An Independent Evaluation of Frozen Precipitation from the WRF model and PRISM in the Olympic Mountains for WY 2015 and 2016

William Ryan Currier

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Civil Engineering

University of Washington
2016

Committee:
Jessica Lundquist
Bart Nijssen

Program Authorized to Offer Degree:
Civil and Environmental Engineering
University of Washington

Abstract

An Independent Evaluation of Frozen Precipitation from the WRF model and PRISM in the Olympic Mountains for WY 2015 and 2016

William Ryan Currier

Chair of the Supervisory Committee:
Professor Jessica Lundquist
Civil and Environmental Engineering

Estimates of precipitation from the Weather Research and Forecasting Model (WRF) and the Parameter-elevation Relationships on Independent Slopes Model (PRISM) are widely used in regions of complex terrain to obtain spatially consistent precipitation data. We evaluated the ability of both WRF and PRISM to estimate frozen precipitation using a hydrologic model (SUMMA) and a unique set of spatiotemporal snow depth observations collected for the OLYMPEX ground validation campaign during water years (WY) 2015 and 2016. We found that when WRF precipitation was partitioned with the commonly used linear-partitioning scheme based on wet bulb temperature (WRF_{LP}) that its estimation of frozen precipitation was biased low on average. However, we found that when SUMMA was allowed to partition WRF total precipitation based on WRF's microphysical scheme (WRF_{MPP}), simulations of snow depth were near equal or better than PRISM. WRF_{MPP} and PRISM had unbiased estimates of snow depth in WY 2016, but both simulated errors in snow depth of up to ~ 1 m (~40-50 cm of SWE) on an annual basis at a few locations. In the winter of WY 2015, which was abnormally warm by about 2.2°C (1 November – 31 March), PRISM was unbiased, and WRF_{MPP} over predicted annual snowfall by ~ 20% on average but showed a similar mean absolute difference to PRISM. Lastly, we hypothesize that PRISM's rain shadow has too sharp of a gradient between the windward and

leeward side of the Olympic Mountains and that the coastal proximity and topographic position weights in PRISM may shift estimates of total annual precipitation too far west of the crest.

Chapter I

An Independent Evaluation of Frozen Precipitation from the WRF model and PRISM in the Olympic Mountains between WY 2015 and 2016

1. Introduction:

Quantifying the amount of precipitation that falls as snow in complex terrain, where we have limited observations, remains a challenge. Methods that produce spatially-distributed precipitation estimates range from physically-based, numerical weather models, such as the Weather Research and Forecasting model (WRF) (Skamarock *et al.* 2008), to statistical models that spatially interpolate surface precipitation observations. A widely-used statistical model is the Parameter-elevation Relationships on Independent Slopes Model (PRISM), which is based on statistical regressions that account for topography and coastal proximity (Daly *et al.* 2008). PRISM has been widely used to spatially interpolate observations of precipitation to a grid in the following spatiotemporal datasets: Hamlet and Lettenmaier (2005), Maurer *et al.* (2002), NLDAS-2, Hamlet *et al.* (2010), Livneh *et al.* (2013), and NCEP stage IV. Previous studies have found that in complex terrain there is significant uncertainty in the estimation of spatially-distributed precipitation (Gutmann *et al.* 2012, Livneh *et al.* 2014, Henn *et al.* 2016) due to a sparse network of gauges (Lundquist *et al.* 2003) and observational uncertainty at the gauge itself (Goodison *et al*. 1998, Rasmussen *et al*. 2012). WRF or PRISM are often used in hydrologic models to influence decisions regarding avalanche control, reservoir storage, and flood forecasting. Therefore, uncertainties in the estimation of spatially-distributed precipitation directly translate into uncertainties in forecasts for agriculture, transportation, hydroelectric power, and recreation.

The Olympic Mountain Experiment (OLYMPEX) was a ground validation campaign for the NASA Global Precipitation Measurement Mission (GPM), on the Olympic Peninsula in Washington, USA (Houze *et al*. 2016), and offered a unique opportunity to compare the performance of WRF and PRISM in a maritime mountain environment. While the Olympic Mountains have been the focus location of previous dynamical (Anders *et al.* 2007, Minder *et al.* 2008) and statistical (Daly *et al.* 2008) precipitation studies before, its historical lack of mountain observations allow us to estimate how both approaches work at higher mountain elevations where data were not previously available for model training and development.

For the OLYMPEX campaign, we collected a unique set of independent snow depth observations (using cameras, poles, and LiDAR, described in section 3) to investigate the ability of a high-resolution (4/3 km) atmospheric model simulation (WRF, Mass *et al*. 2003) and PRISM to determine frozen precipitation, both annually (water years 2015 and 2016) and during individual storm events (focused on the OLYMPEX intensive observational period from November-December 2015).

This chapter is organized as follows: Section 2 provides background information on previous studies that have evaluated WRF and PRISM. Section 3 describes the location of this study and the data used. Section 4 explains our methodology. Section 5 presents the results, and section 6 discusses key model sensitivities within this maritime environment, including issues with comparing simulations of snow depth at a point to different spatial areas. Finally, in section 7 we offer our conclusions.

2. Background:

a. Prior studies of WRF and PRISM

WRF and PRISM are examples of two different approaches to provide spatially consistent estimates of precipitation. WRF is an atmospheric model that combines atmospheric dynamics with a cloud microphysical scheme, which parameterizes processes that control the formation, growth, and fallout of precipitation from clouds. WRF does not require surface gauge observations and represents changing precipitation distributions in different types of storms but is sensitive to boundary conditions and the chosen microphysical scheme. Microphysical schemes are an active area of research, particularly ice microphysics, which account for the diffusional growth, aggregation, riming, melting, and freezing of hydrometeors above the surface. In contrast, PRISM is a gridded climatology map, which estimates the variability in precipitation due to topography and coastal proximity. To obtain a spatiotemporal dataset, the climatology is commonly used with total precipitation observations (solid and liquid) at the surface from nearby gauges to spatially interpolate total precipitation between observations. Therefore, PRISM-derived precipitation in spatiotemporal datasets is highly dependent on the presence and quality of nearby precipitation observations. Due to its basis in statistical climatology, it does not accurately represent storm spatial patterns that differ from the norm but generally does provide an unbiased estimator of the sum over multiple storms.

Previous evaluations of WRF and PRISM in the western United States have reached conclusions varying with time and location. Wayand *et al.* (2013) used observations of streamflow and a set of independent precipitation gauges in the American River Basin, CA to find that the best estimate of precipitation depended on the year. However, in general, the WRF model performed as well as PRISM, which had the advantage of using most of the available precipitation gauges in its model development. Gutmann *et al.* (2012) showed in a regional analysis over Colorado that it was common for differences between PRISM and WRF to occur in areas that were furthest away from the observations of precipitation. Gutmann *et al.* (2012) also showed that winter precipitation estimated by the PRISM climatology was biased about 150% when compared to an observation that was not used in the development of PRISM. Gauge under catch at this location may exacerbate this difference; however, WRF run at 2 km-resolution differed only by ~20% when compared to the independent observations.

Lundquist *et al.* (2015) used observations from over 100 snow pillows in California's Sierra Nevada Mountains to show that gridded datasets, derived largely from PRISM, performed well over a 20-year period. Precipitation estimates resulted in a median ±10% error on a total water year (WY) time-scale. While these results are promising, the median error across all locations could exceed 50% for individual storms. These large differences were associated with storms that brought unusual synoptic conditions. Therefore, the authors suggested the use of a dynamical model during these events.

Daly *et al.* (2008) used United States Geological Survey (USGS) streamflow gauges from eight basins on the windward and leeward slopes of the Olympic Mountains along with an estimate of evapotranspiration to evaluate PRISM against other statistical precipitation models, including Daymet (Thornton *et al.* 1997) and Worldclim (Hijmans *et al*. 2005). Despite having no observations in the core mountain regions of the Olympic Mountains during the development of these products, Daly *et al.* (2008) found that both PRISM and Daymet could estimate annual mean basin precipitation well on the windward slopes. However, on the leeward slopes, Daymet and PRISM estimates diverged. Daymet overestimated mean annual basin precipitation on the leeward side because it was not able to replicate the non-monotonic relationship between precipitation and elevation. On the windward side, where both of these products agreed on mean

annual basin precipitation, the spatial interpolation dependence on topographic facets and coastal proximity used in PRISM causes spatial differences between PRISM and Daymet.

Dynamical models at relatively high resolution have previously been evaluated in the Olympic Mountains and have shown the capability of producing small-scale orographic precipitation enhancement when compared to a network of rain gauges between 50 and 900 m (Anders *et al.* 2007, Minder *et al.* 2008). Both studies revealed that a high-resolution dynamical model can capture small-scale variability and precipitation rates well on an annual or seasonal time-scale; however, individual events contained significant errors. This may pose a problem when simulating snow in this environment as significant over- or underpredictions for large events may result in an over- or under-accumulation of snow. This is because, despite steady winter precipitation, the snowpack in this region is primarily built during a few individual storms because warm rain events typically carry more moisture than cooler snow events. This indicates the importance of rain vs. snow partitioning in this environment. Dynamical models in a similar environment have previously been shown to help resolve issues with rain vs. snow partitioning by using the microphysical scheme output to calculate the fraction of rain and snow in an individual event rather than relying solely on surface temperature to partition rain vs. snow (Wayand *et al.* 2016a).

OLYMPEX offered the unique opportunity to further evaluate both WRF and PRISM's ability to predict frozen precipitation using a unique spatiotemporal snow depth dataset. However, because snow depth is not a direct measurement of frozen precipitation, we were reliant on a hydrologic model to simulate the snow depth. Since precipitation is the greatest source of uncertainty in a snow model (Raleigh *et al.* 2015), we used the hydrologic model (evaluated at four nearby SNOTEL sites) to simulate snow depth from estimates of precipitation by WRF and PRISM. We then evaluated both WRF and PRISM estimates of precipitation by comparing modeled snow depth from each simulation against independent observations of snow depth across the mountain range.

3. Location and Data

a. Location and Climate

The Olympic Peninsula is located in the northwestern corner of Washington State, USA. It is surrounded by the Pacific Ocean, the Strait of Juan de Fuca, and the Puget Sound. The mountain range, noted for its steep jagged peaks, causes significant gradients in precipitation as moisture-laden southwest flow is orographically uplifted. Precipitation estimates are greater than 6.5 m km^{-1} in the western peaks and < 2 m km^{-1} on the leeward side of the mountain range (Figure 1, Daly *et al.* 2008). These precipitation estimates make the windward side of the Olympic Peninsula one of the wettest places in the contiguous United States. Surface based radiosonde data from 1973-2007 at the nearby Quillayute Sounding showed that at around an elevation of 1200 m, winter precipitation often transitions from rain to snow (Minder *et al.* 2010). This is further demonstrated at the SNOTEL sites (elevation range of 1270 – 1527 m) where we found that forty percent of the hours with observed precipitation between 1 November 2015 and 1 April 2016 fell within -1°C and 2°C. This makes rain vs. snow partitioning of total precipitation critical in this environment for accurately simulating snowfall.

We found that during WY 2015 temperatures were 2.2ºC greater than normal (0.2ºC 1 November – 31 March), while precipitation from three NRCS SNOTEL sites was 91% of normal

(1143 mm 1 November – 31 March). These conditions were exacerbated in the months of January through March with temperatures 3.1ºC above normal. This led to a severe snow drought, with an average April 1 snow water equivalent (SWE) that was 3% of normal (622 mm). WY 2015 was therefore as warm as future climate projections, where future warming scenarios are expected to warm 2-5°C over the next century (IPCC 2013, Snover *et al.* 2013). In contrast to this, in WY 2016, temperatures were 0.4°C above normal (1 November - 31 March) and therefore April 1 SWE was 104% of the average. Winter precipitation across the three SNOTEL sites was 150% of the average.

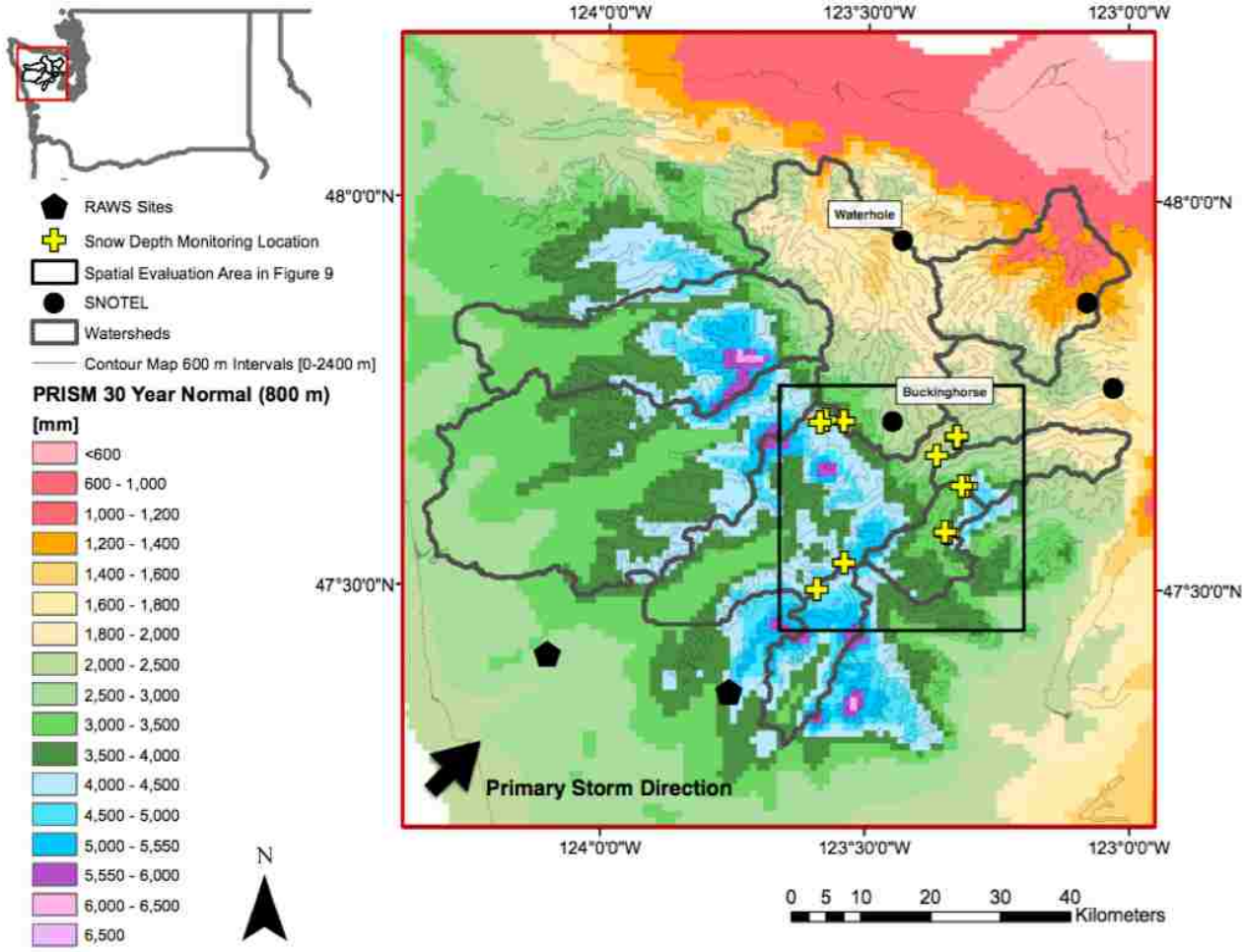

Figure 1: Independent snow depth-monitoring locations used within this study relative to PRISM 30-year precipitation averages on the Olympic Peninsula in northwest Washington State, USA. NRCS SNOTEL sites are located predominantly on the leeward side of the mountain range and are labeled. Buckinghorse was not available during the creation of PRISM but provides a key measurement in the interior region of the Olympic Mountains. PRISM was therefore largely dependent on observations outside of the interior mountain range.

b. Snow Depth Monitoring Sites

Within the Olympic National Park boundaries we set up a network of twelve snow monitoring sites. These sites provided time-series of snow depth, temperature, and relative humidity. At each location, two to three Wingcape® time-lapse cameras were deployed at roughly 2.8 to 6.8 meters high in a nearby tree depending on the site's elevation and the expected

amount of snowfall. Historic snow course measurements in this area showed snow depth values of up to 7.5 meters. Time-lapse cameras took pictures of the snow depth poles every hour during daylight hours (9:00 - 16:00 PDT). Poles ranged in height from 4-6 m, and had black tape every 5 cm and either red or neon orange/pink tape every 50 cm.

In WY 2015 all poles had red tape every 50 cm and we therefore depended on a semi-automated method to measure the snow depth poles rather than our eye because within the camera image, the red tape became indistinguishable from the black tape making it nearly impossible to process the thousands of images by eye. This method is described in Chapter II of this thesis and has been carefully evaluated to provide an uncertainty of ± 4 cm.

In WY 2016 some snow depth poles became buried in snow and were not visible for months at a time. Some snow depth poles also became significantly bent due to snow creep. We developed a method that allowed us to correct for the angle of the snow depth pole and provide an uncertainty measurement for this correction. The correction for the bent pole and the associated uncertainty estimate were based on the azimuth difference between the camera and snow depth pole, as well as the estimated angle of the pole, and is further described in detail in chapter II of this thesis. When and where the uncertainty in our measurements was greater than ± 5 cm, the measurements were not used for evaluation in this study, but were instead shown with uncertainty bounds to provide guidance in the evolution of the snowpack.

Adjacent to the cameras, in the trees we placed HOBO U23 Pro-v2 temperature/relative humidity data loggers within plastic radiation shields following the methods of Lundquist and Hugget (2008). The temperature/relative humidity data loggers were reported by the manufacturer to have an uncertainty of ±0.21ºC at 0ºC, with uncertainty increasing to about ± 0.75 ºC at -40 ºC. Relative humidity accuracy was within ±2.5% between 10% and 90% relative humidity. The uncertainty below 10% and above 90% increases to a maximum of ±3.5% including hysteresis. For quality control and redundancy, an additional HOBO Pendant Temperature Data Logger (UA-001-xx) was deployed adjacent to the temperature/relative humidity logger and had a confidence interval between ±0.53ºC from 0ºC to 50ºC. The uncertainty increases to ±0.75ºC at -20ºC.

c. NRCS SNOTEL data

Observations of hourly temperature, daily incremental precipitation, daily snow depth, and daily SWE data were received from the four National Resource Conservation Service (NRCS) SNOTEL locations on the Olympic Peninsula (Figure 1). Two of the SNOTEL sites, Buckinghorse and Waterhole, also provided hourly observations of relative humidity (RH), and Waterhole provided hourly averaged wind speed. For this study daily incremental precipitation was uniformly distributed to hourly values.

NRCS sites use precipitation accumulation reservoir gauges with antifreeze. Figure 1 shows that the majority of the SNOTEL locations are on the leeward side of the mountains. The Buckinghorse site is located in the center of the Peninsula's mountain range but was installed in 2007. This location was therefore not used in the development of the PRISM climatology.

Current SNOTEL temperature data are biased warm at cold temperatures due to an erroneous conversion from voltages to °C (Julander *et al.* 2007, Oyler *et al.* 2015, Harms *et al.* 2016). The NRCS is actively working on this issue, but to date there is no official correction. During WY 2014 we had co-located HOBO temperature sensors at three of the four SNOTEL sites in the Olympics. Observations from the HOBO measurements also suggested the same warm bias at

cold temperatures. We corrected for the warm bias in the observations by using a linear equation (1) based on a least square regression between the observations and co-located HOBO temperature sensors during WY 2014. Equation 1 is as follows:

$$T_{corr} = 1.03 * T_{SNTL} - 0.90 \quad (1)$$

where T_{SNTL} is the observed raw SNOTEL temperature in °C, and T_{corr} is the corrected SNOTEL temperature in °C.

d. RAWS Data

Additional daily precipitation observations on the windward side were received from two Remote Automated Weather Stations (RAWS) from the Western Regional Climate Center. Again, the daily precipitation data was uniformly distributed to hourly precipitation. RAWS use unheated-tipping bucket gauges, which may be subject to freezing and thus were not used in the creation of the PRISM climatology outside of May-September (Daly *et al.* 2008). However, air temperature at these two locations is above freezing throughout most of the winter as these two sites are at elevations of 470 and 1021 meters, which is below the median 0°C-isotherm height. Furthermore, we have compared the precipitation at these two sites to the WRF model's precipitation to see if there are missed precipitation events due to gauge freezing and we note that on days of precipitation from the WRF model there is precipitation at the RAWS site.

e. PRISM

The PRISM climatology group provides a map of total annual precipitation estimates over a 30-year period. PRISM is largely dependent on a regression between elevation and observations of precipitation. Individual grid cells were further modified based on the coastal proximity and the topographic facet (Daly *et al.* 2008). These two modifications allow PRISM to estimate rain shadows and the orographic enhancement of precipitation. In this study we used the 800-meter, 30-year (1981-2010) climate normal. We combined this with observations of precipitation from the RAWS and SNOTEL sites to estimate precipitation using the nearest PRISM climatology grid cell. See section 4 for more details. The maximum difference between the PRISM elevation map and the elevation of our snow-monitoring sites was 317 m, with a mean difference and absolute mean difference of 65 m and 96 m, respectively.

f. WRF

The WRF data were provided by the Northwest Modeling Consortium (Mass *et al.* 2003), which runs and archives the WRF version 3.6.1 output in three nested domains (36, 12, 4 and 4/3 km). The 4/3 km nested domain encompasses Washington State and uses the Thompson *et al.* (2004, 2008) microphysical scheme without convective parameterizations for numerical weather prediction. WRF was run with 84 h forecasts that were initialized every 12 h. As in Minder *et al.* (2010) and Wayand *et al.* (2016a), the 12-24 h forecasts were extracted from the 84 h forecasts and then pieced together to provide a temporally continuous dataset. Wayand *et al.* (2016a) demonstrated the ability of this WRF configuration to resolve mountain passes and valleys, as well as improve phase prediction in rain snow partitioning at mountain passes. The maximum

difference between the WRF terrain height and the elevation of each snow-monitoring site was 361 m, with a mean difference and absolute mean difference of 70 m and 105 m, respectively.

g. Airborne Snow Observatory Snow Depth

Two spatially complete snow depth data sets from airborne scanning LiDAR were provided by the Airborne Snow Observatory (ASO) team at the Jet Propulsion Laboratory (Painter *et al.* 2016). A snow-off flight for the Olympic Mountains was flown on 4 September 2014. Snow-on flights were flown on 8-9 February 2016 and 29-30 March 2016, and the data were processed to a three-meter gridded resolution of snow depth. The data was provided in an ellipsoidal WGS-1984 datum and was then projected into a NAD 1983 UTM Zone 10N grid using a nearest neighbor interpolation.

The accuracy of ASO in a non-forested, flat, 15 x 15 m area has been shown to have a mean absolute error of less than 8 cm, with an overall bias of less than 1 cm (Painter *et al.* 2016). In this evaluation, we focus on snowfall accumulation, and our hydrologic model was not set up to simulate forest-snow processes. Therefore, in this analysis we use the classification from the CASI 1500 imaging spectrometer, which was aboard the ASO, to remove forested pixels from the analysis. Using the March snow depth map, this removed anywhere from 17-78% of the snow depth pixels within a 60 meter bounding box (mean: 54%).

4. Methodology

As outlined in Figure 2 we first calibrated the snow model at the available SNOTEL sites using observed uniformly distributed precipitation, we refer to this as the model calibration and evaluation phase. We adjusted the precipitation partitioning parameters so that the model was unbiased for SWE until peak SWE (Figure 2: 1. a.). We chose two sets of parameter values taken from the literature for our new snow density and compaction routine to provide a range of uncertainty in modeled snow depth, as neither of these provided an optimal simulation of snow depth at all four sites (Figure 2: 1. b.). We then used the same model structure and parameters at all of our independent snow monitoring stations using precipitation estimated with PRISM and predicted from the WRF model. Since the WRF model has a microphysical scheme, the WRF model has two ways to partition frozen precipitation. We evaluated predicting frozen precipitation from the WRF total precipitation using a calibrated linear temperature threshold and using the WRF's microphysical scheme. Both simulations from WRF and the simulation using PRISM were then evaluated against our observations of snow depth and against the median ASO snow depth values within a 60-meter square-bounding box (Figure 2: 2). The following sections provide more detail on the model choice, forcing data, calibration, evaluation. Furthermore, we explicitly describe how we distribute and partition precipitation using PRISM and WRF.

a. Model set up and description

In this study we used the Structure for Unifying Multiple Modeling Alternatives (SUMMA) (Clark *et al.* 2015a,b,c). SUMMA is a modular, physically based energy balance model with a numerical solver at its core. This allowed additional parameterizations to be added to the model and allowed multiple existing modeling approaches to be vetted against one another.

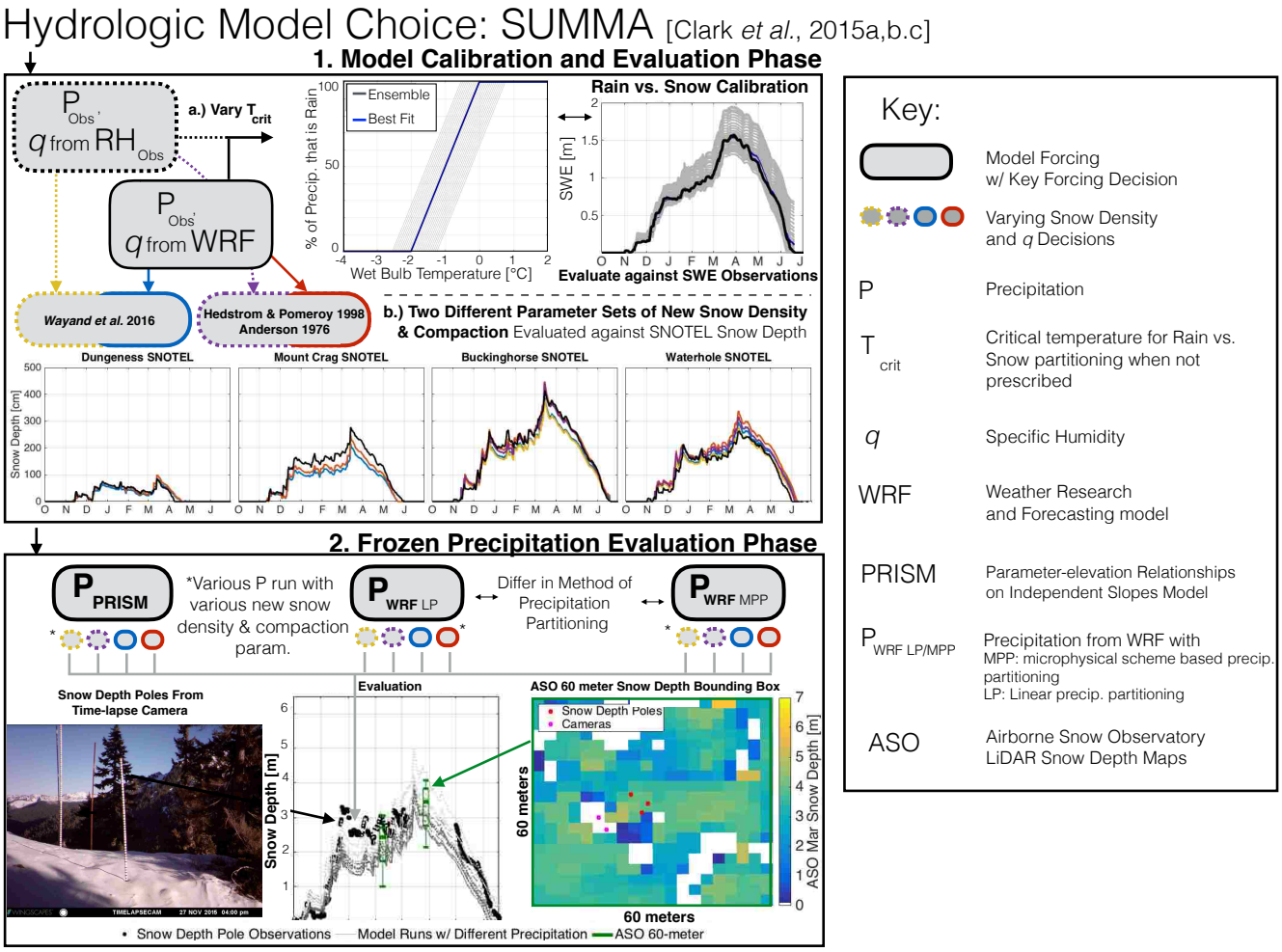

Figure 2: Conceptual figure of the overarching methodology for evaluating PRISM and WRF's ability to predict frozen precipitation in the Olympic Mountains. Methodology is broken into two phases: 1.) The calibration and model evaluation phase against the four available SNOTEL sites, using observed precipitation, two choices in specific humidity, and two different sets of new snow density and compaction parameters. Modeled snow depth colors correspond to a unique specific humidity and snow density decision, as described in the top left of the figure. Black corresponds to observed snow depth. 2.) Evaluating estimated frozen precipitation against our observations (median ASO LiDAR 60 m bounding box and snow depth poles).

b. Model calibration

Rain-snow partitioning is the most critical model parameterization in a warm maritime snow environment (Wayand *et al.* 2016a,b). Therefore, the model was manually calibrated to partition rain vs. snow in WY 2016 by searching for an optimal parameter value, T_{crit} that minimized our objective function (2). T_{crit} is the mid-point value of the wet bulb temperature range, which we held constant at SUMMA's default value, 2°C, wherein precipitation transitions linearly between rain and snow as described in USACE (1956). The objective function ensures that there was a near zero bias until peak SWE. We do not consider the melt period in our model calibration as our study is focused on an evaluation of precipitation and not the ablation season. However, because errors in the energy balance can cause biases in modeled SWE before peak SWE, we

evaluated the model-forcing variables related to the energy balance closely at the nearby Snoqualmie Pass energy balance tower (Wayand *et al.* 2015). We also note that there was little difference in the optimal parameter value, T_{crit}, based on whether we chose to use mean bias over the entire period or the root mean square error (RMSE). The objective function, $\overline{Bias}_{\rightarrow SWE_{Pk}}$, is written as follows:

$$\overline{Bias}_{\rightarrow SWE_{Pk}} = \frac{\sum_{n=1}^{SNTL}\sum_{i=1}^{k}\left(SWE_{m_{i,n}} - SWE_{o_{i,n}}\right)/k}{SNTL} \quad (2)$$

where n is an individual SNOTEL site, and k signifies the number of days between the first continuous snow period and the day of peak SWE. SWE_m is modeled SWE averaged from hourly time-steps to daily values, and SWE_o is observed daily SWE. $SNTL$ is the number of SNOTEL sites used in the evaluation. In this study, $SNTL$ differs between two and four depending on the source of specific humidity used in the model.

Based on prior modeling experience in this environment, we calibrated the model by evenly sampling 25 members of a uniformly distributed T_{crit} parameter space between -1.55°C and -0.25°Ç. The optimal T_{crit} value from minimizing $\overline{Bias}_{\rightarrow SWE_{Pk}}$ is shown in Table A1 along with other model parameters. Each model parameter was either taken from SUMMA's default values, fit to observations at the nearby Snoqualmie Pass energy balance tower, or taken from the literature (Table A1). A description of the model forcing data shown in the model calibration and evaluation phase is shown in Table A2 along with a brief explanation as to why these forcing variable estimates were chosen. Modeled SWE and snow depth from the calibration phase for each forcing data set are shown in Figure 3.

c. Model Calibration Results - SWE

Accurate measurements of specific humidity are essential, as the temperature during snowfall events is typically near the rain vs. snow threshold, and we determine rain vs. snow through the wet bulb temperature. Therefore, we tested the sensitivity of our model calibration to different specific humidity inputs, specifically comparing those measured *in situ* with those calculated by WRF (Fig. 2a). Despite differences in specific humidity and variations at individual SNOTEL sites (Table 1), both sets of forcing data converged on the same T_{crit} value for a minimum $\overline{Bias}_{\rightarrow SWE_{Pk}}$ (Table A2).

	Bias until Peak SWE [cm]	
	Model Forcing: Set 1	**Model Forcing: Set 2**
Dungeness	--	0.4
Mount Crag	--	-6.4
Buckinghorse	-1.7	1.7
Waterhole	1.2	4.8
All (mean)	-0.3	0.1

Table 1: Model evaluation statistics from WY 2016 at four NRCS SNOTEL sites using model parameters from Table A1 and model forcing decisions as described in Table A2. Bias until Peak SWE shows the individual biases in simulating SWE compared to observations for individual sites.

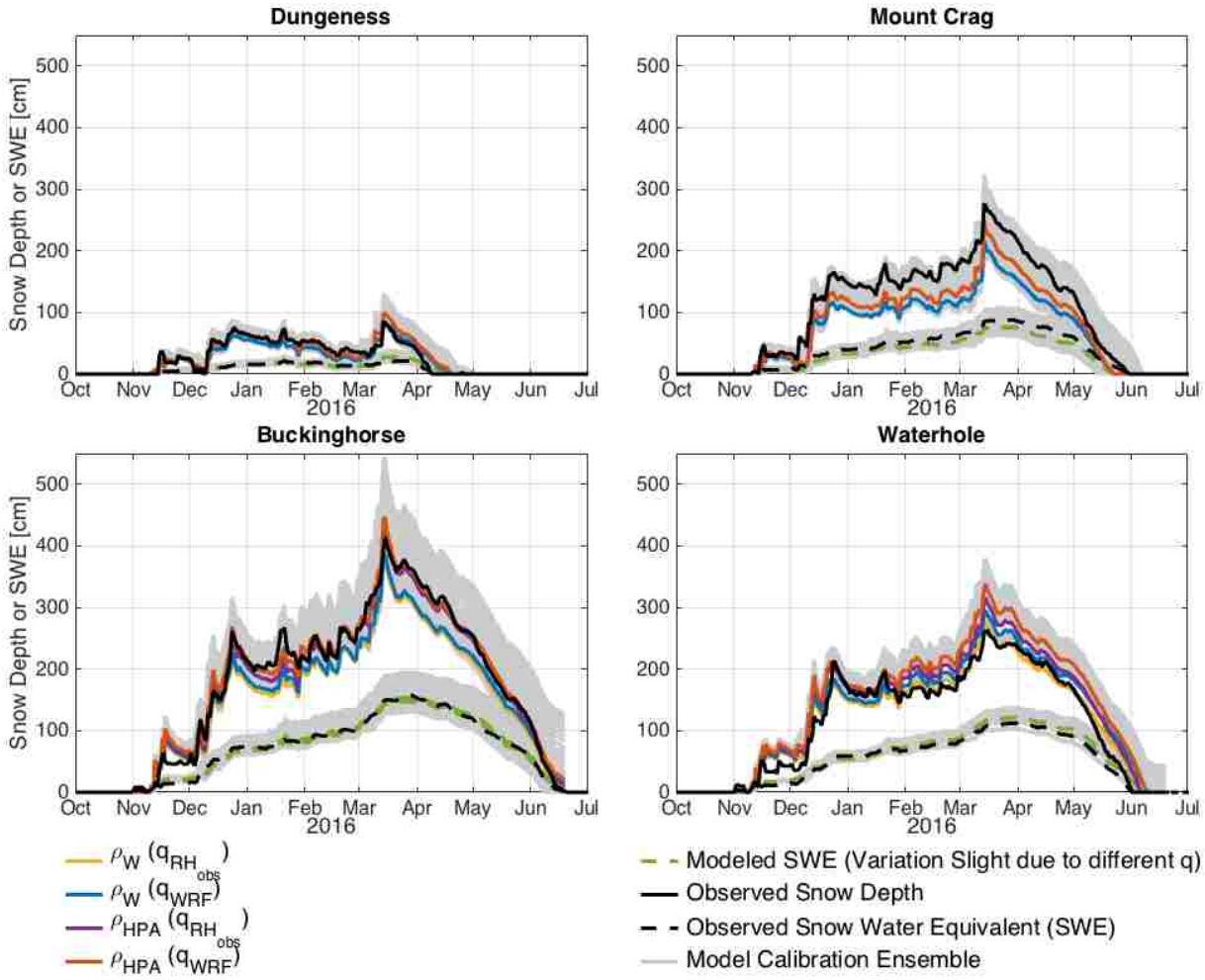

Figure 3: Resulting model SWE (dotted line) and snow depth from the calibration phase for each forcing data set. ρ_W refers to the new snow density and compaction parameters suggested by Wayand *et al.* (2016). ρ_W references the Hedstrom and Pomeroy (1998) new snow density parameters with the Anderson (1976) compaction parameters. See section 4.d. for more details. *q* references the source of specific humidity in the model (WRF or observations).

d. Model Evaluation of Snow Depth

We use two sets of literature values for the new snow density and compaction parameters (Table A1, Figure 2, Figure 3) to generate an ensemble that accounts for the model uncertainty in simulating snow depth. Within these two sets of snow density parameters, the choice in specific humidity is also varied to generate a four-member ensemble for each SNOTEL site. Snow density parameter set 1 used literature values from Wayand *et al.* (2016b) (ρ_w), who ran an ensemble of SUMMA simulations with varying parameters at the nearby Snoqualmie Pass energy balance tower to evaluate new snow density and compaction parameters against observations of snow depth. We chose results from Snoqualmie Pass, WA because of its maritime climate and relatively similar geographic region and elevation (921 m). We chose the parameter set from the bulk calibration method because the new snow density parameterization that showed optimal values from a systematic evaluation of new snow density was a function of wind speed (Boone 2002), and we only have a single observation of wind speed within our study

domain. The other set of snow density parameters comes from two separate studies (ρ_{HPA}). We chose new snow density parameters from the Hedstrom and Pomeroy (1998) study because this represented the opposite side of the new snow density spectrum, and the parameterization from Hedstrom and Pomeroy (1998) is embedded in SUMMA. The parameter values were optimized from new snow density observations in a cold continental climate, which typically provides much lower new snow density values (Judson and Doesken 2000, LaChapelle 1962). Lastly, we used default compaction parameters from the commonly used Anderson (1976) parameterization in ρ_{HPA}.

Errors on 29-30 March 2016 from simulations of snow depth for various snow density and specific humidity decisions are shown in Table 2. We chose 29-30 March 2016 as an evaluation date instead of peak snow depth because 29-30 March is the closest date to peak snow depth (~13 days) that we had Airborne Snow Observatory snow depth data. Results from Table 2 showed that the optimal bulk calibration parameters from ρ_w consistently provided lower snow depth values than the ρ_{HPA} parameters (Figure 3). Neither parameter set provided a perfect model simulation, nor was one consistently under-predicting or over-predicting snow depth (Table 2).

In our snow depth evaluation we use the mean of the ensemble for two reasons: First, there are systematic differences in how model parameters affect modeled snow depth (Table 2, Figure 3). The ρ_w model parameters resulted in lower snow depth values, while ρ_{HPA} resulted in higher snow depth values for the same model forcing data. Therefore, the ensemble mean provides a way to reduce the dimensionality of the model uncertainty. Second, no simulation was shown to be optimal at all evaluation sites.

	Snow Depth Differences Mar. 29-30 [cm / %]				
Specific Humidity Source	***q* from obs. RH**	***q* from WRF**	***q* from obs. RH**	***q* from WRF**	**Ensemble Mean**
New Snow Density and Compaction Parameters	**Wayand et al. (2016)**	**Wayand et al. (2016)**	**Hedstrom and Pomeroy (1998) & Anderson (1976)**	**Hedstrom and Pomeroy (1998) & Anderson (1976)**	
Dungeness	--	3.6 / 7.7	--	11.5 / 24.7	**7.5 / 16.2**
Mount Crag	--	-69.4 / -30.3	--	-45.9 / -20.1	**-57.6 / -25.2**
Buckinghorse	-55.8 / -15.4	-49.4 / -13.6	-12.5 / -3.5	-3.9 / -1.1	**-30.4 / -8.4**
Waterhole	-1.8 / 0.8	17.4 / 7.2	30.7 / 12.8	51.8 / 21.6	**24.5 / 10.2**
All (mean)	**-28.9 / -8.1**	**-24.4 / -7.3**	**9.1 / 4.7**	**3.4 / 6.3**	**-14.0 / -1.8**

Table 2: Difference between modeled snow depth and observed snow depth on March 29-30 2016 at four NRCS SNOTEL sites. Model differed in various snow density and specific humidity decisions (q).

In the model evaluation phase we found the percentage errors (Table 2) to be normally distributed around 0%, with a 95% confidence interval within ±33% and a standard deviation of 17%. Therefore, in the annual precipitation evaluation phase, we determined an over-accumulation or under-accumulation of total annual frozen precipitation to occur when the difference between observed snow depth and the ensemble mean of modeled snow depth was greater than 17 % on 29-30 March 2016.

e. Precipitation Distribution

In the precipitation evaluation phase, model runs that were forced with PRISM used the PRISM climatology with six sources of precipitation (P_i) to estimate precipitation at a snow

monitoring station. Sources of precipitation included the two RAWS stations and four NRCS SNOTEL stations. For example, at a single snow monitoring station, the ratio between the nearest 30-year normal PRISM annual precipitation value at the snow monitoring station and all precipitation sites was determined. These were referred to as the PRISM multiplier (m_i). The relative inverse distance was also calculated to determine a set of weights for an individual snow-monitoring site (w_i), as shown in equation 3:

$$w_i = \frac{1/d_i}{\sum_{i=1}^{n} 1/d_i} \tag{3}$$

where d_i is the distance between a snow monitoring site and a precipitation source. Here n was equal to six as there are six precipitation sources. The sum of all the weights for each snow monitoring site is equal to 1. The total precipitation at a snow monitoring station (P_{sms}) was then calculated in equation 4:

$$P_{sms} = \sum_{i=1}^{n} P_i m_i w_i \tag{4}$$

This allowed a precipitation site that was closer to the snow monitoring station to influence the precipitation at the snow monitoring station more than a source of precipitation farther away. After the precipitation was determined, the calibrated model (section 4 a-d) was run with the model decisions described in Table A1 and Table A2.

Model runs that used WRF precipitation partitioned precipitation in two different ways. In one set of model runs, total precipitation was taken directly from the nearest WRF model grid cell and converted from accumulated to incremental values. These values were aggregated to daily values and then uniformly distributed to the hourly values to be consistent with the methods we used during the model calibration. The uniformly distributed WRF data was then run using the same set of model decisions and model forcing decisions (besides precipitation) as with the PRISM model run. Both of these model runs used the linear precipitation-partitioning scheme from the SWE calibration, and we therefore refer to this as the WRF_{LP} modeling set up from here on.

In the other set of WRF model runs, SUMMA was modified to allow the fraction of rain or snow in any given event to be prescribed using *a priori* calculations based on output from WRF's microphysical scheme. Precipitation partitioning was calculated as shown in equation 5:

$$rain\ fraction = 1 - \frac{Snow+Hail+Graupel}{Total\ Precipitation} \tag{5}$$

where $Snow$, $Hail$, $Graupel$, and $Total\ Precipitation$ were derived from the WRF model's microphysical scheme and converted from accumulated values to incremental values. This modeling scenario, WRF_{MPP}, therefore is forced with WRF precipitation using the hourly incremental data and partitioning as described in the Thompson *et al.* (2008) microphysical scheme and is therefore not dependent on the wetbulb temperature. This differs from WRF_{LP}, which used the calibrated linear threshold based on wetbulb temperature to determine the correct phase of the precipitation most of the time. The WRF_{MPP} has the same model forcing data decisions and model decisions as in both the WRF_{LP} and the PRISM model runs (except for those related to precipitation and precipitation partitioning).

5. Results

a. Frozen Precipitation Evaluation – WY 2016

1. Annual Differences

Modeled and observed snow depth time-series for WY 2016 all represented storm timing well (Figure 4). In tabulating our results (Table 3), we compared the various model simulations with the median 29-30 March 2016 ASO LiDAR snow depth value within a 60 m bounding box rather than the snow depth pole, as the point pole measurements may not be spatially representative of the general area (see further details in Appendix 4), especially in cases where the pole was significantly bent or buried in snow.

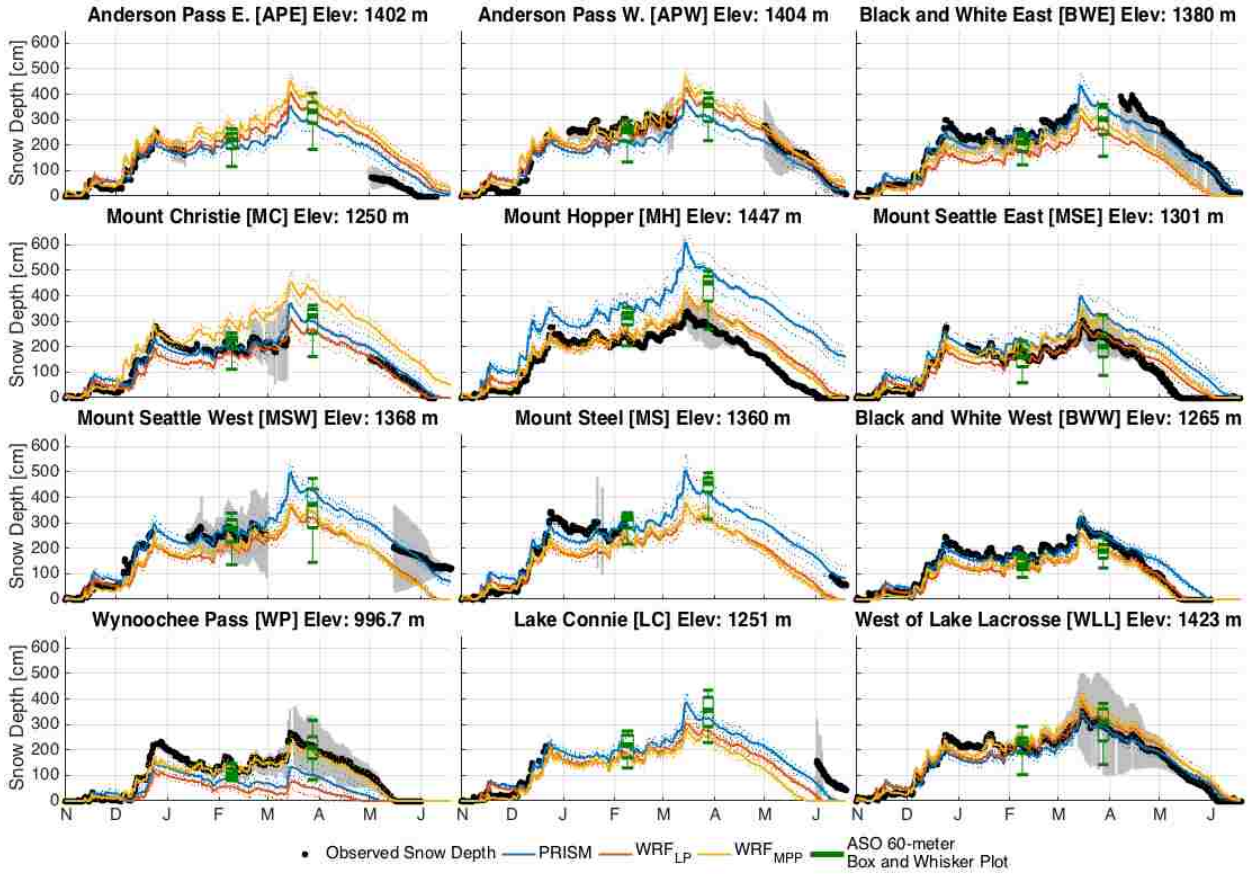

Figure 4: Snow depth time-series at all twelve independent measurement sites. Thick line represents the ensemble mean for PRISM, WRF_{LP}, and WRF_{MPP} while the thin dotted lines represent individual ensemble members. Black dotted lines with gray shading show the angle corrected snow depth measurements from the time-lapse camera network along with their relative uncertainty. Green box and whisker plots show the distribution of values from within a square 60-meter area around the snow depth pole. The minimum and maximum values of the box and whisker plot show the bottom 10% and top 90% of the snow depth values within the 60-meter bounding box, respectively.

We found that all simulations had similar mean absolute differences. However, WRF precipitation, partitioned by the microphysical scheme output (WRF_{MPP}), performed better on

average than WRF precipitation partitioned using the calibrated linear threshold (WRF_{LP}) as WRF_{MPP} had a mean difference closer to zero. WRF_{LP} was generally biased low, with a mean difference across all sites of -50.3 cm (Table 3, Figure 4). This is equivalent to ~4-8% of the estimated annual precipitation not being stored in the snowpack, assuming a bulk density of 400 kg m^{-3}.

	PRISM with linear precipitation partitioning (PRISM)		**WRF with linear precipitation partitioning (WRF_{LP})**		**WRF with prescribed precipitation partitioning (WRF_{MPP})**	
Modeled – Observed	**March Ensemble Mean Difference [cm / %]**	**March Difference Greater Than ±17%?**	**March Ensemble Mean Difference [cm / %]**	**March Difference Greater Than ±17%?**	**March Ensemble Mean Difference [cm / %]**	**March Difference Greater Than ±17%?**
Anderson Pass East	-50 / -15	No	-1 / 0	No	44 / 13	No
Anderson Pass West	-51 / -14	No	-1 / 0	No	34 / 10	No
Black and White East	45 / 15	No	-47 / -16	No	-11 / -4	No
Mount Christie	-26 / -8	No	**-60 / -19**	**Yes**	**92 / 29**	**Yes**
Mount Hopper	43 / 10	No	**-104 / -23**	**Yes**	**-106 / -23**	**Yes**
Mount Seattle East	**117 / 59**	**Yes**	**44 / -22**	**Yes**	**96 / 48**	**Yes**
Mount Seattle West	44 / 12	No	-62 / -17	No	-41 / -11	No
Mount Steel	-49 / -11	No	**-144 / -32**	**Yes**	**-140 / -31**	**Yes**
Black and White West	**60 / 32**	**Yes**	6 / 3	No	4 / 2	No
Wynoochee Pass	**-111 / -53**	**Yes**	**-161 / -77**	**Yes**	-17 / -8	No
Lake Connie	-40 / -11	No	**-88 / -25**	**Yes**	**-113 / -32**	**Yes**
West of Lake Lacrosse	-27 / -9	No	-12 / 4	No	35 / 11	No
Total Sites	*12.0*					
Mean Difference	**-4 / 0**	**--**	**-50 / -15**	**--**	**-10 / 0**	**--**
Mean Absolute Difference	**55 / 21**	**--**	**61 / 20**	**--**	**61 / 18**	**--**
Number of Occurrences Outside the Range	**--**	**3.0**	**--**	**5.0**	**--**	**5.0**

Table 3: Ensemble mean of modeled snow depth from various sources of precipitation compared to observations on 29-30 March 2016. The average modeled snow depth value between 29 and 30 March 2016 from the ensemble mean was compared to the median value within a 3600 m^2 area from the Airborne Snow Observatory data that had the forest values removed.

The WRF_{MPP} model simulated snowfall with similar skill to PRISM in WY 2016, and the best performance was related to how the errors were reported or the metric used. For instance, PRISM and WRF_{MPP} were both generally unbiased. However, PRISM resulted in a mean difference closer to zero than WRF_{MPP}, but WRF_{MPP} resulted in an almost identical mean difference when the number was reported as a percentage. Furthermore, when looking at the mean absolute difference, WRF_{MPP} had a lower mean absolute difference compared to PRISM when the number was reported as a percentage, indicating that WRF_{MPP} generally had larger errors than PRISM, but these were generally located at sites with more observed snow.

This was further shown using a different metric. We found that PRISM's ensemble mean fell outside the range of the snow model's uncertainty (±16.8%) only three times compared to WRF_{MPP} and WRF_{LP}, which both fell outside of this range five times. This suggests that we are more confident that PRISM was able to estimate frozen precipitation at more locations on an annual basis than both simulations of WRF. However, because both WRF_{MPP} and PRISM were both generally unbiased and had a similar mean absolute difference, it is difficult to definitively say which model performed best.

2. Individual Storms

Despite problems with snow depth poles becoming buried later on in WY 2016, we had high quality observations of snow depth from the time-lapse camera network during the intensive observational period of the OLYMPEX campaign (Figure 4). During this time-period we look only at positive differences in snow depth from both the observations (available daily) and the model. During this period there are two significant events: one in late November and a series of snowstorms in December.

We computed cumulative sums of snow accumulation between 1 November and 23 December 2015 and for the December snowstorms (4 December – 23 December 2015). We note that snowfall in December continued until around 25 December 2015, but our observations were no longer valid after 23 December because of dome like structures that formed on the outside of the pole, which resulted in observations that would be biased high. Therefore we limit our analysis to the period until 23 December 2015. A few exceptions to this were made at Black and White East and Lake Connie, where the evaluation ends on 18 December and 22 December, respectively. After these dates the poles become significantly bent towards the camera or buried. Total differences in accumulated snowfall between modeled and observed are shown in Table 4.

2.1: 1 November – 23 December 2015

Again, PRISM and WRF_{MPP} had a similar mean difference and mean absolute difference. PRISM had a slightly lower mean difference (19.1 cm) compared to WRF_{MPP} (24.6 cm) when looking at the total accumulation between 1 November 2015 and mid December 2015. Errors in both PRISM and WRF_{MPP} were not consistent; PRISM may have under-accumulated at one site while WRF_{MPP} over-accumulated at that same site or vice versa (Table 4; Figure 5). Similarly, errors in WRF_{MPP} and WRF_{LP} were not consistent either. WRF_{LP} had the lowest mean absolute difference but again exhibited an overall low bias (-44.9 cm).

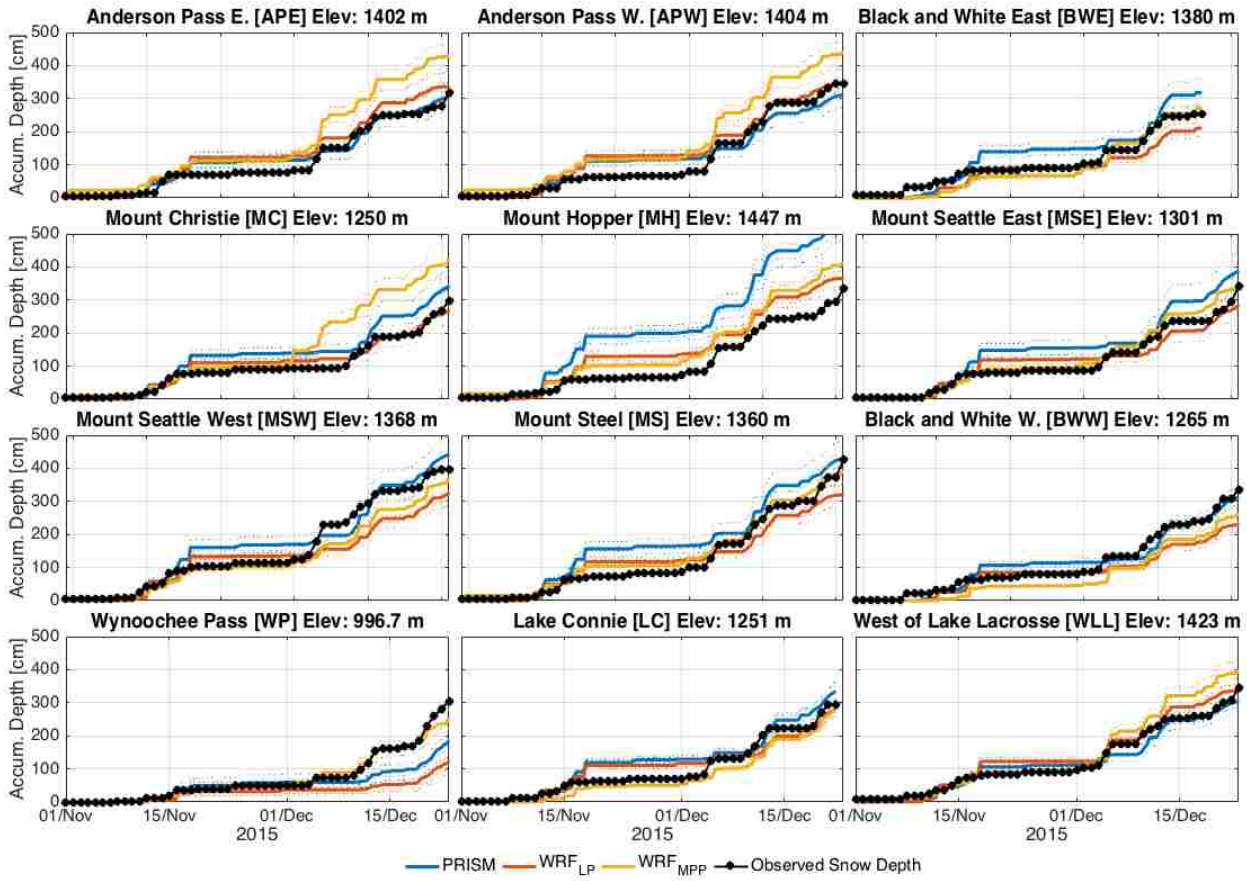

Figure 5: Accumulated snow depth between 1 November and 23 December 2015 during the OLYMPEX intensive observational period. Observed accumulated snow depth is shown with black dots and the ensemble mean of the simulated cumulative snowfall is shown in the thick colored line. Individual ensemble members are again shown with thin dotted lines.

2.2: 1 December – 23 December 2015

Looking at just December, when the Olympic Mountains received most of its snow in WY 2016, the best model performance shifted from PRISM to WRF_{MPP} (Table 4; Figure 6). WRF_{MPP} resulted in the mean difference closest to zero and lowest mean absolute difference. WRF_{MPP} had a mean difference of 12.8 cm and a mean absolute difference of 43.5 cm, while PRISM on average under-accumulated snowfall with a mean difference of -36.4 cm and a mean absolute difference of 49.2 cm. WRF_{LP} also under-accumulated snowfall on average with larger errors than PRISM. In general PRISM over-accumulated the November storm but then under-accumulated the December storm period (Figure 5 and Figure 6). We found similar behavior with WRF_{LP} but not to the same extent as PRISM. WRF_{MPP} generally simulated both periods well but exhibited biases at particular locations.

	1 November 2015 – 23 December 2015			4 December 2015 – 23 December 2015		
	PRISM Minus Obs. [cm/%]	WRF_{LP} Minus Obs. [cm/%]	Diff. WRF_{MPP} Minus Obs. [cm/%]	PRISM Minus Obs. [cm/%]	WRF_{LP} Minus Obs. [cm/%]	Diff. WRF_{MPP} Minus Obs. [cm/%]
Anderson Pass East	-13 / -4	21 / 7	**115 / 36**	-51 / -21	-26 / -11	50 / 23
Anderson Pass West	-34 / -10	2 / 1	**95 / 27**	**-78 / -29**	-52 / -19	23 / 9
Black and White East	**66 / 26**	-42 / -17	21 / 9	18 / 12	-33 / -22	27 / 19
Mount Christie	45 / 15	-27 / -9	**128 / 43**	1 / 0	-47 / -23	**71 / 35**
Mount Hopper	**205 / 61**	35 / 11	**78 / 24**	**73 / 29**	-26 / -11	23 / 9
Mount Seattle East	47 / 14	-56 / -16	11 / 3	-23 / -9	**-92 / -36**	-15 / -6
Mount Seattle West	47 / 12	-69 / -17	-23 / -6	0 / 0	**-81 / -30**	-19 / -7
Mount Steel	5 / 1	**-102 / -24**	-36 / -8	-66 / -20	**-129 / -39**	-73 / -22
Black and White West	-22 / -7	**-103 / -31**	**-73 / -22**	-50 / -20	**-103 / -42**	-39 / -116
Wynoochee Pass	**-119 / -39**	**-180 / -59**	-52 / -17	**-130 / -51**	**-167 / -66**	**-65 / -25**
Lake Connie	41 / 14	-14 / -5	-20 / -7	-12 / -6	**-55 / -25**	-13 / -6
West of Lake Lacrosse	-40 / -12	-5 / -1	51 / 15	-53 / -22	-29 / -12	35 / 15
Mean Difference	**19 / 6**	**-45 / -14**	**25 / 8**	**-31 / -11**	**-70 / -28**	**1 / 2**
Mean Abs. Difference	**57 / 18**	**55 / 17**	**59 / 18**	**46 / 18**	**70 / 28**	**38 / 16**
Average absolute difference from SNOTEL sites using observed precipitation	**--**	**35 / 14**	**--**	**--**	**28 / 15**	**--**
Maximum absolute difference from SNOTEL sites using observed precipitation	**--**	**50 / 20**	**--**	**--**	**44 / 25**	**--**

Table 4: Total difference in accumulated snowfall between the model ensemble mean and observed snowfall during two windows within the intensive OLYMPEX observational period. Average/Maximum absolute difference from SNOTEL sites using observed precipitation is the average/maximum absolute difference from the model ensemble mean and the observations of SNOTEL snow depth accumulation between the two windows. Simulations that are outside of the maximum absolute difference [cm and %] are in bold.

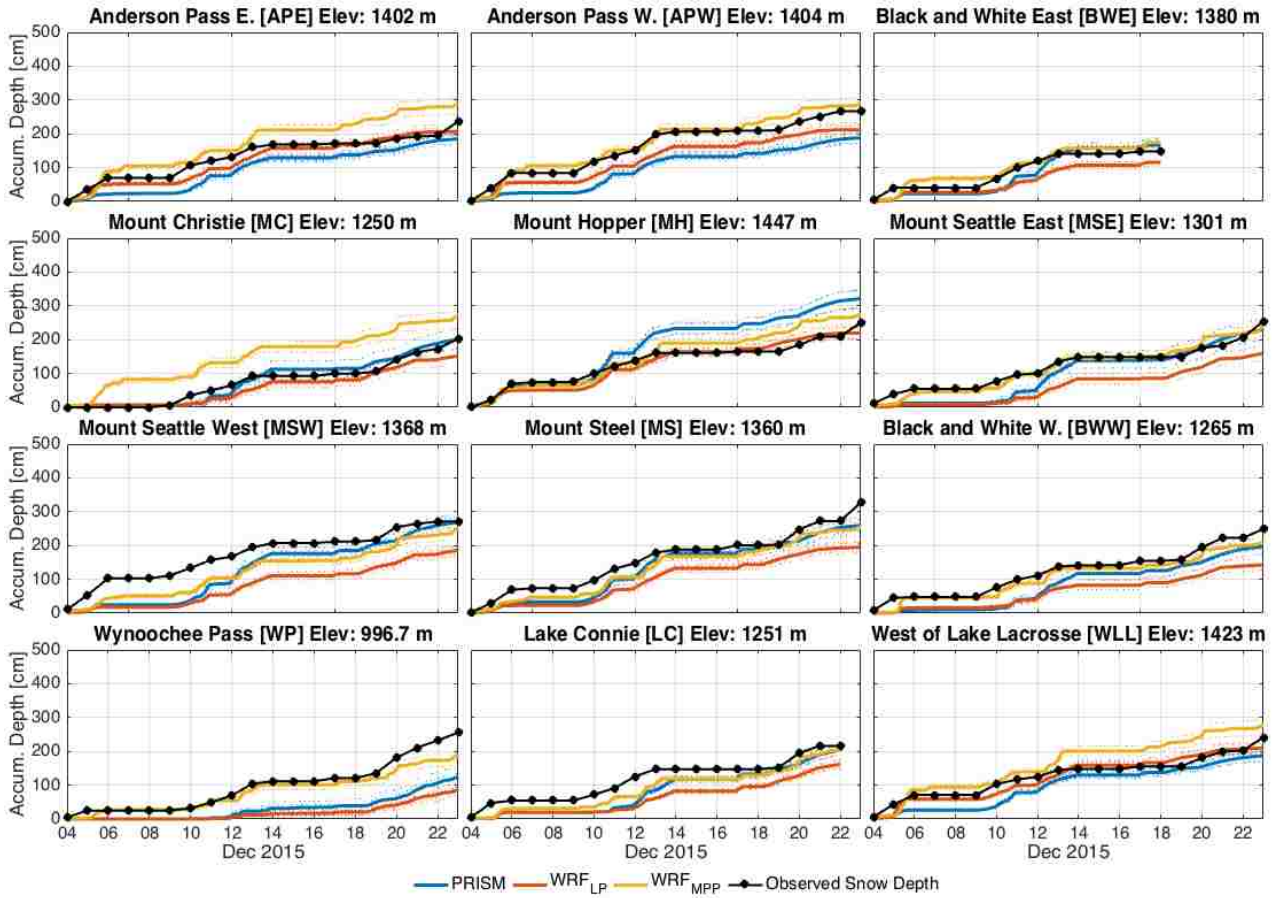

Figure 6: Accumulated snow depth between 4 and 23 December 2015 during the OLYMPEX intensive observational period. Observed accumulated snow depth is shown in black and the ensemble mean of the simulated cumulative snowfall is shown in the thick colored line. Individual ensemble members are again shown in thin dotted lines.

b. Frozen Precipitation Evaluation – WY 2015

The same model ensembles were run in WY 2015 as in WY 2016. In WY 2015 we compared the simulations to observations from seven reliable snow monitoring sites. These observations clearly reflect the abnormally warm conditions, with a maximum snow depth of around 85 cm across all our snow monitoring sites. Using the same linear temperature threshold for precipitation partitioning in WY 2015 as in WY 2016, we see promising results in model transferability to a warmer winter when modeling the four SNOTEL sites with observed precipitation (Figure 7).

In WY 2015 we took advantage of having continuous observations and used the same type of analysis as we did for the intensive observational period in WY 2016 (Figure 8). This allows us to isolate our evaluation to the simulation of snowfall from different precipitation sources. Our analysis showed that PRISM had the lowest mean difference (2.4 cm), but WRF_{LP} had the lowest mean absolute difference (68.2 cm). WRF_{MPP} had similar errors to PRISM, with a mean absolute difference of 89.4 cm. WRF_{LP} continued to be biased low, while WRF_{MPP} was biased high on average. PRISM remained unbiased in WY 2015 as in WY 2016 but we note that our sample size was reduced in WY 2015.

The total WY maximum absolute difference at the SNOTEL sites from the ensemble mean that used the observed precipitation was 28 cm (Figure 7). We found in WY 2015 that 66% of the absolute percent errors at the SNOTEL sites were less than 17%, as they were in WY 2016. PRISM, WRF_{LP}, and WRF_{MPP} sometimes differed in accumulation by the end of the water year by far more than 28 cm or by more than the 17% (Figure 8). Therefore we presume that most of the errors are attributable to precipitation sources. In WY 2015 PRISM continued to overestimate precipitation at Mount Hopper while WRF_{MPP} continued to overestimate precipitation at Mount Christie. Mount Steel was simulated well by PRISM in both years, and Black and White East was simulated well by WRF_{MPP} and PRISM in both years.

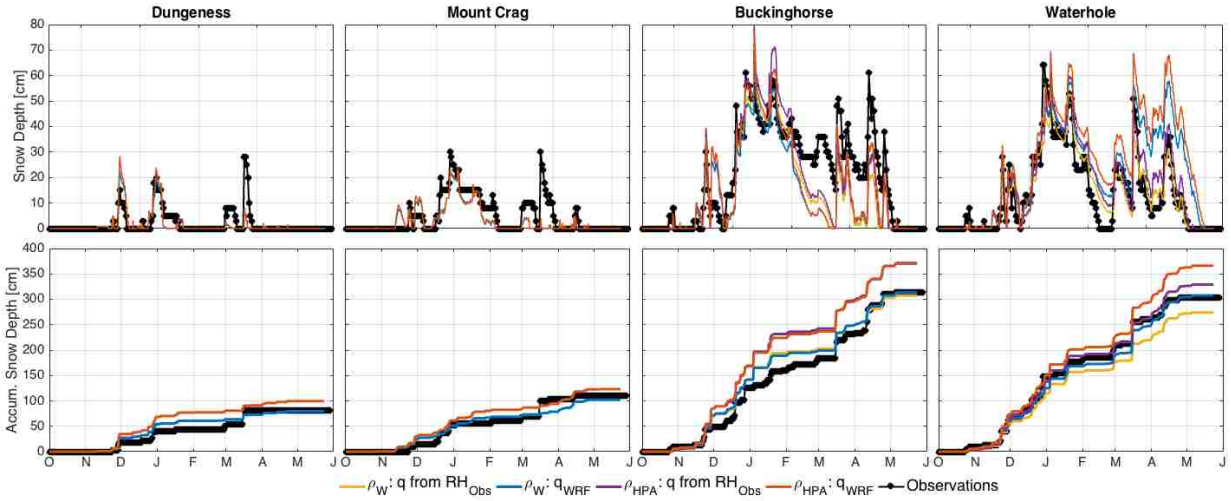

Figure 7: Top panel: Model and observed snow depth in WY 2015 at the four SNOTEL sites. Black dots represent the observations. Bottom panel: Model and observed accumulated snow depth during the entire water year 2015. Black dots represent the positive daily differences from the observations.

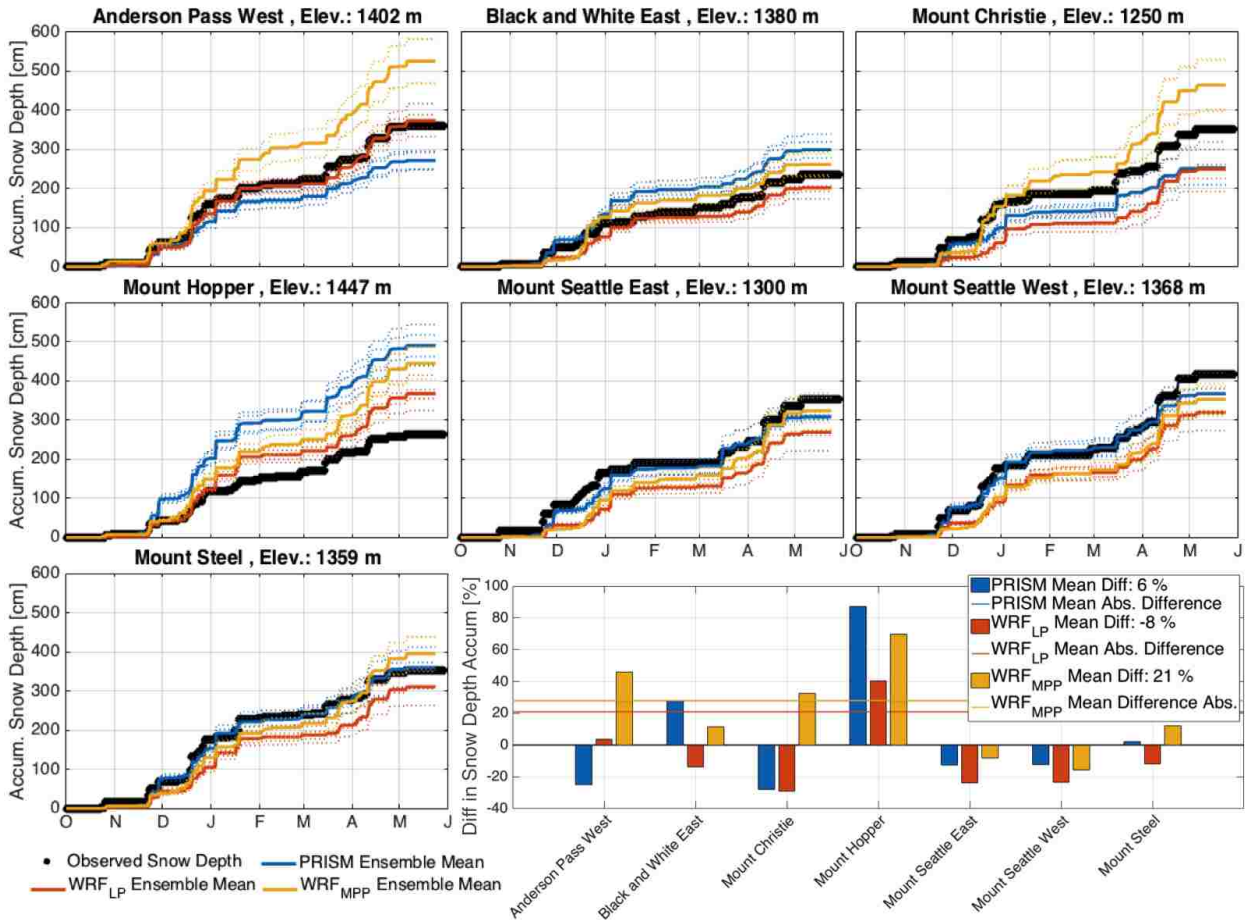

Figure 8: Accumulated snow depth during the entire water year 2015. Observed accumulated snow depth is shown in black and the ensemble mean of the simulated cumulative snowfall is shown in the thick colored line. Individual ensemble members are again shown in thin dotted lines. Percent errors in total accumulated snowfall are shown in the bar chart. Colors are consistent between line charts and bar chart. Horizontal lines in bar chart show the mean absolute percent difference across the seven independent evaluation sites in WY 2015. Gold line (WRF_{MPP}) overlaps the blue line (PRISM).

c. Spatial Distribution of Errors

When considering the directionality of the model differences from the observations in WY 2016, spatially coherent patterns appear (Fig. 9). PRISM shifts from over-accumulating in the Northern Quinault (MSW & MSE) to under-accumulating in the Elwha Watershed (MC, BK). The PRISM climatology (Figure 1) shows that there is a significant drop in PRISM estimated precipitation between the Northern Quinault and the Elwha Watershed. This intuitively makes sense as the mountain range transitions from the windward side to the leeward side. However, the dramatic difference (-101.4 cm) in modeled snow depth using PRISM at Buckinghorse (BK) from observed snow depth suggests that PRISM weights associated with topographic position may be estimating too sharp of a gradient between the windward and leeward side of the mountain range.

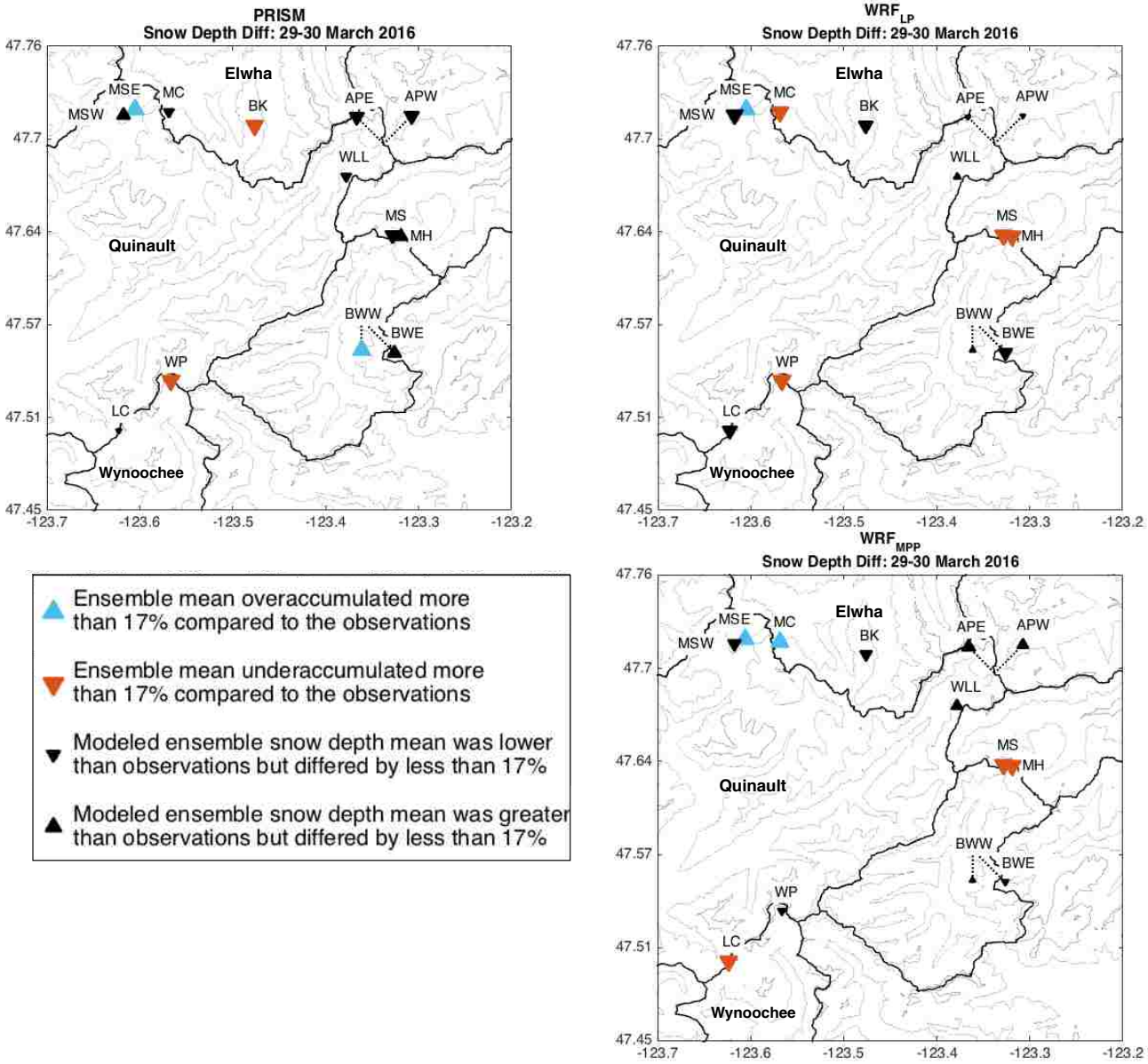

Figure 9: Spatial differences in errors at independent snow depth sites. Black triangles show that the mean of the model ensemble for PRISM, WRF_{LP} or WRF_{MPP} differed by less than 17% of the observations but are scaled for how much they differ from the observations. Black lines show watershed delineations based on USGS streamflow gauges for seven different watersheds.

Furthermore, PRISM under-accumulated frozen precipitation in the Eastern Quinault (APE, APW, WLL). Here, we postulate that the spatial weights due to coastal proximity may cause PRISM's estimate of precipitation to be too low. For instance, Anderson Pass East (APE), Anderson Pass West (APW), and West of Lake Lacrosse (WLL) are inside the Eastern Quinault Watershed, and all indicate that PRISM under-accumulated snowfall. These sites do not sit on the leeward side of the mountain range, as did MC and BK but instead, they are located on the windward side, on the banks of a deep U-shaped valley, referred to as the Enchanted Valley.

Additional support for both of these theories is that WRF_{MPP} over-accumulated frozen precipitation at Buckinghorse, Anderson Pass, and West of Lake Lacrosse. Therefore, the physically-based model and the observations of snow depth indicate that this region is wetter, at

least during cold precipitation events, than PRISMs annual climatology would suggest. Also, the Eel Glacier and Anderson Glacier are located near the Anderson Pass sites and provide physical evidence that on average this is one of the wetter areas in the Olympic Mountains. However, we cannot definitively conclude that PRISM is wrong in the Eastern Quinault because PRISM simulated snow depth differed by less than 17%, which means the difference could be due to the snow model uncertainty.

Model simulations that estimated precipitation with PRISM simulated snow depth errors at Mount Steel (MS, Fig. 9) and Mount Hopper (MH, Fig. 9) with different directionalities, but MS and MH were located about 700 meters apart from each other. However, both sites were located in two separate PRISM grid cells that differ in precipitation estimates by about 10%. MH is located at the edge of the PRISM grid cell, and therefore it may better reflect the MS precipitation estimate.

WRF_{MPP} and WRF_{LP} also showed differences in directionality at Mount Seattle West (MSW) and Mount Seattle East (MSE). Both of these sites were forced with the same WRF precipitation as they were located around 950 meters apart from each other and therefore had the same nearest grid cell. However, both experienced different observed snow accumulation in both water years (Figure 4, 5 and 8). MSW accumulated more snow than MSE and therefore suggests that within a WRF grid cell there is spatial variability in observed snowfall. Therefore the over-accumulation compared to observations at MSE may be balanced out by the under-accumulation seen at MSW, and WRF may have accurately simulated snowfall within the area of this WRF grid cell. This highlights the complexity of evaluating gridded precipitation sources at larger spatial resolutions to smaller domains (point or 60 m area). See discussion for more information.

6. Discussion

a. Spatial Representativeness of LiDAR Observations to PRISM and WRF Grid Cells

There is a significant range in observed snow depth within a spatial scale of 60, 800 (PRISM resolution), and 1333 (WRF resolution) meters (Figure 10). Assuming a snow density of 400 kg m^{-3}, SWE ranges from around 0.5 to 2.5 meters, depending on the site. In general the snow depth distributions were similar, regardless of spatial scale. However, the median values were not always consistent amongst different spatial areas. An analysis showed that the median snow depth value from a 60 m bounding box was generally higher by about 20-60 cm when compared to the median value from an 800 m bounding box. Similarly, we found that the median elevation within a 60 m bounding box was around 40-80 meters higher than the elevation within an 800 m bounding box. Therefore the change in the median elevation value and the median snow depth values at different spatial scales were generally correlated. However, when we used an observed lapse rate and the calculated sensitivity of rain vs. snow to temperature (see section 6. b.) we could not fully explain the differences in median snow depth values across spatial scales based on changes in median elevation alone. We therefore hypothesize that interactions between wind, terrain, vegetation, and the sensitivity of temperature to rain vs. snow partitioning may be responsible for these differences in median values at different spatial scales. We suggest future work use high-resolution distributed modeling (3 m grid cell) that accounts for the local and complex elevation changes within an 800 and 1333 m grid to compare distributions of snow depth from LiDAR observations to those simulated by the distributed snow model.

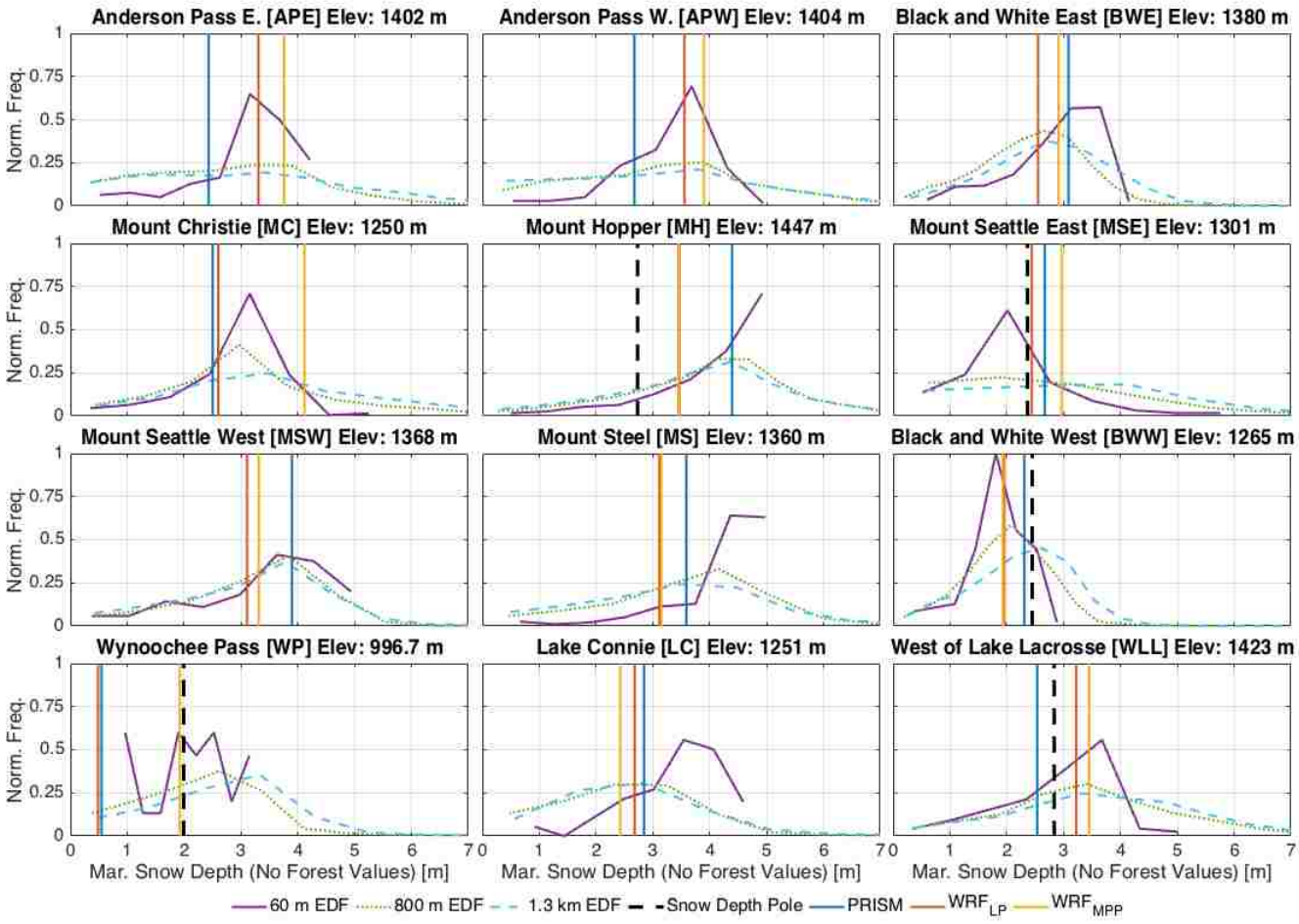

Figure 10: Normalized empirical distribution functions (EDF) showing the distribution of snow depth values within different spatial areas surrounding the location of the snow depth pole. Vertical dotted lines show the corrected pole measurements when available. Solid vertical lines show the mean model ensemble for various precipitation sources on 29-30 March 2016. Note that at Mount Hopper, Mount Seattle West, Mount Steel and West of Lake LaCrosse WRF_{MPP} (gold) and WRF_{LP} (red) overlap.

Lake Connie (LC), which sits near the top of a ridgeline, had the most dramatic difference (~1 meter) between the median observed snow depth values and median elevation when looking at different spatial scales. The distributions show that both WRF_{MPP} and WRF_{LP} were more reflective of the median snow depth value from the 800 m or 1333 m bounding box rather than the 60 m bounding box (Figure 10), and therefore both WRF_{LP} and WRF_{MPP} may not be under-accumulating at LC (Figure 4, Figure 9). Similarly, when PRISM was compared to the median value from the larger spatial domain at LC, the PRISM estimate simulated more snow depth than the median ASO snow depth value from an 800 or 1333 m bounding box, but it agreed with the snow depth value from the median 60 m bounding box. Therefore, it is possible PRISM over accumulated snow in this region.

Furthermore, WRF_{MPP} and WRF_{LP} at MSE over-accumulated snow when compared to the median value from the 60 m bounding box. However, WRF_{MPP} and WRF_{LP} at MSE agreed with the median value from a larger spatial domain (Figure 10). This is consistent with results from section 5.c, where we showed that MSE and MSW had the same WRF precipitation, but both MSE and MSW experienced different amounts of observed snowfall. This highlighted the spatial

variability of snowfall and snow depth within a WRF grid cell and suggested that the over-accumulation at MSE may be balanced out by looking at a larger spatial distribution of snow depth within the domain of a WRF grid cell. We found by incorporating observations from the nearby MSW site (more snowfall; Figure 5) and by looking at the median value from a larger spatial domain (Figure 10) that this was true. Together, this showed that at MSE, WRF_{MPP}, and WRF_{LP} accurately simulated snowfall within the size of a WRF grid cell.

Despite the difficulty and complexity of evaluating gridded precipitation products to both point or larger spatial areas, we highlight that at most locations, like MSW, there was little change in the evaluation of PRISM, WRF_{MPP}, and WRF_{LP} when moving from the 60 m bounding box to a larger spatial area (Figure 10). For instance, PRISM simulations at sites in the Eastern Quinault (APE, APW, MLL; Figure 9) appear to still be biased low compared to the observations at larger spatial areas (Figure 10). Furthermore, WRF_{MPP}, WRF_{LP}, and PRISM at most sites still tend to be simulating snow depth well when compared to the spatial distributions of snow depth from larger and smaller spatial areas.

b. Rain vs. Snow Sensitivity

There were significant differences in model performance based on the method used for rain vs. snow partitioning. Each method had its own advantages and disadvantages. For instance, the calibrated linear temperature threshold method was dependent on uniformly distributing daily total precipitation observations because the calibration locations in the Olympic Mountains were at locations with only daily observations of precipitation. We explored what would happen if the calibrated linear temperature threshold were run with hourly WRF precipitation. We found a systematic and more significant low bias (-109 cm) in the simulations of snow depth compared to when we used the uniform distribution of WRF daily precipitation (WRF_{LP}: -50 cm). However, when this calibrated linear temperature threshold was used with the observed daily precipitation observations and the PRISM climatology, we found that the model was generally unbiased in simulating snow depth on an annual basis, but there were generally errors during individual storms, making this method promising for summer water supply forecasting but maybe not for estimating snowfall during individual storms. This is because this method is designed to correctly partition precipitation on average.

We found that using the output from a microphysical scheme for precipitation and partitioning (WRF_{MPP}) yielded results that were as good or better than the PRISM method. This indicated a promising path forward in hydrology, as WRF_{MPP} was able to simulate changes in rain vs. snow at an hourly time step and was not dependent on using the linear temperature threshold for a non-linear process. For example, using the output from the microphysical scheme, we back calculated the T_{crit} model parameter by finding the wet bulb temperature during events that had a rain fraction (5) that was between 20% and 80%. The back calculated T_{crit} from WRF wet bulb temperature was normally distributed around 0.7°C with a 95% confidence interval that ranged from -2.2°C to 3.6°C and a standard deviation of 1.5°C. The back calculated T_{crit} model parameter was also found to be dynamic in that it changes from hour to hour within a storm. Similar results were found when we constrained the rain fraction to 30-70%.

Additionally, we found T_{crit} to be a highly sensitive model parameter. During the model calibration, we found that a 1°C (-1.4-0.4°C) change in the T_{crit} model parameter resulted in a difference at the Buckinghorse SNOTEL site in over a half a meter (52 cm) of SWE. Errors in modeled SWE at Buckinghorse thus ranged from -14% to 18% for a 1°C change in the T_{crit}

parameter value. Similar results were found across the rest of the SNOTEL sites as well. We determined that for a 0.2°C change in the critical temperature value, the error in peak SWE changed on the order of anywhere between 2 and 8%.

Using temperature sensors that were deployed during the OLYMPEX campaign that ranged from 180 meters to 1,458 meters, we determined the mean lapse rate to be -4.5 °C km^{-1} during precipitation events. This is consistent with previous work (Minder *et al.* 2010) that showed lapse rates within the Pacific Northwest are considerably less than the assumed -6.5°C km^{-1}. This -4.5 °C km^{-1} lapse rate in conjunction with the rain vs. snow sensitivity shows that an elevation change of 100 meters could result in a 4 to 16% change in peak SWE for that elevation band. This sensitivity underscores the importance of using a correct lapse rate in this environment when using a distributed model, as using a commonly applied -6.5 °C km^{-1} lapse rate will result in a significant difference in total SWE per elevation band than what would likely be observed.

When using WRF with the Thompson *et al.* (2008) microphysical scheme to partition precipitation (WRF_{MPP}), a lapse rate does not need to be known. Furthermore, the WRF_{MPP} model set up also does not require model calibration to partition rain vs. snow. This is critical, as calibrated model parameters require observations of SWE/snow depth and precipitation observations, which are not always available. We therefore recommend that future snow model development in maritime environments focus on the best microphysical scheme rather than partitioning schemes based on temperature or wet bulb temperature. Here we only evaluated the Thompson *et al.* (2008) microphysical scheme, but it is promising that in a comparison of microphysical schemes, the Thompson *et al.* (2008) was found to perform best in predicting snow across the cold continental climate of Colorado, USA (Liu *et al.* 2011). Furthermore, this microphysical scheme and model set up helped improve partitioning rain and snow at Snoqualmie Pass, WA, especially during cold air intrusions (Wayand *et al.* 2016a). We recommend that the snow depth data presented herein be used along with other OLYMPEX observations to evaluate how WRF performs with different physics, and hopefully build even better microphysical schemes that can be used in current and future climate studies.

c. Model Sensitivity Unique to Warm Maritime Snow Environments

Throughout this study many non-trivial model-forcing decisions were made to improve model skill. We found one of the most sensitive decisions in this environment to be incoming longwave radiation. We evaluated empirical methods that performed well in Flerchinger *et al.* (2009) but found significant differences in model performance based on the empirical method for longwave radiation. We found that the Dilley and O'Brien (1998) clear sky method with the Unsworth and Montieth (1975) method for cloud correction had the lowest RMSE and mean difference when compared to observed incoming long wave radiation at the nearby Snoqualmie Pass, WA, energy balance tower. We note that this was also one of the most transferrable equations in Flerchinger *et al.* (2009). Other suggested transferrable longwave parameterizations from Flerchinger *et al.* (2009) simulated either too much or too little incoming longwave radiation. For instance, the Dilley and O'Brien (1998) clear sky method with Kimball *et al.* (1982) cloud correction, resulted in too little incoming longwave compared to the Snoqualmie Pass observations and therefore resulted in not enough melt in our model simulations of SWE and snow depth.

Another non-trivial decision was the choice in wind speed. In this region we found modeled turbulent fluxes to be significant energy input to the snowpack, because throughout the season

both the latent and sensible heat fluxes can be oriented in the same direction. Therefore, we found that using the only wind speed measurement within our study domain (Waterhole SNOTEL), despite being located far away from independent locations, offered better model performance than using a constant 2 m s^{-1}. This was because observations of wind speed were < 2 m s^{-1} on an hourly average during most of the ablation season. Therefore, using a constant 2 m s^{-1} wind speed resulted in a higher ablation rate compared to observations at independent locations because sensible and latent heat fluxes were often both simulated to be directed towards to snowpack. We recommend other models set up in a similar environment pay close attention to choices in incoming longwave radiation and wind speed.

7. Conclusions

PRISM and WRF_{MPP} were both unbiased in predicting frozen precipitation across our twelve independent snow depth sites during WY 2016. However, significant individual errors occurred that resulted in a difference of up to over a meter (50-60%) in snow depth. The average error for both PRISM and WRF_{MPP} was around 20%, but individual sites often had percent errors that differed less than 17% (see section 4 d). When output from WRF's microphysical scheme was used to partition precipitation (WRF_{MPP}) instead of the commonly used linear partitioning scheme based on wet bulb temperature (WRF_{LP}), WRF performed similar to PRISM on average. The best model performance depended on the metric and the evaluation period. For instance, WRF_{MPP} had a lower mean absolute difference when the error was presented as a percentage, while in contrast PRISM had a lower mean absolute error, indicating they both performed similarly in WY 2016, but that WRF_{MPP} had larger errors at sites with more snow. Furthermore, WRF_{MPP} outperformed PRISM during the largest storm (4 December 2015 – 23 December 2015), while PRISM performed better over the entire OLYMPEX intensive observational period (1 November 2015 – 23 December 2015) and throughout WY 2015. Since PRISM is based on a precipitation climatology, it was not specifically designed to simulate extreme events like the December 2015 snowstorms, but PRISM over longer time-periods had individual errors balance out (Figure 5 and 6).

Using the observations and the WRF simulations we found that PRISM may be under predicting snow depth near the crest of the Olympic Mountains (Figure 9) and therefore suggest that PRISM shift the boundary of the rain shadow further East in the Quinault Watershed and further into the Elwha Watershed. For instance, WRF_{MPP}, which had the advantage of taking into account atmospheric synoptic conditions for individual storm events, indicated, along with the observations, that these areas generally experienced more snow.

The WRF_{MPP} modeling scenario is an attractive path going forward in snow hydrology for four reasons: 1.) The rain vs. snow threshold does not have to be calibrated, 2.) The lapse rate, which affects precipitation partitioning, does not have to be known *a priori*, 3.) WRF_{MPP} was shown to perform well in a typical maritime winter and in a winter that experienced mean winter temperatures that are expected within the next century (IPCC 2013, Snover *et al.* 2013), making it a valuable modeling tool for climate studies that use dynamical downscaling, and 4.) WRF_{MPP} has the potential to be run anywhere, even in watersheds with no observations or a gridded climatology.

Microphysical schemes, regarding the magnitude and phase of precipitation are an active area of research (Jankov *et al.* 2009, Liu *et al.* 2011, Minder and Kingsmill, 2013) and previous studies have noted significant over- and under-accumulation of precipitation during individual

events (Anders *et al.* 2007, Minder *et al*. 2008), as shown here at a select few sites. We believe that these observations of snow, during WY 2015 and 2016, should be used with other OLYMPEX observations going forward to evaluate other microphysical schemes during various synoptic conditions and possibly develop new schemes so that we can better simulate the phase and magnitude of snow events, especially in this relatively warm maritime environment.

8. Acknowledgements

We gratefully acknowledge funding support from NSF (EAR-1215771) and NASA (Grant numbers NNX13AO58G and NNX14AJ72G). We would also like to thank Olympic National Park for providing us with the permission to install snow depth poles within park boundaries. Furthermore, we would like to thank Clifford Mass and Neal Johnson for providing access to the WRF data archive. We would also like to thank Theodore Thorson, Derek Beal, Colin Butler, Justin Pflug, Max Mozer, Adam Massmann and Brad Gaylor for the tremendous effort they gave to help install the snow depth monitoring sites in remote regions of Olympic National Park and with help processing these data. The author also wishes to express his gratitude to Nicholas Wayand for his help with using WRF data and setting up SUMMA. Lastly, the author would like to express sincere appreciation to Jessica Lundquist and Bart Nijssen for their counsel.

All meteorological and snow depth data collected during the OLYMPEX campaign are archived at the Global Hydrology Resource Center Distributed Active Archive Center (GHRC DAAC) and are publicly available. SUMMA model code is available at https://github.com/NCAR/summa/ along with more information at http://www.ral.ucar.edu/projects/summa. A description of how to derive SW radiation data using MTCLIM can be found at: https://vic.readthedocs.io/en/vic.4.2.c/Documentation/ForcingData/. The PRISM 800 meter, 30-year (1981-2010) climate normal was downloaded from http://www.prism.oregonstate.edu/normals/. The RAWS precipitation data was downloaded from http://www.raws.dri.edu/. All NRCS SNOTEL data were downloaded from: http://www.wcc.nrcs.usda.gov/snow/.

9. References

Anders A. M., G. H. Roe, D. R. Durran, J. M. Minder, 2007: Small-scale spatial gradients in climatological precipitation on the olympic peninsula. *J. Hydrometeorol*. **8**, 1068–1081.

Anderson, E. A., 1976: A point energy and mass balance model of a snow cover. NOAA Tech. Report NWS 19. 172 pp, [Available online at: http://amazon.nws.noaa.gov/articles/HRL_Pubs_PDF_May12_2009/HRL_PUBS_51-100/81_A_POINT_ENERGY_AND_MASS.pdf]

Boone, A. 2002: Description du Schema de Neige ISBA-ES (Explicit Snow). Centre National de Recherches Météorologiques. 63 pp, [Available online at: http://www.umr-cnrm.fr/IMG/pdf/snowdoc.pdf]

Clark, M. P., and Coauthors, 2015a: A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resour. Res.*, **51**, 1–17. doi:10.1002/2015WR017200.A

Clark, M. P., and Coauthors, 2015b: A unified approach for process-based hydro- logic modeling: 2. Model implementation and case studies. *Water Resour. Res.*, **51**, 2515–2542. doi:10.1002/2015WR017200

Clark, M. P., and Coauthors, 2015c: The structure for unifying multiple modeling alternatives (SUMMA), version 1: Technical Description. NCAR Tech. Note NCAR/TN-514 + STR, 54 pp, doi:10.5065/D6WQ01TD

Daly, C., and Coauthors, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, doi:10.1002/joc.1688.

Dilley, A. C., and D. M. O'Brien, 1998: Estimating downward clear sky long-wave irradiance at the surface from screen temperature and precipitable water. *Quart. J. Roy. Meteor. Soc.*, **124**, 1391-1401

Flerchinger, G. N., W. Xaio, D. Marks, T. J. Sauer, and Q. Yu, 2009: Comparison of algortihms for incoming atmospheric long-wave radiation. *Water Resour. Res.,* **45**, W03423, doi:10.1029/2008WR007394.

Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation measurement intercomparison, final report. WMO/TD-872, 212 pp.

Gutmann, E. D., R. M. Rasmussen, C. Liu, K. Ikeda, D. J. Gochis, M. P. Clark, J. Dudhia, and G. Thompson, 2012: A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. *J. Climate*, **25**, 262–281, doi:10.1175/2011JCLI4109.1.

Harms, D. S., J. Weeks, and J. Lea, 2016: Comparing SNOTEL Extended Air Temperature Sensor and Equations to an NSIT Certified Sensor in an Environmental Chamber. Poster session presented at: 84th Annual Western Snow Conference, Seattle WA

Hamlet, A. F., and D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature fields for the continental United States. *J. Hydrometeor.*, **6**, 330–336, doi:10.1175/JHM420.1.

Hamlet, A. F., and Coauthors, 2010: Final report for the Columbia Basin Climate Change Scenarios Project. Pacific Northwest (PNW) Hydroclimate Scenarios Project 2860, Climate Impacts Group, University of Washington. [Available online at http://warm.atmos.washington.edu/2860/.]

Hedstrom, N. R. & J. W. Pomeroy, 1998: Measurements and modelling of snow interception in the boreal forest. *Hydrol. Processes,* **12**, 1611–1625. doi:10.1002/(SICI)1099-1085(199808/09)12:10/11<1611::AID-HYP684>3.0.CO;2-4

Henn, B., M. P. Clark, D. Kavetski, A. J. Newman, M. Hughes, B. McGurk, and J. D. Lundquist, 2016: Spatiotemporal patterns of precipitation inferred from streamflow observations across the Sierra Nevada mountain range. *J. Hydrol.*, http://dx.doi.org/10.1016/j.jhydrol.2016.08.009

Hijmans R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, 2005: Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climate*, **25**, 1965–1978.

Houze, R. A., and Coauthors, 2016: The Olympic Mountains Experiment (OLYMPEX), *Bull. Atmos. Meteor. Soc*., in prep.

IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jankov, I., J. W. Bao, P. J. Neiman, P. J. Schultz, H. L. Yuan, and A. B. White, 2009: Evaluation and Comparison of Microphysical Algorithms in ARW-WRF Model Simulations of Atmospheric River Events Affecting the California Coast. *J. of Hydromet.*, **10**, 847-870.

Judson, A. and N. Doesken, 2000: Density of Freshly Fallen Snow in the Central Rocky Mountains, *Bull. Atmos. Meteor. Soc.*, **81**, 1577-1587

Julander, R. P., J. Curtis, and A. Beard, 2007: The SNOTEL Temperature Dataset, *Mt. Views. Newsl. Consort. Integr. Climate Res. West. Mt.*, **1**, 1-12.

Kimball, B. A., S. B., Idso, and J. K. Aase, 1982: A model of thermal radiation from partly cloudy and overcast skies, *Water Resour. Res.*, **18**, 931-936

LaChapelle, E. R., 1958: Winter snow observation at Mt. Olympus. *Proc. 26th Annual Meeting,* Bozeman, MT, Western Snow Conference, 59–63.

Liu, C., K. Ikeda, G. Thompson, R. Rasmussen, and J. Dudhia, 2011: High-resolution simulations of wintertime precipitation in the Colorado Headwaters Region: Sensitivity to physics parameterizations, *Mon. Wea. Rev.*, **139**, 3533–3553, doi:10.1175/MWR-D-11-00009.1.

Livneh, B., and Coauthors, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate*, **26**, 9384– 9392, doi:10.1175/JCLI-D-12-00508.1.

Livneh, B., J. S. Deems, D. Schneider, J. Barsugli, and N. Molotch, 2014: Filling in the gaps: Inferring spatially distributed precipitation from gauge observations over complex terrain, *Water Resour. Res.*, **50**, 8589–8610, doi:10.1002/2014WR015442.

Lundquist, J. D., D. Cayan, and M. Dettinger, 2003: Meteorology and hydrology in yosemite national park: A sensor network application. *Inf. Process. Sens.*, 518–528

Lundquist, J. D. and B. Huggett, 2008. Evergreen trees as inexpensive radiation shields for temperature sensors, *Water Resour. Res.*, **44**, W00D04, doi:10.1029/2008WR006979.

Lundquist, J. D., M. Hughes, B. Henn, E. D. Gutmann, B. Livneh, J. Dozier, and P. Neiman, 2015: High-elevation precipitation patterns: using snow measurements to assess daily gridded datasets across the Sierra Nevada, California. *J. Hydrometeorol*. **16**, 1773–1792. http://dx.doi.org/10.1175/JHM-D-15-0019.1.

Mass, C. F., and Coauthors, 2003: Regional environmental prediction over the Pacific Northwest. *Bull. Amer. Meteor. Soc.*, **84**, 1353–1366, doi:10.1175/BAMS-84-10-1353.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based data set of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237–3251, doi:10.1175/ 1520-0442(2002)015,3237:ALTHBD.2.0.CO;2.

Minder, J. R., D. R. Durran, G. H. Roe, and A. M. Anders, 2008: The climatology of small-scale orographic precipitation over the Olympic Mountains: Patterns and processes. *Quart. J. Roy. Meteor. Soc.*, **134**, 817-839. doi: 10.1002/qj.258

Minder, J. R., P. W. Mote, and J. D. Lundquist, 2010: Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains. *J. Geophys. Res.*, **115**, D14122, doi:10.1029/2009JD013493.

Minder, J. R., and D. E. Kingsmill, 2013: Mesoscale Variations of the Atmospheric Snow Line over the Northern Sierra Nevada: Multiyear Statistics, Case Study, and Mechanisms. *J. Atmos. Sci.*, **70**, 916-938. doi: http://dx.doi.org/10.1175/JAS-D-12-0194.1

Oyler, J. W., S. Z. Dobrowski, A. P. Ballantyne, A. E. Klene, and S. W. Running, 2015: Artificial amplification of warming trends across the mountains of the western United States, *Geophys. Res. Lett.*, **42**, 153–161, doi:10.1002/2014GL062803.

Painter, T. H. and Coauthors, 2016: The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo, *Remote Sens. Environ.*, **184**, 139-152, doi: 10.1016/j.rse.2016.06.018

Raleigh, M. S., J. D. Lundquist, and M. P. Clark, 2015: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrol. Earth Syst. Sci.*, **19**, 3153-3179, doi:10.5194/hess-19-3153-2015

Rasmussen, R., and Coauthors, 2012: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc*., **93**, 811–829, doi:10.1175/ BAMS-D-11-00052.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp, doi:10.5065/D68S4MVH.

Snover, A.K, G.S. Mauger, L.C. Whitely Binder, M. Krosby, and I. Tohver, 2013. Climate Change Impacts and Adaptation in Washington State: Technical Summaries for Decision Makers. State of Knowledge Report prepared for the Washington State Department of Ecology. Climate Impacts Group, University of Washington, Seattle. 130 pp.

Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev*., **132**, 519–542.

Thompson, G., P. R. Field, W. R. Hall, and R. M. Rasmussen, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev*., **136**, 5095–5115.

Thornton, P. E., S. W. Running, and M. A. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol*., **190**, 214–251, doi:10.1016/ S0022-1694(96)03128-9.

Unsworth, M. H., and J. L. Monteith, 1975: Long-wave radiation at the ground, *Quart. J. Roy. Meteor. Soc*., **101**, 13-24

U.S. Army Corps of Engineers, 1956: Snow Hydrology: Summary Report of the Snow Investigations. North Pacific Division, Corps of Engineers, 462 pp.a

Wayand, N. E., A. F. Hamlet, M. Hughes, S. I. Feld, and J. D. Lundquist, 2013. Intercomparison of meteorological forcing data from empirical and mesoscale model sources in the North Fork American River Basin in Northern Sierra Nevada, California, *J. Hydrometeorology*, **14**, 677–699. doi: http://dx.doi.org/10.1175/JHM-D-12-0102.1

Wayand, N. E., A. Massmann, C. Butler, E. Keenan, and J. D. Lundquist, 2015: A meteorological and snow observational data set from Snoqualmie Pass (921 m), Washington Cascades, U.S. *Water Resour. Res.,* **51**, doi:10.1002/2015WR017773

Wayand, N. E., J. Stimberis, J. P., Zagrodnik, C. F. Mass, and J. D. Lundquist, 2016a: Improving simulations of precipitation phase and snowpack at a site subject to cold air intrusions: Snoqualmie Pass, WA. *J. Geophys. Res. Atmos.*, **121**, 9929-9942, doi:10.1002/2016JD025387

Wayand, N. E., M. P. Clark, and J. D. Lundquist, 2016b: Diagnosing snow accumulation errors in a rain-snow transitional environment with snow board observations. *Hydrol. Processes*, doi: : 10.1002/hyp.11002

Appendix 1: Adjusted Model Parameters:

Model Parameter	**Observed Relative Humidity**	***q* from WRF**	**Observed Relative Humidity**	***q* from WRF**
Rain vs. Snow Parameters	-	-	-	-
tempCritRain [°C]	-1.0	-1.0	-1.0	-1.0
tempRangeTimestep [°C]	2	2	2	2
Albedo Parameters	-	-	-	-
albedoDecayRate [s]	500,000	500,000	500,000	500,000
Density Parameters	**Wayand *et al.* (2016) (ρ_w)**	**Wayand *et al.* (2016) (ρ_w)**	**Hedstrom and Pomeroy (1998) & Anderson (*1976*) (ρ_{HPA})**	**Hedstrom and Pomeroy (1998) & Anderson (*1976*) (ρ_{HPA})**
newSnowDenMin [kg m^{-3}]	100	100	67.92	67.92
newSnowDenMult [kg m^{-3}]	50	50	51.25	51.25
newSnowDenScal [°K]	1	1	2.59	2.59
denScalOvrbdn [kg^{-1} m3]	0.02	0.02	0.023	0.023
tempScalOvrbdn [$°K^{-1}$]	0.06	0.06	0.08	0.08

Table A1: Adjusted model parameter decisions. The albedo decay rate was fit to observations of albedo at Snoqualmie Pass, WA (Wayand *et al.* 2015). All other model parameters were taken as default values.

Appendix 2: Model Forcing Variables

Forcing Variable	Source of Model Forcing Variable		Reason
	Set 1	**Set 2**	
Precipitation [kg m^{-2} s^{-1}]	Daily Observed Values Uniformly Distributed	--	Observed
Temperature [°K]	Observed Corrected for using equation 1	--	Biased SNOTEL Temperature Sensors
Specific Humidity (*q*) [g/g]	Observed Relative Humidity Converted to q using observed T and WRF Pressure	WRF	Possible Hysteresis and model biases
Shortwave Radiation [W m^{-2}]	MTCLIM v. 4.2 $f(T_{obs.}, \underline{RH_{Obs.}}, Ppt_{Obs.}$, Waterhole Wind Speed data)	MTCLIM v. 4.2 $f(T_{obs.}, \underline{RH_{WRF.}}, Ppt_{Obs.}$, Waterhole Wind Speed data)	Evaluated at Snoqualmie Pass for performance. Albedo decay rate also calibrated based on observed albedo.
Longwave Radiation [W m^{-2}]	Dilley and O'Brien (1998) with Unsworth and Montieth (1975) Cloud correction $f(T_{obs.}, \underline{RH_{Obs.}}$ SW)	Dilley and O'Brien (1998) with Unsworth and Montieth (1975) Cloud correction $f(T_{obs.}, \underline{RH_{WRF.}}$ SW)	Evaluated at Snoqualmie Pass for performance.
Pressure [Pa]	WRF	--	
Wind Speed [m s^{-1}]	Waterhole Observed Hourly Averaged Wind Speed	--	Using a constant 2 m s^{-1} wind speed caused too much melt during ablation season. L_e+H in the same direction.

Table A2: SUMMA Model Forcing Data used in the evaluation phase. "--" indicates that the model forcing variable in set 2 is the same as it is in set 1. The major difference between set 1 and set 2 is the difference in specific humidity (*q*), which is then followed through into the calculation of shortwave radiation and longwave radiation.

Appendix 3: Model Configuration

SUMMA model options used in all simulations are reported below in Table A3. See Table 1 of Clark *et al.* (2015b) for a complete description of SUMMA modeling options.

Table A3. SUMMA model configurations used in all simulations.

Summa Decision (short name)	Summa Decision (definition)	Default option used in study (short name)	Default option used in study (definition)
soilCatTbl	soil-category dateset	ROSETTA	merged Rosetta table with STAS-RUC
vegeParTbl	vegetation category dataset	USGS	USGS 24/27 category dataset
soilStress	choice of function for the soil moisture control on stomatal resistance	NoahType	thresholded linear function of volumetric liquid water content
stomResist	choice of function for stomatal resistance	BallBerry	Ball-Berry
num_method	choice of numerical method	itertive	iterative
fDerivMeth	method used to calculate flux derivatives	analytic	analytical derivatives
LAI_method	method used to determine LAI and SAI	specified	LAI/SAI computed from green vegetation fraction and winterSAI and summerLAI parameters
f_Richards	form of Richard's equation	mixdform	mixed form of Richards' equation
groundwatr	choice of groundwater parameterization	noXplict	no explicit groundwater parameterization
hc_profile	choice of hydraulic conductivity profile	constant	constant hydraulic conductivity with depth
bcUpprTdyn	type of upper boundary condition for thermodynamics	nrg_flux	energy flux
bcLowrTdyn	type of lower	zeroFlux	zero flux

	boundary condition for thermodynamics		
bcUpprSoiH	type of upper boundary condition for soil hydrology	liq_flux	liquid water flux
bcLowrSoiH	type of lower boundary condition for soil hydrology	drainage	free draining
veg_traits	choice of parameterization for vegetation roughness length and displacement height	CM_QJRMS1998	Choudhury and Monteith (QJRMS 1998) "A four layer model for the heat budget..."
canopyEmis	choice of parameterization for canopy emissivity	difTrans	parameterized as a function of diffuse transmissivity
snowIncept	choice of parameterization for snow interception	lightSnow	maximum interception capacity an inverse function of new snow density
windPrfile	choice of wind profile through the canopy	logBelowCanopy	logarithmic profile below the vegetation canopy
astability	choice of stability function	mahrtexp	Mahrt (1987) exponential function
canopySrad	choice of canopy shortwave radiation method	CLM_2stream	Community Land Model
alb_method	choice of albedo representation	conDecay	constant decay rate (e.g., BATS approach, with destructive metamorphism + soot content)
compaction	choice of compaction routine	anderson	semi-empirical method of Anderson (1976)
snowLayers	choice of method to combine and sub-divide snow layers	CLM_2010	CLM option: combination/sub-division rules depend on layer index
thCondSnow	choice of thermal conductivity	jrdn1991	Jordan (1991)

	representation for snow		
thCondSoil	choice of thermal conductivity representation for soil	funcSoilWet	mixture of constituents
spatial_gw	choice of method for the spatial representation of groundwater	localColumn	separate groundwater representation in each local soil column
subRouting	choice of method for sub-grid routing	timeDlay	time-delay histogram

Appendix 4: LiDAR Snow Depth comparison to Time-Lapse Snow Depth

A comparison on 8-9 February 2016 between a 3600 m^2 area around the snow depth pole from the LiDAR and our snow depth pole measurements showed that both measurements generally agree or are within each other's interquartile (ASO) or uncertainty range (Snow Depth Poles) (Figure A1). At many sites such as Mount Seattle East, Mount Seattle West, and West of Lake Lacrosse, the difference from the snow depth pole and the median ASO value was less than 15 cm. At other sites, more significant differences appeared. For instance, at Mount Hopper, camera images showed a significant snowdrift, which was formed due to preferential deposition of precipitation. Since the snow depth poles were located outside the snowdrift the median ASO value was higher than the snow pole measurement by 83.5 cm in February 2016 and 172.9 cm in March. In contrast to this, at Black and White West the ASO snow depth maps indicated that our snow depth poles were located within a snowdrift, and therefore the median ASO value was 41.8 cm lower in February and 59.5 cm lower in March when compared to the snow depth pole measurement.

At these two sites we found that the median ASO value from a larger spatial area (>3600 m^2) was more similar to the median ASO value from the 3600 m^2 area than it was to the snow depth pole measurement. This indicates to us that the median ASO value within a 60-m bounding box was a better representation of the snow within this region than the snow depth poles. Furthermore, by March many snow depth poles contained significant uncertainty in their measurements because they became bent or buried with snow. Therefore, we chose to compare our model simulations with different sources of precipitation, on an annual basis, to the median March ASO value within a 60-meter bounding box, rather than the snow depth pole measurements.

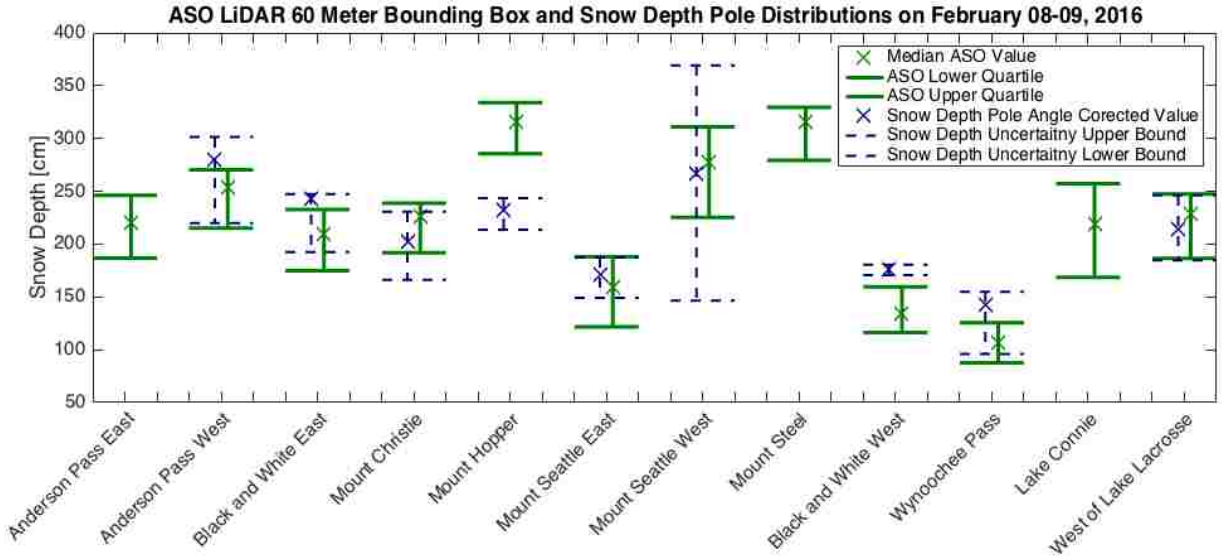

Figure A1: The ASO LiDAR distribution of snow depth values within a 60-meter bounding box (3600 m^2) and the snow depth pole measurements with their associated uncertainty on 8-9 February 2016. Sites that do not have blue bars (dotted lines) shown are sites with no snow depth measurement at this time because the poles became buried with snow.

Chapter II

A cost effective and low impact measurement technique for annual snow depth

1. Introduction

Recently, time-lapse photography has been proposed as an innovative way to measure snow depth (Floyd and Weiler, 2008, Parajka *et al.* 2012, Gavelmann *et al.* 2013, Fortin *et al.* 2015). The advantage of time-lapse photography for measuring snow depth is that it is inexpensive, costing at a minimum of around $150 per location, and that it can provide a time-series of snow depth as long there is sunlight. It also has a low impact on the environment, allowing for it to be set up in remote locations and in protected wilderness areas. Furthermore, time-lapse photography can be installed in the summer months and does not require maintenance during the fall, winter, or spring months.

Published methodologies and associated uncertainties to obtain snow depth from photographs vary. For instance, Nievinski and Larson (2014) read the pole "manually," by looking at pole markings relative to snow depth and found the uncertainty in their measurements to be ± 3 cm. Parajka *et al.* (2012) developed a fully automated algorithm that identifies and counts the number of evenly spaced markers on their snow depth pole. Floyd and Weiler (2008) were able to create an automated linear mask that identified the number of pixels between the top and bottom-most visible part of the snow depth pole. This mask was then used to determine the length represented by a single pixel, which was then multiplied by the total number of pixels between the top of bottom-most visible part of the pole to determine the snow depth.

Gavelmann *et al.* (2013) proposed a similar methodology to Floyd and Weiler (2008) but instead of using a linear mask to determine the top and bottom of the pole, they manually selected the pixel coordinates at the top and bottom. However, Gavelmann *et al.* (2013) noted a consistent offset in their measurements compared to manual measurements as snow depth increased. Therefore, they resorted to manual measurements, similar to Nievinski and Larson (2014) for the rest of their study.

The remainder of this chapter is organized as follows: In section 2 we further describe our installation set up in the Olympic Mountains for the study in Chapter I and introduce the errors that we came across in water year (WY) 2015 and 2016. In Section 3 we describe a snow depth measurement technique similar to that of Garvelmann *et al.* (2013) but that also accounts for perspective distortion. In section 4 we show the results from an evaluation of this tool. In section 5 we describe how to angle correct for bent poles and provide uncertainty with this measurement. In Section 6 we provide suggestions on how to set-up a time-lapse photography network.

2. Installation

A total of around 1000 kilometers were hiked amongst five people who carried the installation supplies within or on their backpacks (Figure 1).

Figure 1: Colin Butler, Max Mozer and Justin Pflug (from right to left) carrying the necessary equipment to measure snow depth in remote locations during the OLYMPEX campaign.

Backpack weights ranged from 27 to 32 kilograms. The entire installation process took around four weeks to complete, making labor the most significant financial cost of a snow depth measurement network. All installation equipment is shown in Figure 2. We note that our camera network did not provide data in real-time, making annual maintenance trips necessary.

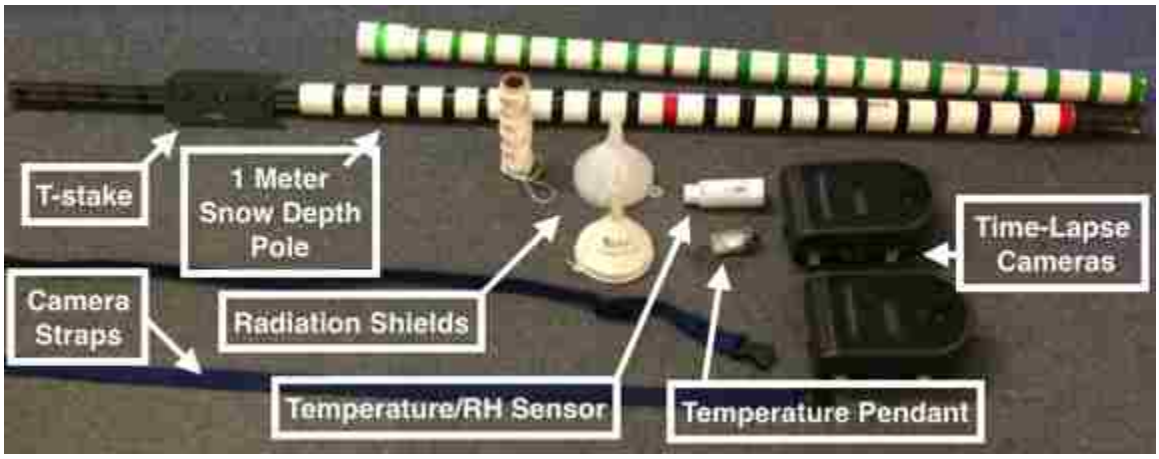

Figure 2: An example of the supplies that were installed at a snow depth measurement site. Note that in WY 2016 red tape was replaced with neon orange/pink tape as seen in Figure 1.

Cameras were strapped to trees roughly 2.8 to 6.8 m above the ground depending on the site's elevation and the expected amount of snowfall. Historic snow course measurements in this area reported snow depth values of up to 7.5 m. Since the cameras were placed high in the tree, wooden shims were used to tilt the camera into view of the snow depth poles. The declination angle from a plane that runs normal to the surface ranged from 0º to 33º and was dependent on the distance from the pole to the camera, the local topography, and the height of the camera in the tree. Since the cameras were placed high in the trees, they were subject to movement associated with tree sway.

Snow poles ranged in height from 1 to 6 m depending on elevation and anticipated snow depth. In WY 2015, black tape marks were placed every 5 cm, and red tape marks were placed every 50 cm. Poles were designed to allow the person processing the data to count the number of black tape marks from the red tape closest to the top of the snowpack. However, we found that when the camera was placed at approximately 11 m from the pole, it became nearly impossible to distinguish the red tape from the black tape, and thus the person processing the data could no longer manually read the snow depth poles to obtain a manual measurement without counting every piece of black tape, every hour there was an image. To address this issue we developed a semi-automated method for measuring snow depth from time-lapse photographs. This method is described in section 3, and an evaluation is shown in section 4.

In WY 2016 we corrected for this by placing neon orange and neon pink tape over the red tape. This allowed for the 50 cm marker to maintain its true color in low light conditions and for the poles to be manually read by eye similar to Nievinski and Larson (2014). However, in WY 2016 we had substantially more snow than in WY 2015, and the poles became bent or buried due to snow creep. We therefore developed a method that allowed us to correct for the angle of the poles (section 5). We note that when installing our snow depth poles, we specifically looked for areas that were flat. However, in environments with deep snowpacks, even slight slopes can cause the snow to creep.

3. WY 2015 – Semi Automated Method

First we identified the coordinates for the bottom-most visible part of the pole by zooming in on the image. If there was no snow, then we flagged this image as having no snow. At these times this provided a measurement of truth and allowed us to evaluate errors in our calibration. After we identified the bottom of the pole in all of the images, we repeat, for the top part of the pole. Then the number of pixels from the bottom of the pole to the top of the pole was calculated using the distance formula (1).

$$Pixel\ Distance = \sqrt{(x_t - x_b)^2 + (y_t - y_b)^2} \quad (1)$$

where x_t and x_b correspond to the x-coordinates for the top and bottom of the pole, respectively. Similarly, y_t and y_b correspond to the y-coordinates at the top and bottom of the pole, respectively.

Lastly, we calibrated the measurements so that the pixel distance between the top and bottom of the pole could become a snow depth measurement. Our program was unique because it calibrated the measurements by identifying the pixel coordinates associated with the pole's tape marks from the bottom portion of the pole to the top of the pole (Figure 3 a.). The blue dots show

the location of where the user clicked to identify the pixel coordinates of the tape marks for the bottom 150 cm on an example image. These clicks generated a look-up table, which associated the pixel distance to the known measurement on the pole. For instance, the orange arrows in Figure 3 a. show that 100 cm from the bottom is associated with a pixel distance of 1128.3 pixels from the top of the pole. Therefore an image that produced a distance close to 1128.3 pixels from the top of the pole was assigned a snow depth measurement of 100 cm. To refine our measurements we linearly interpolated between each tape mark to produce a look-up table that had pixel distances associated with 0.5 cm intervals.

To evaluate camera movement throughout the year we superimposed other calibration markers onto the image from September 13, 2014. For instance, the other markers correspond to the same black tape marks but on different dates (see legend). This shows that the camera was not fixed, and that the camera moved randomly with the tree throughout the winter due to interactions with the wind and snow. Furthermore, it appeared that the camera began to slide down the tree later in the season. Notice that the same movement was also found at the top of the pole and not just the bottom (Figure 3 b.).

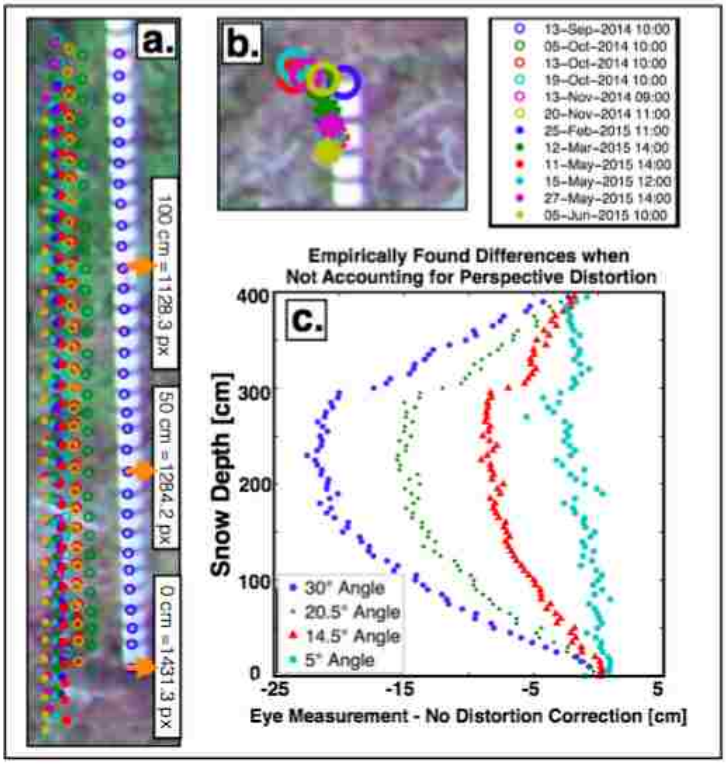

Figure 3: a.) Calibration locations at various times in the year along the bottom 150 cm of a 400 cm pole. Background image was taken on September 13, 2014 at 10:00 AM PDT. Orange arrows indicate the height above the ground for 0, 50 and 100 cm markers along the snow depth pole and the associated number of pixels from the top of the pole. b.) The location of the top of the pole on the same dates. c.) Empirically found differences if distortion is not corrected for from the true snow depth based on a 400 cm snow depth pole.

The look-up table allowed for the program to empirically correct for perspective distortion, which was often found on cameras that had a declination angle of more than 10º. We derived differences from using a single length per pixel instead of the lookup table to calibrate the pixel distance calculations to snow depth measurements. These empirically-derived differences were a

function of declination angle and the snow depth. These were found based on lab experiments, where we varied the camera declination angle and then used a single length per pixel value, as in Garvelmann *et al.* (2013), to calibrate our measurements at 5 cm intervals. We then subtracted these values from the known measurement on the pole. We found when the camera has a 30º-declination angle that using a single length per pixel value will cause the measured snow depth to be approximately 20 cm less than the true snow depth at 200 cm. As the declination angle decreases or snow depth decreases, distortion has less of an impact (Figure 3 c.). The results shown here likely explain the biases that Garvelmann *et al.* (2013) experienced in their study.

When we conducted the semi-automated process described above, it became apparent that many images were not visible due to condensation, while other images displayed physical conditions that led us to make a series of subjective decisions (Figure 4). To address theses issues, our program flags each type of subjective measurement (Figure 4 a-d) and also flags images that were not visible due to fog or condensation on the camera.

We believe the remaining uncertainty within this technique comes from random tree movement throughout the year as well as tall poles (>=5 m) bending in the wind. Our program asks the user to produce multiple calibrations and use the measurements for when there isn't any snow present. Here the program calculates a mean difference throughout the year that is near zero centimeters. To avoid poles that bended throughout the year, we avoided six-meter tall poles that did not have the support from a t-stake or snow (in the low snow year) and thus moved with the wind.

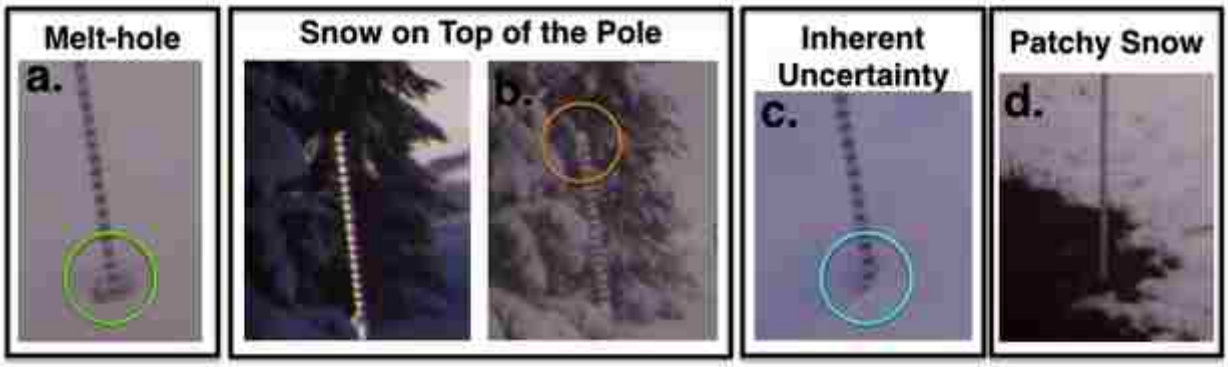

Figure 4: a.) Melt hole during the ablation season can increase the uncertainty. b.) Snow on top of the pole could cause an increase in 5-10 cm, depending on the height of the snow on top of the pole, and whether the user recognized that this was snow. Note that the image on the left, without the orange circle, had no snow on top of the pole. c.) Similar to the melt hole, the users interpretation could cause a 5-7.5 cm difference in measured snow depth. This differs from the melt hole in that this is during a period of snow accumulation, not melt. d.) Here, there is no snow at the bottom of the pole but there is still snow in the area. The user should determine whether the pole was designed for a point measurement or if the time-lapse cameras were designed to measure snow depth in the visible domain.

3. WY 2015 Semi Automated Method Results

We used snow depth measurements from Mount Seattle West and Black and White East to compare images that were processed by eye to images that were processed using the semi automated method (Figure 5). We selected these two time-series because they could be processed "manually," where as the remaining images were taken by cameras located too far from the pole for the red tape marks to be identified. These two cameras had a declination angle of 32º and 10º, respectively, which allowed us to evaluate our calibration method.

The blue violin plots show the relative frequency of using the semi-automated method to receive a given mean difference from an eye measurement (Figure 5). Red violin plots show the relative frequency of using the semi-automated method to estimate zero snow depth on the dates when we said there was no snow. Blue violin plots therefore only analyze times when there was snow on the ground. The distribution of different mean differences comes from different combinations of a given number of calibrations. For example, at Black and White East, there were 13 total calibrations. We chose to average 7 of the 13 calibrations. There were thus 3432 unique combinations of how we could average the 7 of 13 calibrations. As the number of calibrations is increased the likelihood of any bias in the snow depth measurement from the semi-automated method decreases to ~ 0 cm. The program continued to calibrate the snow depth measurements based on randomly selected images when there was no snow on the ground until the mean difference from all calibrations was less than 1 cm.

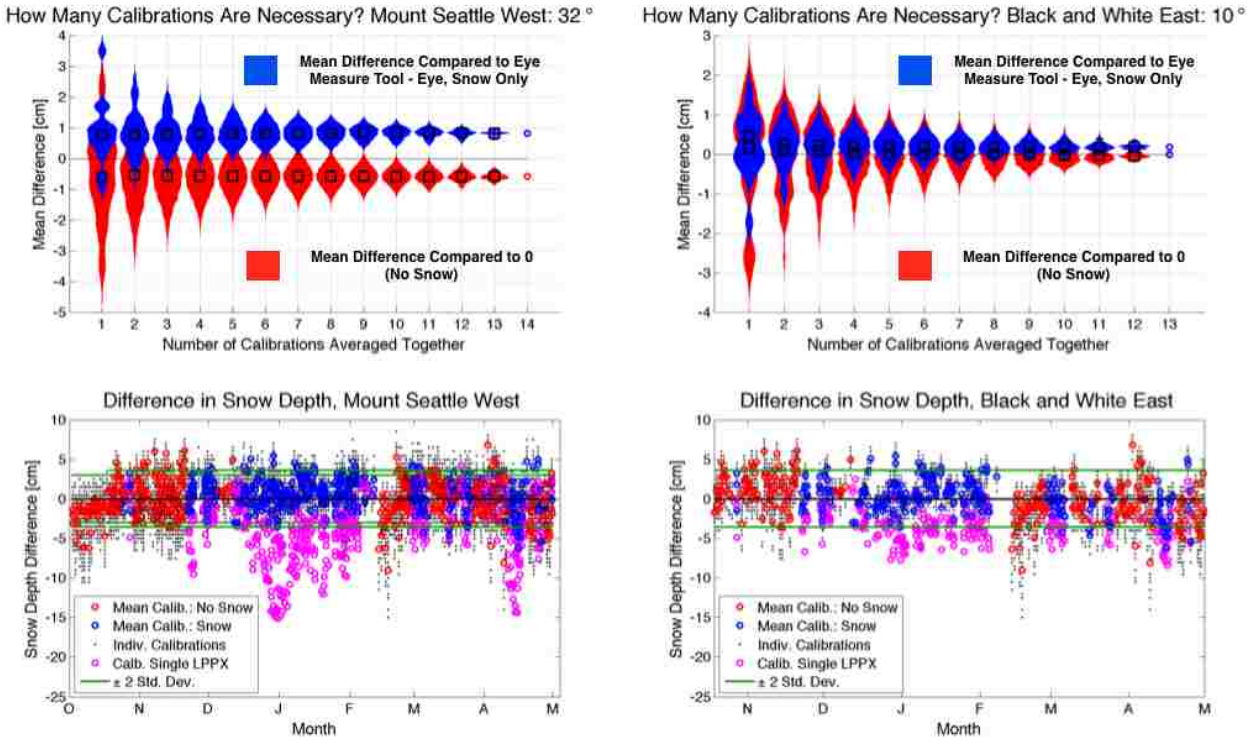

Figure 5: Top pannel) Violin plots demonstrating that as the number of calibrations is increased the likelihood of having any bias in our measurements from the semi-automated program is decreased. Bottom Pannel: Individual differences from the semi-automated tool snow depth measurement or zero if there was no snow.

The bottom panel (Figure 5) shows the individual differences from zero when there was no snow (red) and the differences from an eye measurement when there is snow (blue) after averaging together the optimal number of calibrations (14 at Mount Seattle West and 13 at Black and White East). The black dots show the individual differences that arise from each calibration. Magenta dots show the results from using a single length per pixel value similar, to Garvelmann *et al.* (2013). Note that large differences appeared only when the snow depth was significant (Late December – Mid January) and when we had a higher camera declination angle. This is consistent with findings from the lab experiment that derived differences from known measurements on the pole (Figure 3.c.).

Therefore the calibration process, presented herein, was essential and the more calibrations the user did, the higher probability that the mean difference of the true measurement was close to 0 cm (Figure 5). The mean difference for when there was snow was 0.83 cm and 0.18 cm for Mount Seattle West and Black and White East, respectively. The standard deviation for when there was snow was 1.92 cm and 1.79 cm for Mount Seattle West and Black and White East, respectively. We therefore determine that the 95% confidence interval in the semi-automated method was ± 4 cm. This uncertainty was only ± 1 cm greater than that of Nievinski and Larson, (2014) who processed time-lapse images by eye. The red dots, which show the calculated snow depth measurement when there wasn't any snow was forced to zero in practice but was shown to exemplify that there are random errors that are similar in nature to the ones that appeared when there was snow.

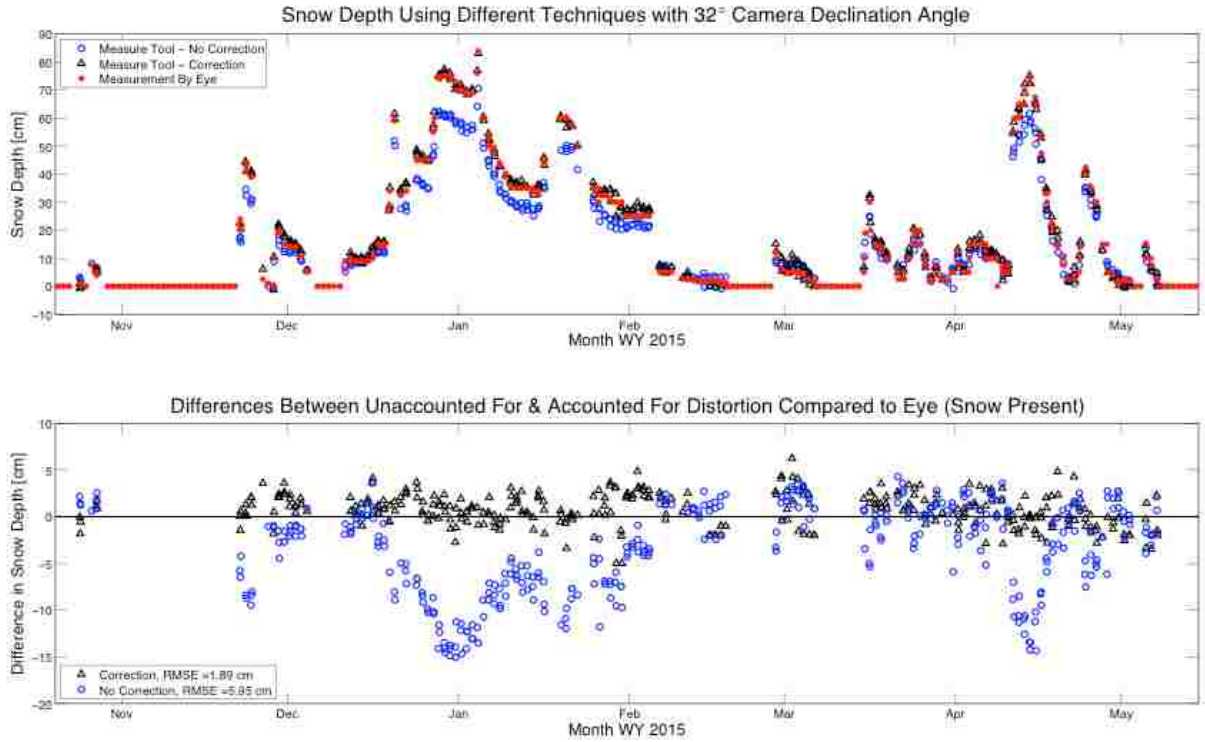

Figure 6: Evidence that this technique corrects for the perspective distortion, which can occur if only a single length per pixel value was used. For both the corrected and not corrected images 14 random calibrations were done and the mean was taken. The second plot shows the difference from eye for both techniques when there was snow. Large differences from an eye measurement occur when snow depth increases and distortion is not accounted for (Blue). The differences greater than 5 cm are systematic errors stemming from the methodology that used a single length per pixel.

5. WY 2016 – Angle Corrections and Associated Uncertainty

a. Measuring pole angles and computing angle corrected snow depth

Many snow depth poles upon retrieval in WY 2016 were bent due to snow creep (Figure 7). This resulted in inaccurate measurements of snow depth when measuring the poles by eye. To improve snow depth measurements, it was necessary to account for how the poles were bent at different times throughout the year.

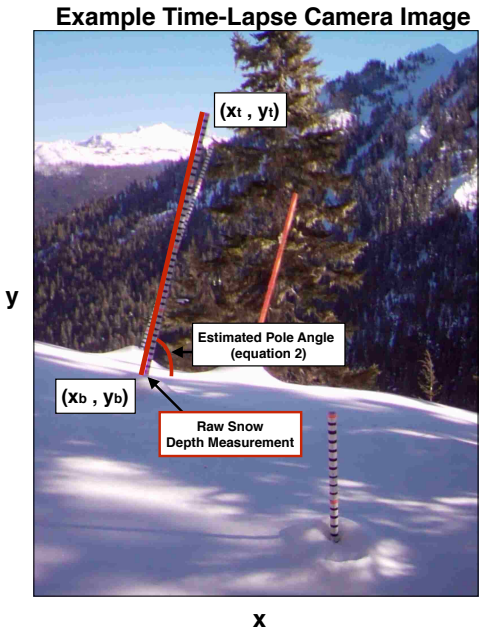

Figure 7: Example time-lapse image of the view from Camera 2 at Mount Seattle East

In WY 2016, images of the bent snow depth poles were processed to retrieve raw snow depth measurements by counting tape marks from the top of the pole. Afterwards images were reviewed again to quality control any erroneous measurements. Then the images were looped through to provide the pixel coordinates from the top and bottom most visible part of the pole. Using the determined coordinates (Figure 7), an estimated angle for the pole was found (2):

$$estimated\ pole\ angle\ clicked = \left| \tan^{-1}\left(\frac{y_t - y_b}{x_b - x_t}\right) \right| \quad (2)$$

With the estimated pole angle, we corrected the snow depth measurement (3):

$$angle\ corrected\ snow\ depth = raw\ snow\ depth * \sin\left(estimated\ pole\ angle\right) \quad (3)$$

In this process, we made three assumptions when interpreting camera images. The major assumption was that the poles were bent at a constant angle, meaning the pole's angle from vertical under the snow was identical to the angle observed above the snow. This assumption may not hold true at higher pole angles (>50°) when the pole likely bends with curvature rather than along a straight line. We also made the assumption that the camera provided a constant fixed frame of reference when capturing images of the poles. Lastly, we made the assumption that the poles were bent in the same azimuthal direction, the entire year. The azimuthal direction was determined when the time-lapse camera images were retrieved. However, in some cases the poles were standing straight, pointed towards zenith when we arrived. Here, we estimated the azimuth direction that the poles were bent using the measured azimuth from the cameras and by using time-lapse images from multiple cameras to determine the azimuthal direction the pole was bent towards. This led to a large source of uncertainty in the estimated azimuthal direction at some sites. Sites where this was estimated were documented and included in the metadata.

b. Quality Controlling the Data

False accumulation events occurred when correcting for the angle of the snow depth pole during the ablation season. False accumulation events occurred when the pole became less bent but the actual snow depth remained constant, meaning the algorithm saw a straighter pole and therefore the measured snow depth increased. Since snow depth was not increasing at that time in reality, the false accumulation events were quality controlled. This involved selecting two points in the time series that were outside the false accumulation event. From here, the ablation rate between these two points from the raw data time-series was translated to fit the time series that had corrected for snow depth.

c. Creating Uncertainty Bounds for our final data set.

There are two different sources of uncertainty in the snow depth measurements. The first source of uncertainty was from measuring the poles by eye. Since the poles had tick marks every 5 cm, all of the non-zero measurements were subject to a ± 3 cm uncertainty, to account for human error and inherent uncertainty in the measurement itself (Figure 4, a&c).

The second source of uncertainty came from the methods used to angle correct the poles. When a pole was bent due to snow creep, it bent in a certain azimuthal direction (Figure 8).

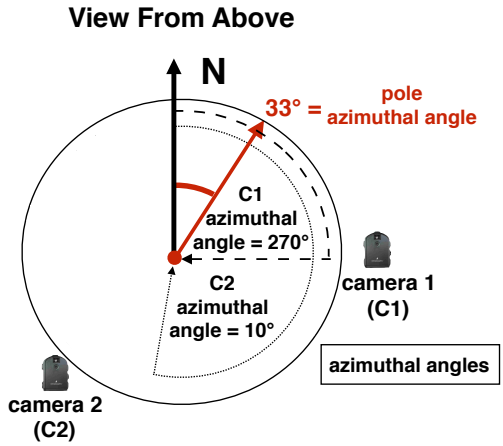

Figure 8: Conceptual diagram of the azimuthal directions for the camera set up in Figure 7.

To determine how bent a pole was at a given time, the pole should ideally be viewed from an angle that is orthogonal to the direction the pole is bent. Due to the setup of the snow depth measurement sites, the view of the poles were limited to angles at which the cameras were pointed. This setup resulted in most cameras not being able to fully capture the bend of a pole from an orthogonal view. We calculated the number of degrees the camera was from orthogonal (DFO) as follows (4 & 5):

$$Differnce\ in\ Azimuth = DIA = |Pole\ Azimuth - Camera\ Azimuth| \quad (4)$$

$$DFO = \begin{cases} 90° - DIA,, & 0° < DIA < 90° \\ DIA - 90°, & 90° < DIA < 180° \\ 270° - DIA, & 180° < DIA < 270° \\ DIA - 270°, & 270° < DIA < 360° \end{cases} \quad (5)$$

In order to quantify the potential error from not having an orthogonal view of the bent poles, a lab test was performed. In this test, a pole was set up at a known angle so that the pole angle and azimuth at which it was leaning were controlled. Images of the pole were taken from different camera view angles (Figure 9).

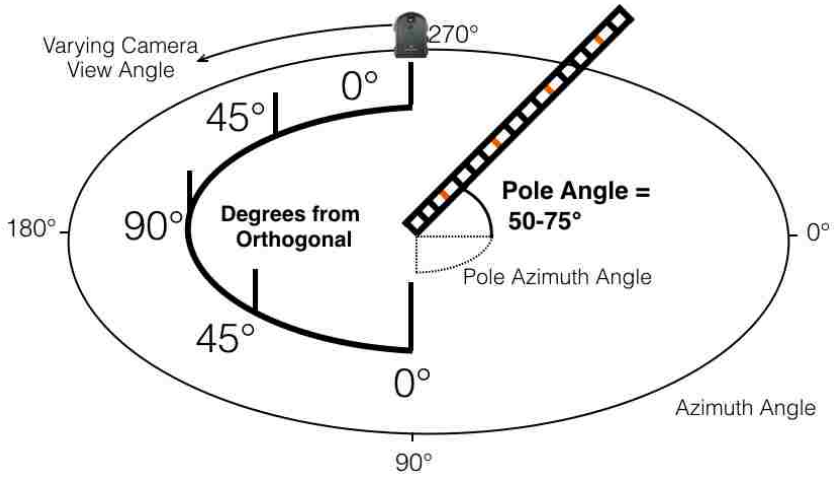

Figure 9: Conceptual figure of the lab experiment from an oblique view.

In the lab experiment, the camera view angle started at 0 degrees from orthogonal (orthogonal to the pole azimuth angle) and ended at 90 degrees from orthogonal (directly in line with the pole azimuth angle), using 10 degree increments. Images were repeated going from 90 degrees to the opposite 0 degrees from orthogonal, so that we had two images for each azimuth angle. The angle in each image was then estimated (2), and the difference from the actual angle was computed and plotted against camera azimuth angle (Figure 10). This process was performed for pole angles of 50-75 degrees in 5-degree increments (with 90 degrees being a pole pointed towards zenith). The 50-75 degree range was chosen because when the pole was angled at greater than 75 degrees, the pole was almost vertical, so the pole angle had a negligible impact when measuring the snowpack. When a pole was bent at more than 50 degrees, the assumption that the pole was bent at a constant angle, above and below the snow surface, appeared to be false based on visual inspection. Therefore the data during these periods was not used in our analysis.

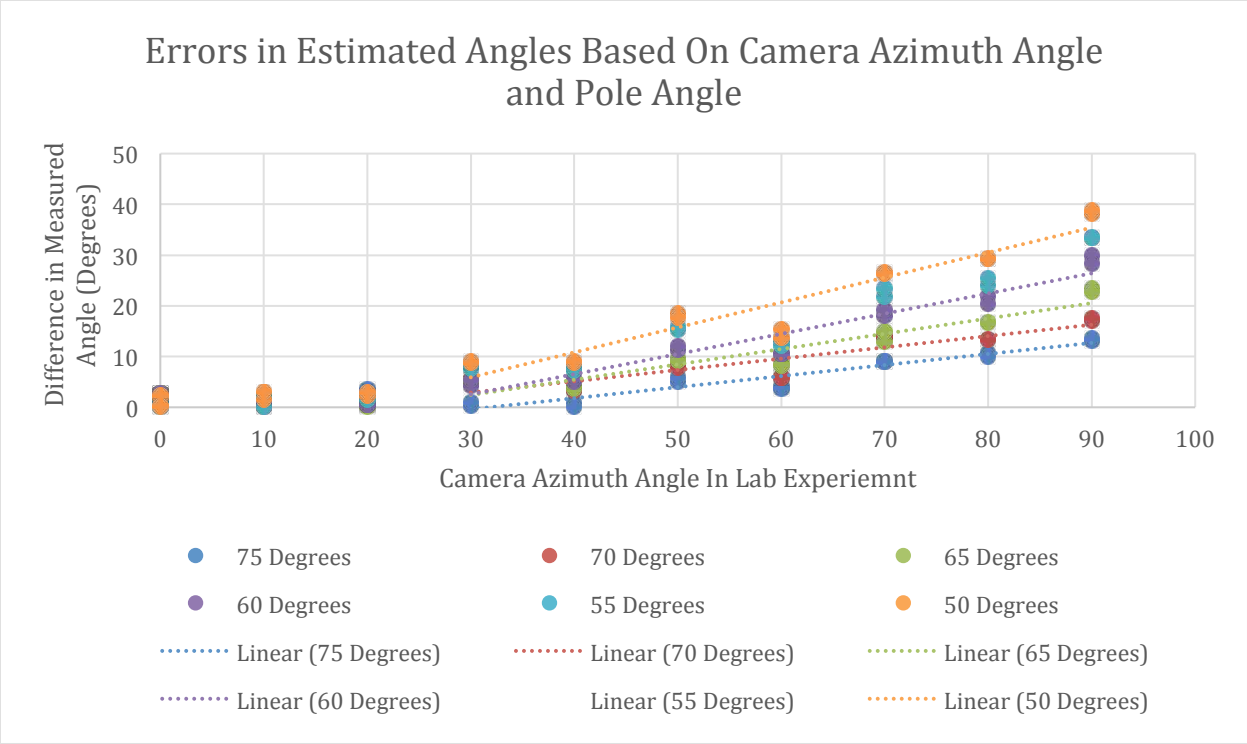

Figure 10: Difference in estimated clicked angle (pole angle estimate from clicking minus actual angle) vs. the camera azimuth angle. Each line or series of data points represents a different pole angle.

Errors in estimated angles were minimal when the camera view angle was within 20 degrees of orthogonal from the direction at which the pole was bent. However, beyond 30 degrees, there was a linear trend in the errors for each unique pole angle from 50-75 degrees. When the pole angle was greater than 75 degrees, the uncertainty from the pole angle was taken to be 0 cm as the uncertainty was similar to the already accounted for natural processing uncertainty (±3 cm).

When the pole angles were less than 75 degrees and the camera view angle was not within 25 degrees of an orthogonal view of the pole, we estimated the uncertainty using the results from the lab experiment. We developed a set of least squares regressions (Table 1) between the degrees the camera view angle was from orthogonal (Figure 9) and the estimated pole angle clicked (2) for each actual pole angle (Figure 11).

Actual Pole Angle	Regression Equation	R^2
75°	EAA = 0.19 * DFO + 70.17	0.91
70°	EAA = 0.21 * DFO + 66.93	0.93
65°	EAA = 0.26 * DFO + 70.00	0.94
60°	EAA = 0.33 * DFO + 55.22	0.93
55°	EAA = 0.36 * DFO + 51.93	0.94
50°	EAA = 0.41 * DFO + 46.81	0.94
ECA = Estimated Actual Angle & DFO = Degrees From Orthogonal		

Table 1: Fitted least squares regressions between the estimated angle clicked and the degrees the camera is from orthogonal.

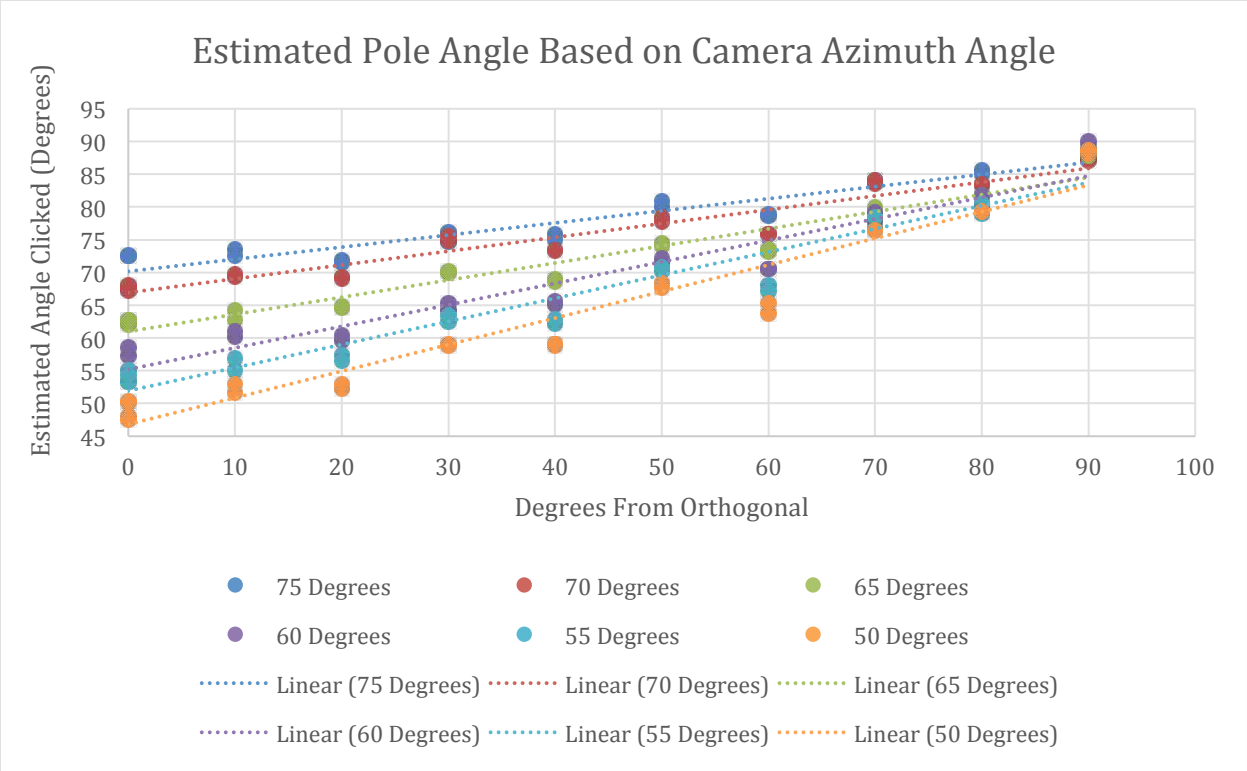

Figure 11: Linear relationships between the estimated angle clicked (2) and the DFO developed from lab experiments for each known pole angle (50-75 degrees – see legend). Note that when the camera view angle is 0 degrees from orthogonal the pole angle is near equal to the estimated pole angle.

Once the regression lines were determined, it was possible to estimate the actual pole angle for a given moment in time. The known DFO was plugged into each linear regression to generate a series of estimated actual pole angles. The absolute value of the difference between the estimated clicked angle (2) and the values from each regression line at the given DFO was taken. Then the two smallest absolute differences were stored. This represented the two regression lines between which the estimated clicked angle fell for a given DFO (Figure 11). We then linearly interpolated between the two regression lines to determine a more refined (<5° increment) estimated actual pole angle, from which, we estimated the correction error factor (6).

$$correction\ error\ factor = \sin(estimated\ clicked\ angle) - \sin(estimated\ actual\ angle). \quad (6)$$

The uncertainty was then estimated (7)

$$uncertainty = correction\ factor\ error * raw\ snow\ depth \quad (7)$$

The estimated uncertainty was then added to the ±3 cm from human uncertainty to create a total uncertainty. The total uncertainty for each individual pole is shown in Figure 4 of Chapter I of this thesis. We reiterate from Chapter I that the uncertainty in snow depth was at many times significant (>±15 cm). Therefore we did not use angle-corrected data when evaluating estimates of frozen precipitation. Instead we used the corrected snow depth measurements and their associated uncertainty as a guide for the snow depth accumulation throughout the year (gray

shading in Chapter I, Figure 4). We did use the uncertainty and corrected snow depth measurements to compare with the interquartile range from the Airborne Snow Observatory data (Chapter I, Appendix 4).

6. Discussion and Conclusions

We presented the methods that we used to measure the snow depth poles for both WY 2015 and 2016 during the OLYMPEX ground validation campaign. We found that previous studies, which used a single length per pixel value in automated snow depth readings from time-lapse photography may suffer from significant errors depending on the snow depth and the camera declination angle. We also found a dramatic improvement in the ease of measuring snow depth from time-lapse cameras when using neon pink and orange tape on our snow depth poles instead of red tape. This was because the red tape became indistinguishable from the black tape in low light conditions in instances where the camera was placed far (more than ~11 m) from the pole. Lastly, we demonstrate that snow depth poles can become significantly bent throughout the winter due to snow creep. This translates into a large uncertainty range at many locations because the camera azimuth angle was not orthogonal to the direction that the pole is bent. We therefore, despite these efforts, did not use the snow depth measurements to evaluate precipitation estimates in WY 2016 once the poles became bent. In WY 2015 the poles did not become bent because snow depth was abnormally low.

We suggest that future time-lapse photography networks for snow depth find locations that are protected from snow creep and use orange/pink tape over the red tape. We also suggest the use of fully automated methods, as manually measuring snow depth from time-lapse images is a time consuming process. Preliminary work suggests using a fully painted orange pole in order to distinguish the pole from the background environment. In our setup the background was often black (trees) or white (snow) making it difficult to visualize or detect any contrast between certain segments of the pole and the background. Furthermore, it is critical that the cameras have a declination angle of less than 10°, making it difficult to use the orange poles in our study, as these measurements would need to be corrected for perspective distortion and (Section 3 and 4) for the angle at which the poles were bent (section 5). Using orange poles thus poses a challenge for deep snowpacks, where the cameras are forced to be placed high in a tree and are therefore angled down to ensure the poles are in the camera's field of view.

7. References

Floyd, W., and M. Weiler, 2008: Measuring snow accumulation and ablation dynamics during rain-on-snow events: innovative measurement techniques. *Hydrol. Process.*, **22**, 4805–4812, doi:10.1002/hyp.7142

Fortin, V., M. Jean, R. Brown, and S. Payette, 2015: Predicting Snow Depth in a Forest–Tundra Landscape using a Conceptual Model Allowing for Snow Redistribution and Constrained by Observations from a Digital Camera. *Atmosphere-Ocean*, **53***,* 200-211

Garvelmann, J., S. Pohl, and M. Weiler, 2013: From observation to the quantification of snow processes with a time-lapse camera network, *Hydrol. Earth Syst. Sci.*, **17**, 1415–1429, doi:10.5194/hess-17-1415-2013

Nievinski, F. G., and K. M. Larson, 2014: Forward and inverse modeling of GPS multipath for snow depth estimation, II: Application and validation, *IEEE Trans. Geosci. Remote Sens.*, **52**, 6564–6573, doi:10.1109/TGRS.2013.2297688.

Parajka, J., P. Haas, R. Kirnbauer, J. Jansa, and G. Blöschl, 2012: Potential of time-lapse photography of snow for hydrological purposes at the small catchment scale, *Hydrol. Process.*, **26**, 3327–3337, doi: 10.1002/hyp.8389