the study's focus on the unique terrain and climate of the northeastern United States, future work should extend to other snow-dominated regions as well. This will provide insights into the factors leading to the differences in the snow and its effect on the seasonal snowmelt, which will be valuable for the future of numerical water prediction modeling.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The publicly archived meteorological and snow dataset from CUNY SAFE, Caribou, ME can be accessed from a web location [36]. The 'namelist.hrldas', 'hydro.namelist' and parameter files and tables can be accessed at a repository [33].

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Dawson, N.; Broxton, P.; Zeng, X. Evaluation of Remotely Sensed Snow Water Equivalent and Snow Cover Extent over the Contiguous United States. *J. Hydrometeorol.* **2018**, *19*, 1777–1791. [CrossRef]
2. Tomasi, E.; Giovannini, L.; Zardi, D.; de Franceschi, M. Optimization of Noah and Noah_MP WRF Land Surface Schemes in Snow-Melting Conditions over Complex Terrain. *Mon. Weather Rev.* **2017**, *145*, 4727–4745. [CrossRef]
3. Hall, A. The Role of Surface Albedo Feedback in Climate. *J. Clim.* **2004**, *17*, 1550–1568. [CrossRef]
4. Flanner, M.G.; Shell, K.M.; Barlage, M.; Perovich, D.K.; Tschudi, M.A. Radiative Forcing and Albedo Feedback from the Northern Hemisphere Cryosphere between 1979 and 2008. *Nat. Geosci.* **2011**, *4*, 151–155. [CrossRef]
5. Lawrence, D.M.; Slater, A.G. The Contribution of Snow Condition Trends Trends to Future Ground Climate. *Clim. Dyn.* **2010**, *34*, 969–981. [CrossRef]
6. Dudley, R.W.; Hodgkins, G.A.; McHale, M.R.; Kolian, M.J.; Renard, B. Trends in Snowmelt-Related Streamflow Timing in the Conterminous United States. *J. Hydrol.* **2017**, *547*, 208–221. [CrossRef]
7. Musselman, K.N.; Clark, M.P.; Liu, C.; Ikeda, K.; Rasmussen, R. Slower Snowmelt in a Warmer World. *Nat. Clim. Chang.* **2017**, *7*, 214–219. [CrossRef]
8. Clow, D.W. Changes in the Timing of Snowmelt and Streamflow in Colorado: A Response to Recent Warming. *J. Clim.* **2010**, *23*, 2293–2306. [CrossRef]
9. Barlage, M.; Chen, F.; Tewari, M.; Ikeda, K.; Gochis, D.; Dudhia, J.; Rasmussen, R.; Livneh, B.; Ek, M.; Mitchell, K. Noah Land Surface Model Modifications to Improve Snowpack Prediction in the Colorado Rocky Mountains. *J. Geophys. Res. Atmos.* **2010**, *115*. [CrossRef]
10. Wayand, N.E.; Hamlet, A.F.; Hughes, M.; Feld, S.I.; Lundquist, J.D. Intercomparison of Meteorological Forcing Data from Empirical and Mesoscale Model Sources in the North Fork American River Basin in Northern Sierra Nevada, California. *J. Hydrometeorol.* **2013**, *14*, 677–699. [CrossRef]
11. Viterbo, F.; Mahoney, K.; Read, L.; Salas, F.; Bates, B.; Elliott, J.; Cosgrove, B.; Dugger, A.; Gochis, D.; Cifelli, R. A Multiscale, Hydrometeorological Forecast Evaluation of National Water Model Forecasts of the May 2018 Ellicott City, Maryland, Flood. *J. Hydrometeorol.* **2020**, *21*, 475–499. [CrossRef]
12. Lahmers, T.M.; Hazenberg, P.; Gupta, H.; Castro, C.; Gochis, D.; Dugger, A.; Yates, D.; Read, L.; Karsten, L.; Wang, Y.-H. Evaluation of NOAA National Water Model Parameter Calibration in Semi-Arid Environments Prone to Channel Infiltration. *J. Hydrometeorol.* **2021**, *22*, 2939–2969. [CrossRef]
13. Verri, G.; Pinardi, N.; Gochis, D.; Tribbia, J.; Navarra, A.; Coppini, G.; Vukicevic, T. A Meteo-Hydrological Modelling System for the Reconstruction of River Runoff: The Case of the Ofanto River Catchment. *Nat. Hazards Earth Syst. Sci.* **2017**, *17*, 1741–1761. [CrossRef]
14. Powers, J.G.; Klemp, J.B.; Skamarock, W.C.; Davis, C.A.; Dudhia, J.; Gill, D.O.; Coen, J.L.; Gochis, D.J.; Ahmadov, R.; Peckham, S.E.; et al. The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1717–1737. [CrossRef]
15. Salas, F.R.; Somos-Valenzuela, M.A.; Dugger, A.; Maidment, D.R.; Gochis, D.J.; David, C.H.; Yu, W.; Ding, D.; Clark, E.P.; Noman, N. Towards Real-Time Continental Scale Streamflow Simulation in Continuous and Discrete Space. *J. Am. Water Resour. Assoc.* **2018**, *54*, 7–27. [CrossRef]

16. Barnett, T.P.; Pierce, D.W.; Hidalgo, H.G.; Bonfils, C.; Santer, B.D.; Das, T.; Bala, G.; Wood, A.W.; Nozawa, T.; Mirin, A.A.; et al. Human-Induced Changes in the Hydrology of the Western United States. *Science* **2008**, *319*, 1080–1083. [CrossRef]

17. Holtzman, N.M.; Pavelsky, T.M.; Cohen, J.S.; Wrzesien, M.L.; Herman, J.D. Tailoring WRF and Noah-MP to Improve Process Representation of Sierra Nevada Runoff: Diagnostic Evaluation and Applications. *J. Adv. Modeling Earth Syst.* **2020**, *12*. [CrossRef]

18. Zeng, X.; Broxton, P.; Dawson, N. Snowpack Change From 1982 to 2016 Over Conterminous United States. *Geophys. Res. Lett.* **2018**, *45*, 12940–12947. [CrossRef]

19. Kampf, S.K.; Lefsky, M.A. Transition of Dominant Peak Flow Source from Snowmelt to Rainfall along the Colorado Front Range: Historical Patterns, Trends, and Lessons from the 2013 Colorado Front Range Floods. *Water Resour. Res.* **2016**, *52*, 407–422. [CrossRef]

20. Garousi-Nejad, I.; Tarboton, D.G. A Comparison of National Water Model Retrospective Analysis Snow Outputs at Snow Telemetry Sites across the Western United States. *Hydrol. Processes* **2022**, *36*, e14469. [CrossRef]

21. Somos-Valenzuela, M.A.; Palmer, R.N. Use of WRF-Hydro over the Northeast of the US to Estimate Water Budget Tendencies in Small Watersheds. *Water* **2018**, *10*, 1709. [CrossRef]

22. Chezik, K.A.; Anderson, S.C.; Moore, J.W. River Networks Dampen Long-Term Hydrological Signals of Climate Change. *Geophys. Res. Lett.* **2017**, *44*, 7256–7264. [CrossRef]

23. Painter, T.H.; Rittger, K.; McKenzie, C.; Slaughter, P.; Davis, R.E.; Dozier, J. Retrieval of Subpixel Snow Covered Area, Grain Size, and Albedo from MODIS. *Remote Sens. Environ.* **2009**, *113*, 868–879. [CrossRef]

24. Bennett, K.E.; Cherry, J.E.; Balk, B.; Lindsey, S. Using MODIS Estimates of Fractional Snow Cover Area to Improve Streamflow Forecasts in Interior Alaska. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 2439–2459. [CrossRef]

25. Dong, C. Remote Sensing, Hydrological Modeling and in Situ Observations in Snow Cover Research: A Review. *J. Hydrol.* **2018**, *561*, 573–583. [CrossRef]

26. Bales, R.C.; Molotch, N.P.; Painter, T.H.; Dettinger, M.D.; Rice, R.; Dozier, J. Mountain Hydrology of the Western United States. *Water Resour. Res.* **2006**, *42*. [CrossRef]

27. Dozier, J.; Painter, T.H.; Rittger, K.; Frew, J.E. Time-Space Continuity of Daily Maps of Fractional Snow Cover and Albedo from MODIS. *Adv. Water Resour.* **2008**, *31*, 1515–1526. [CrossRef]

28. Hall, D.K.; Riggs, G.A. *MODIS/Terra CGF Snow Cover Daily L3 Global 500 m SIN Grid, Version 61 USER GUIDE*; NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA, 2021. [CrossRef]

29. Niu, G.Y.; Yang, Z.L.; Mitchell, K.E.; Chen, F.; Ek, M.B.; Barlage, M.; Kumar, A.; Manning, K.; Niyogi, D.; Rosero, E.; et al. The Community Noah Land Surface Model with Multiparameterization Options (Noah-MP): 1. Model Description and Evaluation with Local-Scale Measurements. *J. Geophys. Res. Atmos.* **2011**, *116*. [CrossRef]

30. NLCD 2016 Land Cover (CONUS). Available online: https://www.mrlc.gov/data/nlcd-2016-land-cover-conus (accessed on 23 April 2021).

31. Gochis, D.J.; Barlage, M.; Cabell, R.; Casali, M.; Dugger, A.; Fitzgerald, K.; Mcallister, M.; Mccreight, J.; Rafieeinasab, A.; Read, L.; et al. *The WRF-Hydro® Modeling System Description, (Version 5.1.1)*; NCAR Technical Note: Boulder, CO, USA, 2020.

32. Yang, Z.L.; Niu, G.Y.; Mitchell, K.E.; Chen, F.; Ek, M.B.; Barlage, M.; Longuevergne, L.; Manning, K.; Niyogi, D.; Tewari, M.; et al. The Community Noah Land Surface Model with Multiparameterization Options (Noah-MP): 2. Evaluation over Global River Basins. *J. Geophys. Res. Atmos.* **2011**, *116*, 1–16. [CrossRef]

33. Gochis, D.J.; Barlage, M.; Cabell, R.; Casali, M.; Dugger, A.; Fanfarillo, A.; FitzGerald, K.; McAllister, M.; McCreight, J.; RafieeiNasab, A.; et al. WRF-Hydro Model Code. Available online: https://zenodo.org/record/3625238#.YsNpaHbMK38 or https://github.com/NCAR/wrf_hydro_nwm_public/tree/v5.1.1 (accessed on 14 March 2020).

34. 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.; et al. 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.* **2004**, *109*. [CrossRef]

35. Xia, Y.; Mitchell, K.; Ek, M.; Sheffield, J.; Cosgrove, B.; Wood, E.; Luo, L.; Alonge, C.; Wei, H.; Meng, J.; et al. Continental-Scale Water and Energy Flux Analysis and Validation for the North American Land Data Assimilation System Project Phase 2 (NLDAS-2): 1. Intercomparison and Application of Model Products. *J. Geophys. Res. Atmos.* **2012**, *117*. [CrossRef]

36. CUNY Snow Analysis and Field Experiment. Available online: https://www.star.nesdis.noaa.gov/smcd/emb/snow/caribou/microwave.html (accessed on 31 August 2019).

37. Nadolski, V.L. *Automated Surface Observing System User's Guide*; National Weather Service: Silver Spring, MD, USA, 1998.

38. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 2nd ed.; Academic Press: Amsterdam, The Netherlands, 2006.

39. Moriasi, D.N.; Arnold, J.G.; van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans ASABE* **2007**, *50*, 885–900. [CrossRef]

40. Riggs, R.; Hall, D.; Roman, M.O. *VIIRS Snow Cover Algorithm Theoretical Basis Document (ATBD), Version 1.0*; Nasa Goddard Space Flight Center: Greenbelt, MD, USA, 2015.

41. Salomonson, V.V.; Appel, I. Development of the Aqua MODIS NDSI Fractional Snow Cover Algorithm and Validation Results. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1747–1756. [CrossRef]

42. Raleigh, M.S.; Rittger, K.; Moore, C.E.; Henn, B.; Lutz, J.A.; Lundquist, J.D. Ground-Based Testing of MODIS Fractional Snow Cover in Subalpine Meadows and Forests of the Sierra Nevada. *Remote Sens. Environ.* **2013**, *128*, 44–57. [CrossRef]

43. Heilig, A.; Mitterer, C.; Schmid, L.; Wever, N.; Schweizer, J.; Marshall, H.P.; Eisen, O. Seasonal and Diurnal Cycles of Liquid Water in Snow—Measurements and Modeling. *J. Geophys. Res. Earth Surf.* **2015**, *120*, 2139–2154. [CrossRef]

44. Toure, A.M.; Reichle, R.H.; Forman, B.A.; Getirana, A.; de Lannoy, G.J.M. Assimilation of MODIS Snow Cover Fraction Observations into the NASA Catchment Land Surface Model. *Remote Sens.* **2018**, *10*, 316. [CrossRef]

45. Aas, K.S.; Gisnås, K.; Westermann, S.; Berntsen, T.K. A Tiling Approach to Represent Subgrid Snow Variability in Coupled Land Surface-Atmosphere Models. *J. Hydrometeorol.* **2017**, *18*, 49–63. [CrossRef]

46. Clark, M.P.; Hendrikx, J.; Slater, A.G.; Kavetski, D.; Anderson, B.; Cullen, N.J.; Kerr, T.; Örn Hreinsson, E.; Woods, R.A. Representing Spatial Variability of Snow Water Equivalent in Hydrologic and Land-Surface Models: A Review. *Water Resour. Res.* **2011**, *47*. [CrossRef]

47. Niu, G.Y.; Yang, Z.L. An Observation-Based Formulation of Snow Cover Fraction and Its Evaluation over Large North American River Basins. *J. Geophys. Res. Atmos.* **2007**, *112*. [CrossRef]

48. Currier, W.R.; Thorson, T.; Lundquist, J.D. Independent Evaluation of Frozen Precipitation from WRF and PRISM in the Olympic Mountains. *J. Hydrometeorol.* **2017**, *18*, 2681–2703. [CrossRef]

49. Wang, Y.H.; Broxton, P.; Fang, Y.; Behrangi, A.; Barlage, M.; Zeng, X.; Niu, G.Y. A Wet-Bulb Temperature-Based Rain-Snow Partitioning Scheme Improves Snowpack Prediction Over the Drier Western United States. *Geophys. Res. Lett.* **2019**, *46*, 13825–13835. [CrossRef]

50. Wrzesien, M.L.; Pavelsky, T.M.; Kapnick, S.B.; Durand, M.T.; Painter, T.H. Evaluation of Snow Cover Fraction for Regional Climate Simulations in the Sierra Nevada. *Int. J. Climatol.* **2015**, *35*, 2472–2484. [CrossRef]

51. Marshall, A.M.; Abatzoglou, J.T.; Link, T.E.; Tennant, C.J. Projected Changes in Interannual Variability of Peak Snowpack Amount and Timing in the Western United States. *Geophys. Res. Lett.* **2019**, *46*, 8882–8892. [CrossRef]