DEVELOPMENT OF A DAILY GRIDDED SNOW WATER EQUIVALENT ANALYSIS

MD. MAKAMUM MAHMOOD

Development of a Daily Gridded Snow Water Equivalent Analysis

by

© Md. Makamum Mahmood

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

April, 2013

St. John's Newfoundland

ABSTRACT

Currently, snow analysis for weather prediction in Canada is conducted using snow depth measurements alone. The current effort is intended to revisit this analysis using both snow depth and density measurements from snow course sites previously unused during weather prediction analysis. The purpose of this reanalysis is to produce a gridded daily Snow Water Equivalent (SWE) hindcast within the transect from the Great Lakes through Quebec and into Labrador.

The final SWE prediction was produced by combining output from existing deterministic snow density models and developed statistical prediction models.

Statistical Models were developed based on Universal Kriging (UK) interpolation technique on measured data and by considering the background fields. These fields include a number of physiographic variables and the Canadian Meteorological Centre (CMC) snow depth analysis product.

Finally, the new product was evaluated from the calculated Root Mean Square Error (RMSE) of the predicted SWE at both validation and cross validation points, simulation vs. observation comparison, and also from a visual consistency check.

The research produced a methodology for SWE prediction and daily gridded SWE product, which is a valuable attempt to improve hydrological prediction from snow melting. The average RMSE of SWE prediction was around 30-35 mm although, the validation of results were challenged by limited quantity of snow course data.

ACKNOWLEDGEMENTS

I want to thank the following people and groups for their supports and contributions over the past two years, without which the completion of this work would not have been possible.

My supervisor Dr. Ken Snelgrove (Associate Professor, Faculty of Engineering and Applied Science, Memorial University) for his support, guidance, and assistance throughout the research.

Faten Jasim (M. Eng. student, Faculty of Engineering and Applied Science, Memorial University) and Jonas Roberts (Ph. D. student, Faculty of Engineering and Applied Science, Memorial University) for their help, suggestion and guidance.

I appreciate that Environment Canada is interested in my research and wants to thank Vincent Fortin (Environment Canada) for his involvement with the research.

Finally, Natural Sciences and Engineering Research Council of Canada (NSERC) and School of Graduate Studies (SGS), Memorial University of Newfoundland (MUN) for the funding to conduct my research.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANCOVA	Analysis of Co-variance
CaPA	Canadian Precipitation Analysis
CLASS	Canadian Land Surface Scheme
CMC	Canadian Meteorological Center
DEM	Digital Elevation Model
DOY	Day Of Year
EC	Environment Canada
ECMWF	European Center for Medium-Range Weather Forecasts
GCM	General Circulation Model
GIS	Global Information Center
GLDAS	Global Land Data Assimilation System
GRADS	Grid Analysis and Display System
IDW	Inverse Distance Weighting
IMS	Image Mapping System
KED	Kriging with External Drift
MSC	Meteorological Service of Canada
MVR	Multivariate Linear Regression
NCDC	National Climate Data Center
NESDIS	National Environmental Satellite, Data, and Information Services
NOAA	National Oceanic and Atmospheric Administration
NOHRSC	National Operational Hydrologic Remote Sensing Center

NRCS	Natural Resources Conservation Services
NSIDC	National Snow and Ice Data Center
NWP	Numerical Weather Prediction
OK	Ordinary Kriging
RK	Regression Kriging
RMSE	Root Mean Square Error
SA	Special Aviation
SAB	Satellite Analysis Branch
SI	Statistical Interpolation
SK	Simple Kriging
SNODAS	Snow Data Assimilation System
SSD	Satellite Services Division
SWE	Snow Water Equivalent
UK	Universal Kriging
UMD	University of Maryland
USDA	U. S. Department of Agricultures
WMO	World Meteorological Organization

Chapter 1 Introduction

1.1 Background

Snow is an important factor in weather forecasting and the study of climate change. Snow cover variability, snow depth and snow water equivalent are three important components related to snow hydrology. In Canada, snow analysis for weather prediction is mainly conducted by snow depth data because of its frequency and ease of measurement.

In Canada, snow density is mainly measured bi-weekly during the snow period. Standard snow tube survey methods are used to measure both snow depth and density at snow course sites. These methods are popular among water resource managers but have not been used fully by Environment Canada (EC) for weather and hydrological forecasting.

The goal of this study is to develop an assimilation of snow-on-the-ground based on snow course observations. The main objective is to develop a daily gridded snow water equivalent product. A SW-NE transect from the Great Lakes through Quebec and into Labrador is proposed for initial use. The main use of this Snow Water Equivalent (SWE) product will be in weather forecasting but it can also be used to initialize water resource models and to evaluate land surface snow models.

There are currently operational snow analysis products that are produced through analysis, remote sensing and hind casting. There are known deficiencies with these products that can be enhanced by the incorporation of snow survey data into an operational data framework. This project would also seek to demonstrate this.

List of Appendices

A dataset based on representative measurements would be useful for snow model evaluation (such as Canadian Land Surface Scheme). This snow course data product would combine a large pool of point data into a spatially consistent data set. A large area stretching from the Great Lakes to Labrador would allow area-based assessment of land surface snow models that would avoid problems associated with representativeness of point comparisons and provide a more appropriate transect based comparison.

In Canada, hydro-power is one of the most important sources of renewable energy. Development of hydroelectric projects can significantly reduce the greenhouse gas emissions (GHG) and also could displace carbon dioxide emissions every year from thermal, coal and fossil fuel power generation. This will reduce the effect of climate change. For the operation and planning of hydroelectric projects, it is important to know the availability of water. In Canada, a significant amount of runoff comes from snow melting. Snow is a vital component in reservoir inflow. It is expected that reservoir inflows can be determined more precisely from this developed SWE product, and forecasting of reservoir inflows is an important part of the planning, operation and maintenance of any hydroelectric development.

In different areas of the world, snowmelt is the main source of surface water supply and ground water recharge. It is also one of the main causes of flooding where a significant portion of precipitation falls as snow. The daily gridded SWE product developed from this study can be used to improve present flood forecasting and also surface or ground water assessment.

1.2 Study Area

For this study, a large transect stretching from Ontario to Labrador has been selected. The study area uses data from within Quebec, the Great Lakes region, and Labrador, primarily. The spatial extent of this study area is bounded by 55^0 W to 97^0 W and 38^0 N to 60^0 N. However, there are large regions within this domain that have little to no snow course measurements included, particularly in the North-West and South-East of the domain.

1.3 Objectives

The main objectives of the study include the following:

1. To construct a snow course database within the transect;

2. To alter the Canadian Meteorological Center (CMC) snow depth data product to reflect snow depth conditions from the dominant vegetation coverage of the region ;

3. To develop statistical SWE prediction models by incorporating snow course measurements with a number of background fields;

4. To produce the final SWE analysis product from the combination of deterministic and statistical models output;

5. To graphically represent the final SWE product by producing spatial maps;

6. To develop an automated SWE prediction tool for Canada.

The first objective of this research project is to construct a snow course database within the transect by collecting data from different sources. This database should be an extension of Ross Brown's snow CD (MSC 2000) and have options for regular updates and visual representations.

In this study, the CMC snow depth analysis product was shown to be an excellent predictor of SWE since it was well correlated with SWE. It was known that the CMC snow depth product was biased since most measurements of snow depth were drawn from open areas such as airports. In this study, the CMC snow depth data product needs to be altered to reflect snow depth conditions from the dominant vegetation of the region. It is expected that this bias will correct the CMC product while maintaining its spatial variability.

The third objective is to develop statistical models by incorporating point SWE measurements. In this study, Universal Kriging (UK) interpolation technique is applied for the statistical model development. Different background fields with good co-relation to SWE are used as predictors in these UK models.

There are some existing deterministic snow density prediction models. These deterministic models need to be evaluated. The next objective of the current research is to evaluate the existing deterministic models in the study area, and combine the output from deterministic and statistical models to produce the final daily gridded SWE.

Graphical representation of the predicted SWE is necessary. The fifth objective of the study is to generate daily gridded SWE maps.

The final objective is to develop a SWE prediction tool, which can automatically incorporate all the information, data, and models to produce the final SWE product for a selected time period. The tool should also be able to produce the final SWE maps.

The main goal of this current study is to produce a daily gridded SWE product within the transect. All the objectives mentioned above need to be fulfilled for reaching the final goal. It is expected from the research that the final SWE product should contain less errors and also be consistent in variation in both time and space. It is also expected that the SWE product should not contradict with the basic nature of snow hydrology.

1.3 Organization of Thesis

The thesis has been organized into 8 chapters as follows. Chapter 2 describes a review of the literature related to snow hydrology. Previous works related to this study are also summarized shortly in this chapter. Chapter 3 describes about different types of data collection and their usage. A description of the constructed data base is also provided. Chapter 4 describes about the spatial analysis, interpolation and data assimilation techniques. Different spatial interpolation techniques are discussed and evaluated in this chapter. Chapter 5 describes the CMC snow depth analysis product and necessary adjustment to reduce the associated bias. In chapter 6, the methodology of spatial and temporal SWE prediction is described. Chapter 6 also describes about the existing deterministic snow density models and developed statistical models. Chapter 7 describes the generated SWE product and error associated with the prediction. In this chapter, a discussion about the research approach and outcome is also presented. Chapter 8 draws the conclusion and summarizes some recommended areas for future improvement.

Chapter 2 Literature Review

2.1 Snow and Hydrology

Snow is a form of precipitation, the flakes of which come in a variety of sizes and shapes. It is a granular porous medium which contains both ice and pore spaces. When the temperature is below 0^0 C these pore spaces contain air but if the temperature is above 0^0 C then the pore spaces can also contain liquid water, resulting in a three phased system (Dingman, 2002). Snow is typically stored on the land surface for hours to months, or even years before melting.

Snow is an important component of the hydrosphere influencing global, regional and local climates, hydrology, and water resources. More specifically, snow plays an important role in the following:

- Water supply
- Agricultural projects
- Hydro-electric power generation
- Soil moisture estimation
- Ecological and environmental needs
- Flood prediction and forecasting
- Climate change
- Numerical weather prediction

Basic terminology for the study of snow includes snow pack, snow depth, snow density and snow water equivalent. Snowpack refers to the accumulated snow on the ground at the time of measurement. Snow depth is the depth of snowpack and snow density is defined as mass per unit volume of snow. For hydrologists, the most important property of a snowpack is the amount of water substance it contains (Dingman, 2002). This amount of water is known as the snow water equivalent (SWE) and denotes as the depth of water resulting from the complete melting of snow. If we denote snow water equivalent as S_{water} , snow depth as S_{depth} , snow density as ρ_{snow} and density of water as ρ_{water} then the relation is

$$S_{water} = \frac{\rho_{snow} S_{depth}}{\rho_{water}}$$

The density of new-fallen snow mainly depends on air temperature (Mellor, 1964) and wind speed (Dingman, 2002). Relative densities of fresh snow with respect to water can range from 0.004 to 0.34 $gmcm^{-3}$ (McKay, 1970) depending on the temperature and wind condition (Dingman, 2002), though relative snow density usually ranges from 0.07 to 0.15 $gmcm^{-3}$ (Garstka, 1964; Dingman, 2002). For the sake of simplicity, an average relative snow density of 0.1 is often assumed (Dingman, 2002). After snow accumulates on a land surface, a metamorphism process continues until it has completely melted. Snow density increases as a function of time to reach the maximum limit.

Snow cover distribution as outlined in Arsenault (2010), can be described in three major spatial scales: macroscale, mesoscale, and microscale. The depth of seasonal snow cover depends mostly upon topography, specifically elevation, slope, aspect, barrier height, vegetation mass, temperature, wind, and predominant weather systems. Snow distribution depends upon the vegetation type, density, and open areas, which are exposed to more accumulated snowfall (Arsenault, 2010). Wind blows snow from one place to another, and thus plays an important part in the transportation of snow. Redistribution of snow water equivalent and loss of water by sublimation occurs during this transportation process (Arsenault, 2010). Additionally while, temperature contributes to the snow melting process, rainfall can contribute more significantly.

Measurements of snow and snowmelt are important for many snow related studies. Taking accurate snow measurements can be quite difficult due to differing scenarios and instrumental limitations. Measurements of snow can be done by in-situ ground measurements or measurements from remote sensing. In the case of in-situ measurements, it is difficult to obtain the spatial variation in snow distribution. There are also problems associated with remote sensing, the primary problem being the coarse resolution. Cloud obstacles are also a challenge with regard to satellite measurements.

The amount of snow on the ground is measured using a number of methods (Dingman, 2002 ; Arsenault, 2010). Snow depth can be measured directly at the time of snowfall, and snow depth and SWE can also be measured from the snowpack. In the case of snowfall measurements, station networks are used to measure snowfall over a certain time interval. Manual and automatic gauges are used to measure both snow depth and water equivalent. A manual gauge can record snow fall in 6-hourly time intervals and an automatic gauge can record snowfall rate in hourly intervals (Arsenault, 2010). Snow rulers, snow boards, and snow pillows are also used for measurement Pictures of various snow gauges are shown in figure 2.1.

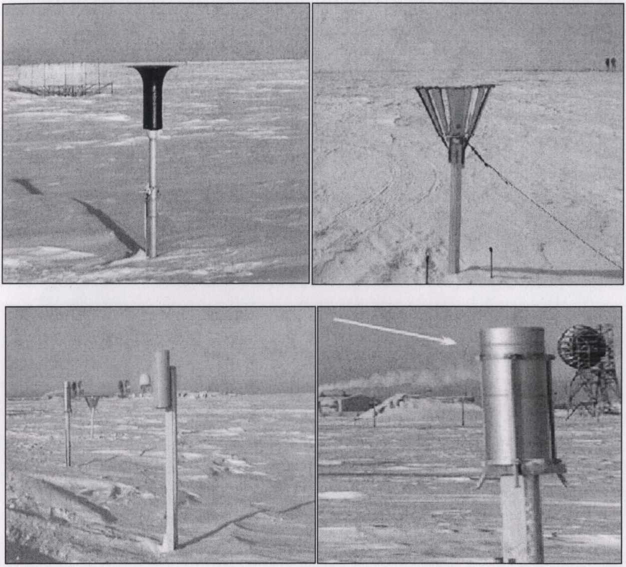


Figure 2.1 Various Gauges for Snow Fall Measurements (Arsenault, 2010)

The United States has developed snow telemetry, known as SNOTEL, to record and transmit snow depth, water equivalent, precipitation, temperature and soil moisture from more than 100 sites in the western US. It is an example of a gauge network. Figure 2.2 presents a SNOTEL site at Smiley Mountain in Big Lost River basin. It contains a variety sensors for measuring different parameters, as well as snow pillows for SWE measurements.

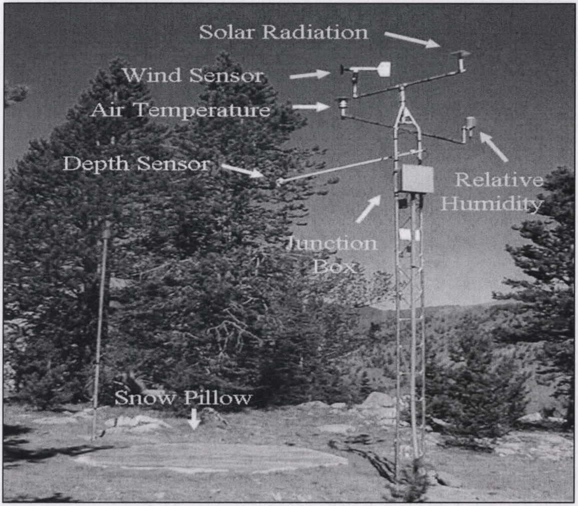


Figure 2.2 SNOTEL Site at Smiley Mountain (Natural Resources Conservation Services, US)

There are a number of ways to conduct snowpack measurements. A snow survey is the a common method. They are conducted periodically at fixed locations, known as snow course sites. Manual measurements of snow depth and SWE are conducted by trained professionals and typically using a standard snow tube method. Figure 2.3 illustrates this method. Snow pillows and acoustic gauges invented by Chow (1992), can also be used to perform snowpack measurements (Dingman, 2002).

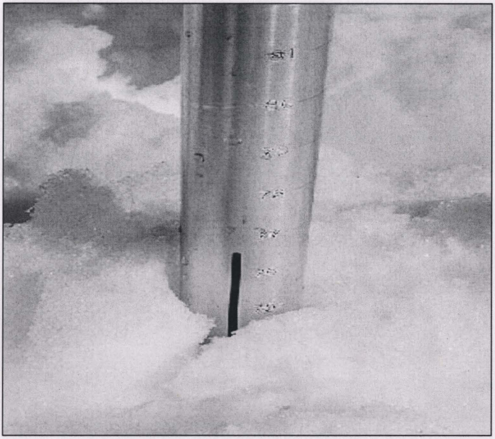


Figure 2.3 Snow Tube Used at a Snow Course Site

According to Dingman (2002), snow can also be measured using remote sensing technologies. For instance, microwave radiation can be used to measure areal extent, SWE and other snowpack properties. Snowpack emitted microwave radiation depends on its temperature, grain size and soil condition, which can be used to estimate SWE (Foster et al., 1987). Presently, there is an Advanced Microwave Scanning Radiometer —Earth Observing System (AMSR--E) with snow depth and SWE products available at a 25 km resolution.

Satellite imagery using visible and infrared light can also provide information about snow cover extent. It should be kept in mind that careful interpretation is required to distinguish snow from clouds, and to identify snow in forested areas and highly reflective land surfaces (Dingman, 2002).

Table 2.1, from Dingman (2002), summarizes snow measurements for different parameters. Here (G) means ground based measurement, (A) means aircraft based measurements, and (S) means satellite based measurements.

Table 2.1 Summary of Methods for Snow Measurements (Dingman, 2002)

Parameter	Depth	SWE	Areal Extent
Precipitation		Standard Storage Gage (G)	Gage Network (G)
		Universal Gage (G)	Radar (G)
Snowfall	Ruler, board (G)	Melt Snow on board (G)	Observation networks (G)
		Use estimated density (G)	Radar (G)
		Universal gage (G)	Visible / Infrared (S)
		Snow pillow (G)	
Snowpack	Snow Stake (G, A)	Universal gage (G)	Snow surveys (G)
	Snow tube (G)	Snow tube (G)	Visible / Infrared (S)
	Acoustic gage (G)	Snow pillow (G)	Microwave/radar (A,S)
		Artificial radio isotope gage (G,A)	
		Microwave/radar (A,S)	
		Natural gamma radiation (G, A)	
Snowmelt		Snow pillow (G)	Snow-pillow network(G)
		Lysimeter (G)	
		Universal gage (G)	

For the purpose of this research, it is necessary to know about snow melt as change of SWE is closely related to melted snow leaving the snowpack. There are four phases of snow melt (Dingman, 2002). The first phase, when an increase of water equivalent occurs within the snow pack, is the accumulation period. The next phase is the warming phase where snowpack temperature increases fairly steadily until the snowpack is isothermal at 0^0 C. The melting phase, where melting water is retained in the snowpack, occurs next. The final phase is the output phase where water drains out from the snowpack.

2.2 Snow Analysis in Canada

Snow is an important component of Canadian hydrology, and is represented as such in published literature. Ross Brown is a prominent person in the Canadian snow field. He has conducted a number of studies in snow cover variability (Brown, 2000; Brown and Mote, 2009; and Brown, 2010). Many climate change and hydrological patterns can be obtained from the analysis of snow cover and other parameters related to snow. That being said, snow depth and density, not snow cover, are the main focus in the current study.

Canadian Snow Depth Analysis

Real time snow information of snow extent, snow depth, snowpack density, and snow water equivalent are necessary for accurate numerical weather prediction. The Canadian Meteorological Center (CMC) has developed a global daily gridded snow depth analysis product. Brasnett (1999) describes details about this snow depth analysis where, two types of information were used: snow depth observation obtained from the SYNOP observing network and an estimated background field from the snowpack model. According to Brasnett (1999), the measured data were screened in a three step process: 1) false reports of snow 2) systematically understated snow depth and 3) stations that violate temporal continuity. A background field was generated from a complete snow pack model by considering all types of precipitation, sublimation/condensation, snowmelt and snowpack metamorphosis (Brasnett, 1999). The final snow depth was produced by combining background fields and observational data using a statistical interpolation technique (Daley, 1991).

The produced snow depth was verified with independent data points, and has been shown to provide better prediction of snow than simple climatology. From the analysis, it was found that the global NWP model suffers from a precipitation deficit in some regions. For instance, snow depth was underestimated by an average of 3.0 cm for the Northern Hemisphere (Brasnett, 1999). According to the author, bias in the analysis could occur due to two main reasons: the underestimation of precipitation by NWP models, and because most measurement stations are located in urban areas or valleys. In other words, it is a problem associated with the representativeness of the data.

The analysis scheme presented in Brasnett (1999) was also used to produce snow density though no observed snow density data was used. For verification, density was estimated based on the age of the snowpack, though, snow density estimation was likely affected by the NWP model's underestimated precipitation.

The CMC snow depth product can be used for a number of purposes other than weather prediction, such as detecting the effects of climate change and validating algorithms for satellite data processing (Brasnett, 1999). The CMC product is of great value to this thesis as it can be used as the main predictor of SWE. Figure 2.4 presents a sample CMC snow depth analysis map.

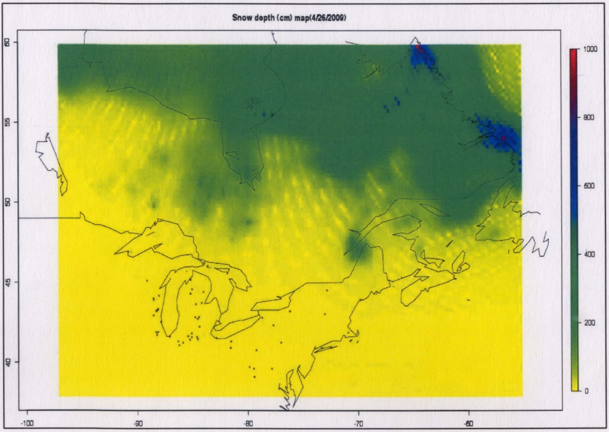


Figure 2.4 Sample CMC Snow Depth Analysis Map within the Study Area

<u>Gridded Monthly Snow Depth and Snow Water Equivalent</u>

Gridded estimation of snow water equivalent is necessary for the evaluation of snow cover in General Circulation Models (GCMs). For this purpose, Brown et al. (2003) developed a gridded monthly snow depth and SWE product for North America. The developed snow depth analysis in CMC (Brasnett, 1999) was applied to generate the monthly mean snow depth and corresponding snow water equivalent. The developed gridded mean snow depth and SWE product can be used for snow cover analysis in GCM models, as well as satellite data evaluation and validation.

The effects of climate change can be detected by using snow cover analysis from GCM simulated snow cover. It should be noted that evaluation of snow cover simulation by GCM models are hampered over the Northern Hemisphere (NH) due to lack of SWE data (Brown et al., 2003). The aim of Brown et al. (2003) was to develop monthly snow depth and SWE product for North America to evaluate GCM snow cover simulation during the Atmospheric Model Intercomparision Project II by using the snow depth analysis scheme developed by Brasnett (1999).

According to Brown et al. (2003), two types of information were required: snow depth observation and a guess background field. Snow depth data in Canada was collected from Canadian synoptic stations, and snow depth data reported weekly or biweekly at Canadian snow course sites was also used in the analysis. US data from National Climate Data Center (NCDC) was also incorporated. Brown et al. (2003) attempted to generate a more accurate background field from the snowpack than that of Brasnett (1999) as their simple snow aging scheme did not provide adequate results for the mean variation of snow densities over North America (Brown et al., 2000). In Brown et al. (2003), the background snow depth field was generated by a snowpack model using European Center for Medium-Range Weather Forecasts (ECMWF) precipitation and analyzed 2-m air temperature as input (Brown et al., 2003). Additional terms have also been included in this model including mixed precipitation type, rain melt, detailed treatment of snow aging, melt factor, and canopy sublimation loss. The observational data and background field were combined by using the statistical interpolation system (Daley, 1991).

SWE was estimated in this study from the analyzed snow depth and calculated snow density from the snowpack model (Brown et al., 2003). It should be noted that at no point are SWE observations used and SWE had been completely estimated from the snowpack model considering ECMWF model precipitation and temperature as the model input. The estimated monthly SWE was validated with respect to snow course sites, though the validation snow course dataset was not totally independent as it had been already used for the estimation of monthly average snow depth. After evaluation, it was found that the density was overestimated in the boreal forest zone and underestimated in other places. To improve the situation the author suggests using more detailed snowpack model.

Brown (2000) also compared his research with findings from some previous studies and found that snow depth data provided a good indication of inter annual variability of SWE using fixed seasonal snow density. A previous study (Brown, 2000) suggested that detailed simulation of snow density might not be necessary to predict SWE from snow depth. However, Brown et al. (2003) has mentioned that by using climatologically average snow density in estimating SWE, one is not able to take into account events such as winter thaws and rain. It was also observed that prediction of snow density was improved in mountainous region, but it was also observed that the snowpack model did not perform as well on the Labrador coast (Brown et al., 2003). The information extracted from Brown et al. (2003) can be used with the current SWE analysis, while the dataset developed in Brown et al. (2003) can also be used for the validation of SWE product.

2.3 Canadian Precipitation Analysis

Canadian Precipitation Analysis (CaPA) converts real time point measurements of precipitation into a gridded data set. The methodology and preliminary results of CaPA were discussed in Mahfouf et al. (2007). Though the CaPA project was created primarily for rainfall precipitation, the information extracted from this project can be useful in the current study. In CaPA, a statistical interpolation technique was used to convert point rainfall into gridded data set.

The goal of CaPA is to produce 6-h rainfall accumulations at a resolution of 15 km over North America in real time (Mahfouf et al., 2007), which is important for NWP models, hydrological forecasting, flood forecasting, soil moisture analysis, and water management. Preliminary steps about the project were described by considering the Quebec region in August 2003. The goal of this project was to produce a framework where real time precipitation analysis could be easily produced and evaluated (Mahfouf et al., 2007).

The spatial interpolation technique used in CaPA as well as information related to the skewness of the data, are both relevant to this thesis.

The Statistical Interpolation (SI) technique has the option to use an initial guess background field while techniques like kriging (Finkelstein, 1984) and the successive correction method (Bussieres and Hogg, 1989) do not have this option. This guess field can improve prediction capacity. The SI has been used in Brasnett (1999) and Brown et al. (2003) and has been widely used as a data assimilation technique to combine observational data with model output. In CaPA, a short term forecast of 6-h accumulated precipitation from the regional GEM model is used as a background field to combine with the observational data to produce the final product.

Dealing with zero values is a challenge in the current research, as it is in CaPA. According to Mahfouf et al. (2007), the analysis in CaPA was performed on a transformed variable x= ln (p+y), where x is the transformed 6-h accumulated precipitation in mm, p is 6-hr accumulated precipitation in mm and y is a constant at 1mm. The value of the constant was chosen to be 1 mm, so that after transformation the zero precipitation values became zero again in the log scale (Mahfouf et al., 2007). The main reason for the transformation was to make the data normal and to remove the skewness, which is required for various statistical analysis techniques.

A 6-hr accumulated precipitation map produced by the CaPA objective analysis is shown below as Figure 2.5.

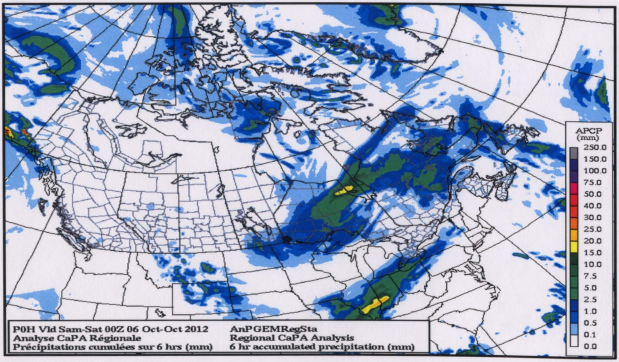


Figure 2.5 Precipitation Map Generated from Regional CaPA Analysis (Environment Canada)

2.4 SNODAS

American researchers have developed the Snow Data Assimilation System named SNODAS, which contains an assimilation of different variables related to snow. SNODAS was developed by the National Operational Hydrologic Remote Sensing Center (NOHRSC) to support hydrological modelling and weather forecasting, by integrating snow data from satellites, airborne platforms, and ground stations with the model estimates of snow cover (Carroll et al., 2001). SNODAS includes procedures to ingest and downscale output from NWP models, and procedures to assimilate satellite-derived, airborne and ground-based observations of snow covered extent and snow water equivalent (Barrett, 2003). An assimilation of snow depth and snow water equivalent has also been generated from the SNODAS project.

According to Barrett (2003), SNODAS has three components: a data ingest, quality control, and a downscaling component for meteorological information from numerical weather prediction models; a snow mass and energy balance model; and data assimilation routines to update snow model estimates of snow pack variables with observed snow cover, snow depth, and snow water equivalent data.

Differences between observed and estimated values were computed first, then the model generated fields were updated by simple nudging or a relaxation data assimilation technique (Barrett, 2003). Figure 2.6 presents a spatial SWE map generated by SNODAS on 29 February,2004.

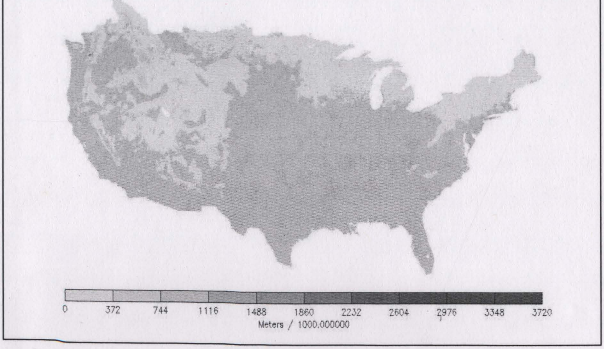


Figure 2.6 Sample SWE map Generated from SNODAS (NSIDC)

From SNODAS, it is important to note that a framework of snow assimilation systems with snow cover, snow depth, and snow water equivalent as variables can be useful for hydrological analysis and weather prediction. In Canada, CMC has developed a snow depth assimilation system, though a SWE assimilation system would be useful. This thesis aims to fill this gap.

2.5 Snow Cover Classification System

A number of classification systems have been developed to classify snowflakes (Magono and Lee, 1966) and snow on the ground (Somerfield, 1969 ; UNESCO, 1970). These classification systems describe snowflakes at scales ranging from millimeters to centimeters (Sturm and Holmgren, 1995). In Sturm and Holmgren (1995), a snow classification system for seasonal snow cover consisting of six classes (tundra, taiga, alpine, maritime, prairie, and ephemeral) was proposed. This classification system is based on collected field observations and has local to global applications. A simple snow density model has also been developed based on this snow cover classification system, and is used in the current research. Simple snow density model used here due to computational simplicity and it was also found from previous studies that complex density models are data intensive (Brown et al., 2003) and difficult to setup for large area. According to Sturm and Holmgren (1995), each class in the new classification system was defined by a unique ensemble of textural and stratigraphic characteristics including information about the snow layers, thickness, density, as well as morphological and grain characteristics within each layer. The classes have been also derived using variables such as winter wind, precipitation, and air temperature (Sturm and Holmgren, 1995).

Sturm and Holmgren's (1995) classification system was based on physical characteristics and not vegetation type or geographical locations. Climate, vegetation, and snow were interrelated but this three way relationship was quite complex. Due to this complexity, snow class was first defined by its physical characteristics, and only subsequently related with the climate, vegetation, and geographic location. Figure 2.7 presents a global snow classification system map derived from Sturm and Holmgren (1995).

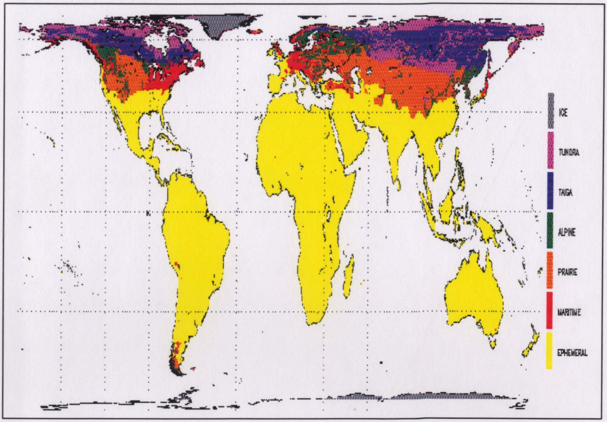


Figure 2.7 Global Map of Snow Classification System (Sturm and Holmgren, 1995)

The accuracy of this snow classification is strongly dependent on the atmospheric and vegetation dataset used. Error in this classification system also occurs due to the lack of station density at some points (Sturm and Holmgren, 1995).

The classification system developed by Sturm (1995) was verified with observed data points. Despite some inconsistency, maps produced from the classification system were accurate enough to use in climate applications. A comparison of the maps produced with observed data indicated accurate classification in roughly 62 to 90% of the area studied (Sturm and Holmgren, 1995).

Two deterministic models are used in the current thesis: a snow climate class model (Sturm et al., 2010) and a snow aging model. The snow climate class model is based on the snow cover classification system , discussed above. Certain parameters in the snow density model such as maximum snow density were also based on the snow cover classification system (Sturm et al., 2010). The values of maximum snow density within each climate class can be used as threshold values to predict density in the SWE analysis.

2.6 SWE Prediction Using Climate Class

According to Sturm et al. (2010), snow water equivalent is the most significant term to understanding global snow water trends. It is already known that measurements of SWE are more difficult and complex than snow depth and snow cover. As satellite measurements of SWE are not up to date, it would be useful if SWE could be predicted from other variables such as snow depth (Sturm et al., 2010). No comprehensive data is available regarding snow depth and SWE measurements worldwide, but it is known that snow depth is measured more frequently than the SWE.

Sturm et al. (2010), attempted to find a suitable way to convert snow depth to snow water equivalent using snow climate classes (Sturm and Holmgren,1995). Using a large training set, bulk density was predicted as a function of snow depth, snow climate class, and day of year, and used to convert snow depth to SWE. Non-linear analysis of covariance model (ANCOVA) was used to develop an equation to predict bulk density based on the predictors mentioned above. Observational data were collected from research surveys in Canada, Switzerland and Alaska with the anticipation that the collected data covered the spatial variability related to snow density (Sturm et al., 2010). The accuracy of the model was evaluated against a large independent data set collected from the MSC (2000).

Sturm et al. (2010) found that, it was more appropriate to model bulk density rather than SWE and subsequently convert this bulk density to SWE using snow depth. Bulk density is a complex function of snow depth, snowpack temperature, aging, and initial density of snow layer. The bulk density was modeled using Bayesian statistical methods to develop a nonlinear ANCOVA model (Sturm et al., 2010). Bayesian methods work well in modelling complex systems (Gelman et al., 2004). Snow depth, snow climate class and day of year were used as model inputs, while predictors of snow density, like temperature and initial density, were captured using the snow climate class (Sturm et al., 2010). Various functional forms were used to choose the best model based on deviance information criterion (DIC; Spiegelhalter et al., 2002).

The model was validated based on probability distribution functions and error statistics. The error in predicted SWE was found to be in the range of -70 mm to 90 mm of the observed (Sturm et al., 2010). The bulk density error was nearly constant, but the probable error in SWE increased with depth while the relative error decreased with increasing SWE. The model results were also compared to collected data from three midlatitude basins in the US and were found to be satisfactory (Sturm et al., 2010).

The strength of the Sturm et al. (2010) model is in its simplicity and ease of use. As mentioned earlier, the model can be applied over an area or region by inputting the climate class of the area, snow depth and day of year, and as such will be used as a deterministic model to predict SWE from the gridded CMC snow depth data in this thesis.

2.7 Spatial Interpolation

One of the objective of the current study is to convert point SWE measurements into a spatial dataset. Spatial interpolation plays an important role in the conversion process, and in this section the advantages and disadvantages of various spatial interpolation methods will be discussed.

Spatial interpolation is a procedure which uses points with known values to estimate values at unknown points. In other words, spatial interpolation is a method used to estimate the value of properties at unsampled points using sample points within the area covered by observations (Goodchild et al., 1990). In most cases, spatial interpolation is used to estimate grid values, convert point data to surface data, and prepare finer grid estimation from the coarser grid. Spatial interpolation is important in the fields of geosciences and Global Information Systems (GIS) , as well as water resources, hydrology, and water management.

Spatial interpolation procedures can be classified according to whether they are useful on a global and local scale. In global interpolation a single function is used to interpolate the whole region (Goodchild et al., 1990), and a single change in input can affect the whole region. The advantage of using global interpolation is that it produces a smooth surface. Local interpolation applies to a small portion of the total set of points. Abrupt changes can be found when one combines local interpolations for mapping the full area of interest (Goodchild et al., 1990). The main difference between global and local methods lie in use of control points, defined as points with known values. In global interpolation all the control points are used to derive a model, but in local interpolation only a sample of control points are used. Additionally, there are exact and approximate interpolation methods (Goodchild et al., 1990). Exact interpolators honor the data points upon which the interpolation is based, while approximate interpolation is used when there is some uncertainty with the given surface values. Spatial interpolation can also be classified as stochastic or deterministic, where the concept of randomization is used in stochastic interpolation while deterministic methods use no concept of probability (Goodchild et al., 1990).

The accuracy of any spatial interpolation method mainly depends on the density of known control points. There are a variety of spatial interpolation techniques. The basic assumptions in each are very similar and the accuracy is always dependent upon the number of control points and their distributions. The primary assumption is that the prediction is influenced by nearby control points and as one goes further afield the influence diminishes. In the final section of this chapter, various spatial interpolation processes are described.

Inverse Distance Weighted Interpolation

Inverse Distance Weighted Interpolation is one of the most common techniques used in spatial interpolation. Here the unknown value of a point is influenced by nearby control points than those further away, and the degree of influence is expressed by the inverse of the distance between points.

Nearest Neighbor Interpolation

Nearest neighbor interpolation is a simple interpolation method which can also be used in spatial prediction. In this process all values at unknown points are assumed to be equal to the nearest known points. Thiessen polygons are typically used with this method.

Thin-Plate Splines

This method is similar to spline for line generalization, but applies to surfaces. Minimum curvature can be achieved by using thin-plate spline, and it is useful for producing smooth surfaces. It is recommended for spatial interpolation of elevations, water tables and climate data.

Kriging

Kriging is a geo-statistical method that assumes the spatial variation of an attribute is neither totally random nor deterministic. Kriging was developed by Georges Matheron, as the "theory of regionalized variables", and D.G. Krige as an optimal method of interpolation for use in the mining industry. The basis of this technique is that the rate of variance between points changes spatially (Goodchild et al., 1990). This can be defined as a variogram calculated from semi-variance. Different types of variograms can be fitted for use in kriging; with Spherical, Exponential and Gaussian being the most common.

Kriging can be simple kriging, ordinary kriging, universal kriging, and co-kriging. Simple kriging assumes that the surface has a constant mean with no trend (Goodchild et al., 1990). Ordinary kriging focuses on spatially related components but assumes that there is no trend, while universal kriging incorporates a trend. In co-kriging, one or more secondary variables, that are correlated with the response are used.

Trend Surface Analysis

In this method the surface is approximated by a polynomial equation, and trend surface models are used to estimate values at unknown points by fitting a trend with the sample points. The goodness of fit for the trend surface is measured by R^2.

Regression

Multivariate regression can also be used for spatial prediction and interpolation. These models relate dependent variables to a number of independent spatial predictors. The accuracy of a regression model depends upon its prediction capacity as denoted by R^2.

Summary

Among the above methods, universal kriging, multivariate regression and co-kriging are the most promising, as there are option to use background fields or other variables in the interpolation. While there is ample software which can be used for the interpolation, in the current study R statistical analysis software is used.

Spatial interpolation techniques are used in various fields of water resources and environmental study. Additionally, they can be used in precipitation, temperature, elevation, snow depth and snow water equivalent prediction. Different spatial interpolation techniques are evaluated in Chapter 4 in order to determine the best technique to carry out the final analysis.

Chapter 3 Data Collection

3.1 Introduction

Different types of data are necessary for conducting the SWE analysis. These data include Canadian snow course data, USA snow data, National Snow and Ice Data Center (NSIDC) snow extent data, Canadian Meteorological Center (CMC) snow depth analysis data, Physiographical data, elevation data, and snow cover classification data. All these data were collected from a number of sources. The quality of the data was unknown, but it seems to have some error on it which is denoted by the Nugget effect on variogram fitting. All these data were organized in a fixed format to use in the analysis. In most cases the gridded data are varied in spatial scales and inverse distance interpolation technique was used in this cases to convert all the gridded data into CMC grid resolution used in snow depth analysis. Our final SWE product is also produced in the same grid as CMC snow depth analysis. In the following part of this chapter, these data types will be discussed in details.

3.2 Construction of Snow Course Database

A Canadian snow course database was constructed within the transect as part of the study. Construction of this snow course database builds on the work of the "Snow CD" product developed by Environment Canada (MSC, 2000). In addition to these data, the time base of the measured data has been extended from 2003-2004 to 2010-2011 with data from Newfoundland Hydro, the Ontario Power Generation Authority, the Ontario Ministry and the Quebec Ministry. Figure 3.1 shows the location of the snow course throughout the period and some stations have stopped functioning in the recent years. In fact, most Canadian stations record snow course only twice per month during the snow accumulation period, and many of the stations have short records.

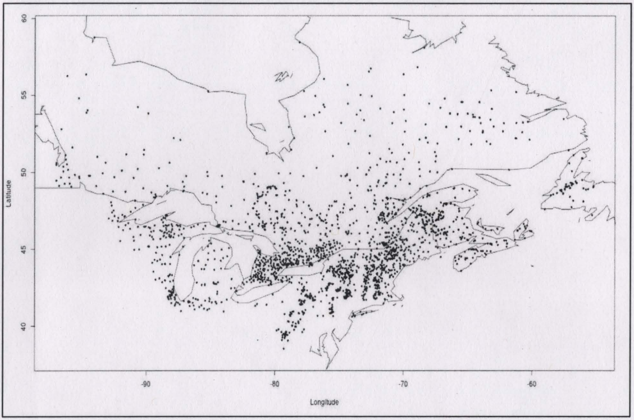


Figure 3.1 The Study Domain and Snow Course Station's Location

A snow course is a permanent site, where manual measurements of snow depth and snow water equivalent are conducted (Arsenault, 2010). Measurements are usually taken during the winter and spring seasons. Snow course sites are usually selected at wind protected locations. They are mainly located in valleys and mountainous regions. At snow course sites, measurements are mainly conducted using the standard snow tube survey method. Snow pillow measurements are also used at some locations. Figure 3.2 details a sample snow course survey conducted by the U. S. Department of Agriculture's (USDA) Natural Resources Conservation Service (NRCS).

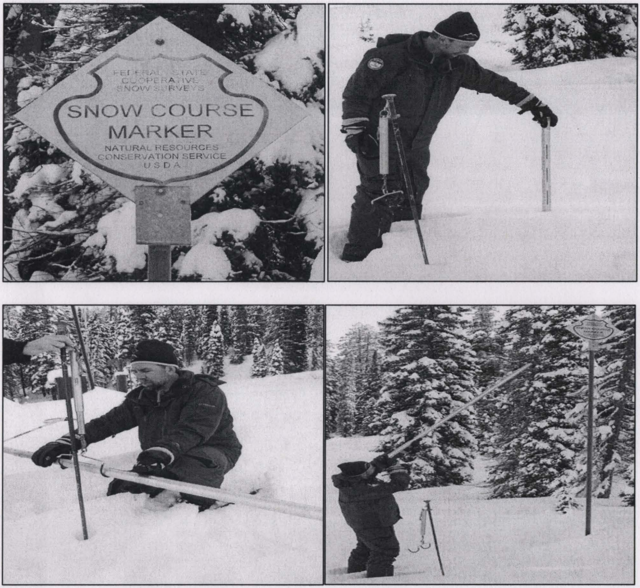


Figure 3.2 Snow Course Survey in USA (NRCS-USDA)

One of the main objectives of the current study is to construct a snow course database within the transect. The database was constructed in a fixed format, which makes it easy to read and update on a regular basis. The database was compiled by FORTRAN code to read and update on a regular basis. The database was compiled by FORTRAN code and visually represented by Grid Analysis and Display System (GRADS) software packages. Two database files were created. One was for the station database which contained name, id, location, Lat, Lon, and elevation of the snow course stations. The other file was for the snow course database and which contained snow depth and snow water equivalent data with corresponding lat, lon and date. Both these files are in text format. FORTRAN codes have been written to update the database with new available data. Real time update of the constructed database is also possible by using real time snow course data. The database was later converted in binary format for visual representation using GRADS software package. Details about the data collection are provided in Appendix-A.

GRADS scripts have been written for the visual representation of the compiled data. Three types of visual representation were generated in the current study. Figure 3.3 presents a sample time series variation of the SWE data for the snow course station at Philip Lake, BC, while Figure 3.4 represents a sample spatial variation of SWE data over Canada on December 01, 1986. In GRADS, spatial interpolation of data is also possible. Figure 3.5 represents Creesman spatial interpolation of the point data on December 18, 1982. These data are plotted as a visual guide. All dates could be represented in a simpler manner.

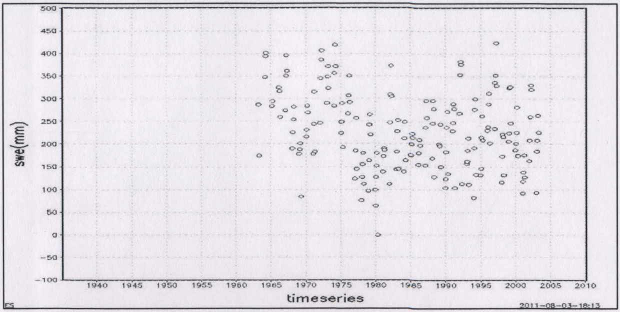


Figure 3.3 Time Series Variation of SWE Data for Philip Lake, BC

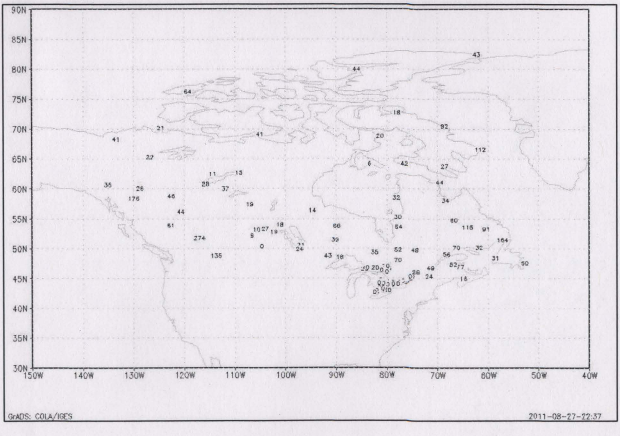


Figure 3.4 Spatial Distribution of Measured SWE Data on 01 December, 1986

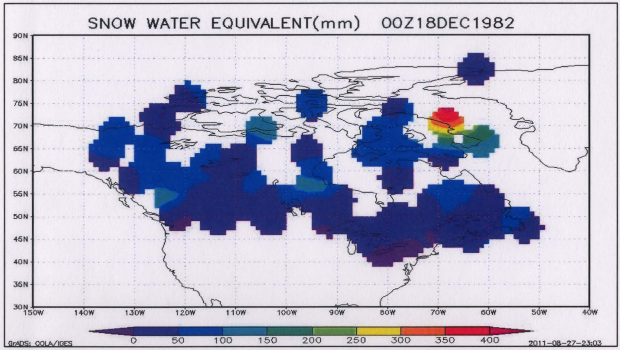


Figure 3.5 Creesman Interpolation of the Point Data on 18 December, 1982

3.2 USA Snow Data

USA snow data was also incorporated into the present study. USA snow course data can play an important role to predict SWE around the Great Lakes area. Only data from the USA stations close to the Canadian border and from the North East region were incorporated in the current SWE analysis. The intent here was to develop a more complete database of snow measurements within the region to enhance prediction and assess the value of these additional data.

Northeastern United States data were drawn from online sources as required and have not been directly incorporated into the database. The National Operational Hydrologic Remote Sensing Center (NOHRSC) contains all the USA snow data. Daily snowfall, snow depth and snow water equivalent data are available on their website (http://www.nohrsc.noaa.gov/nsa/).

Point SWE measurements are the primary interest for the research. SWE and only the corresponding snow depth data for the analysis period were collected from the NOHRSC database. Data only from the North-East region of the USA, the Northern Great Lakes, the Southern Great Lakes and the Allegheny Front were collected in the current study. It should be kept in mind that all incorporated USA data are not collected from the standard snow tube survey method. There are some data which are measured by the snow pillow measurement technique.

3.3 Snow Extent Data

Snow cover and snow extent data could have been helpful in the current SWE analysis. The snow extent data could have been used to validate the predicted SWE. The Ice Mapping System (IMS) daily Northern Hemisphere snow and ice analysis data was used in the analysis as the source of snow cover or snow extent data. This data is available at a resolution of 4 km and 24 km and can be downloaded from National Snow and Ice Data Center (NSIDC) (NOAA, 2004). 4 km resolution data is used in our analysis.

The National Environmental Satellite, Data, and Information Service (NESDIS), part of the National Oceanic and Atmospheric Administration (NOAA), is responsible for the monitoring of snow and ice cover. By inspecting environmental satellite imagery, analysts from the Satellite Analysis Branch (SAB), Satellite Services Division (SSD) created a Northern Hemisphere snow and ice map from November 1966.

The speed, accuracy, and resolution of the dataset was improved by applying IMS with which maps could be also produced by adding additional data. Passive microwave data were used as well the IMS dataset to improve snow detection under cloudy or nighttime conditions. However, automated snow detection via passive microwave is subject to error in certain situations. Therefore, for the greatest possible accuracy, creation of a manual analysis product from a variety of sources continues (Ramsay, 1998).

The snow extent data are in ASCII format and available from the NSIDC website (http://nsidc.org/data/g02156.html). Snow extent data for the simulation period was downloaded and used in the validation of the SWE analysis product. The data can also be visually presented by using GRADS software packages. Figure 3.6 represents a sample snow and ice image of the Northern Hemisphere mapped by NOAA.

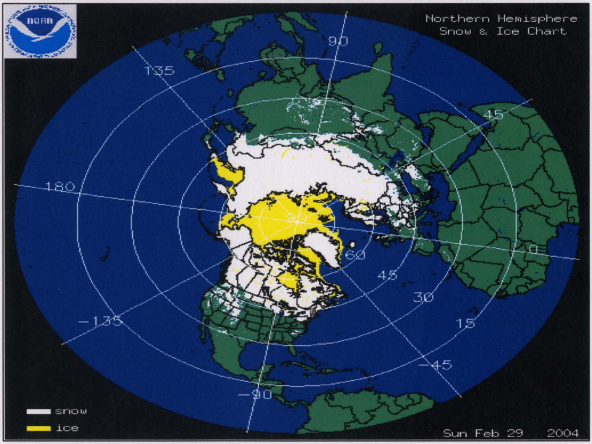


Figure 3.6 24 km Resolution Northern Hemisphere Snow and Ice Chart at Feb 29, 2004 (Satellite Services Division, USA)

3.4 Canadian Meteorological Center Snow Depth Analysis

The Canadian Meteorological Center (CMC) has developed a gridded snow depth analysis product (Brasnett, 1999). This product has a global spatial resolution and daily temporal resolution. It is a gridded product with a resolution of 24 km. Snow depth data collected from surface synoptic observations (synops), meteorological aviation reports (metars), and special aviation reports (SAs) were acquired from the World Meteorological Organization (WMO) to use in the CMC analysis. This CMC data set includes daily observations from 1998 through 2011 and will be updated annually. This CMC snow depth analysis data is of great interest in the current study as it is the main predictor of SWE. The advantage of using CMC snow depth data is that it is a gridded and daily data set.

The CMC dataset can be downloaded from the NSIDC website (Brown and Brasnett, 2010). The data set are in gridded format and provided in tab-delimited ASCII format. There are separate files for location and snow depth data. CMC snow depth data was downloaded for the simulation period and later Fortran codes were used to input these data in the SWE prediction tool.

A monthly gridded snow depth and SWE data set has also been developed for weather prediction (Brown, 2003). This dataset is also available in the NSIDC (Brown and Brasnett, 2010) website. This dataset can be used in future analysis of a predicted SWE product. Validation of the newly developed product can also be done using this monthly average dataset.

3.5 Physiographical Data

Physiographic variables can be used as a very good predictor of SWE. Previous studies show that different physiographical variables such as elevation, slope, aspect, easting, northing, vegetation, forest density, and distance from ocean have good correlations with SWE (Fassnacht et al., 2003). To make a better prediction of SWE, a number of physiographical variables were used in this research. These physiographical variables included elevation, slope, aspect, local slope, eastness, northness, barrier height, vegetation type, vegetation mass, distance from ocean, and distance from water bodies.

Elevation is one of the most important physiographic variables evaluated in the current research. Elevations have been extracted from the Global Land Data Assimilation System (GLDAS) (Rodell et al. 2004). The goal of GLDAS is to combine satellite and ground-based observational data products, using advanced land surface modeling and data assimilation techniques to generate optimal fields of land surface states and fluxes (Rodell, 2004a). Elevation data are available in two resolution formats. 1/4 degree resolution was used in the current study. Though, there are more accurate elevation data available, this data set has been chosen as it can be extracted easily by using Fortran codes and GRADS software packages and the SWE prediction grid resolution in the current research is around 24 km.

Barrier height is the elevation difference between the maximum barrier in direction of the ocean and the grid, and derived from the GLDAS elevation data. Land surface slope are derived from the GTOPO30 DEM (Bliss et al., 1996). GTOPO30 is a global digital elevation model with a resolution of 30 arc second. Figure 3.7 presents the global slopes derived from GTOPO30 dataset. Aspect, local slope, eastness, and northness data used in the current analysis, were calculated from the slope and elevation dataset.

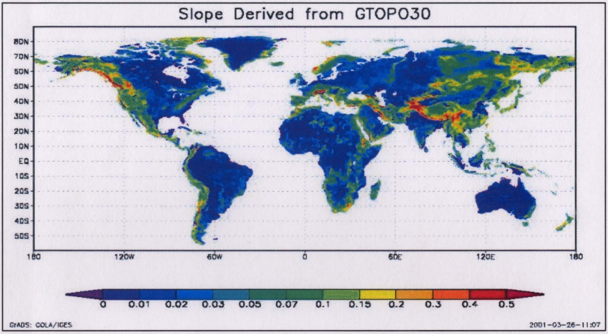


Figure 3.7 Slope Derived from GTOPO30 (NASA, USA)

Vegetation density is another important physiographical variable. Vegetation mass has been calculated by combining the global vegetation cover data from the Global Land Data Assimilation System (Rodell et al., 2004) and unit mass look-up table used in Canadian Land Surface Scheme (CLASS) (Verseghy et al., 1993). Data of Vegetation types are available in GLDAS in 1/4 degree and 1 degree format. 1/4 degree vegetation type data was used in the current study. 14 vegetation types classified by the University of Maryland (UMD) are used in GLDAS. Figure 3.8 details the global distribution of different vegetation types. The frequency of each vegetation type at each 1/4 degree grid is also available. By using the frequency of vegetation types within the grids and vegetation mass table in CLASS, the final relative vegetation mass at each 1/4 degree grid was produced within the transect.

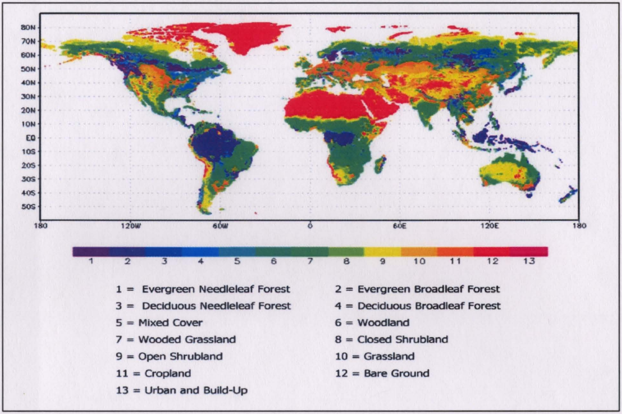


Figure 3.8 Global Map of Vegetation Type (NASA, USA)

Distance from ocean has been calculated from the Global Self-consistent, Hierarchical, High-resolution Shoreline Database (Wessel et al., 1996) and distance from water bodies was calculated from the land-water mask.

Chapter 4 Spatial Interpolation

4.1 Introduction

The goal of the current research is to develop a spatial daily Snow Water Equivalent (SWE) dataset. One of the challenges is to convert the point data into a spatially consistent gridded dataset. To do this, spatial interpolation techniques are used, which could also be used for data assimilation by incorporating background fields.

Previously, statistical methods have been used to interpolate SWE over large areas, but within areas that experience limited variation in topography (Carroll et al., 1999). Kriging (Carroll, 1995), elevation-detrended kriging (Carroll and Cressie, 1996) and binary regression tree analysis (Elder et al., 1998) have also been used for spatial interpolation of SWE (Fassnacht et al., 2003). To predict SWE within the Colorado River basin, a number of spatial interpolation techniques were evaluated (Fassnacht et al., 2003) including inverse distance averaging, optimal distance averaging, hypsometric (HYP) and multivariate physiographic regression (MVR) techniques.

Based on these previous studies, five spatial interpolation methods including Inverse distance weighting (IDW), Multivariate Linear Regression, Universal Kriging (UK), Ordinary Kriging (OK), Regression+IDW were considered for SWE map production in this study. Details about these methods are provided in the following section. These various methods were compared over 14 test days spanning 2008-2010.

4.2 Interpolation Techniques

Five spatial interpolation techniques were evaluated and are presented here. Details about each techniques are described within.

4.2.1 Inverse Distance Weighted

Inverse distance weighted (IDW) is commonly used in spatial analysis and follows the Tobler (1970) first law of geography: "Everything is related to everything else, but near things are more related than distant things". Known points or control points are used to predict at unknown points, and the closer points have more influence than those further away. This influence is denoted as weight, and hence "inverse distance weighted" method.

IDW assumes that each measured point has a local influence that decreases with distance and weighs the points closer to the prediction location greater than those farther away (Brusilovskiy, 2009). In other words, weights of each measured point are proportional to the inverse distance. The decreasing rate of weight depends upon the power. Spatial analysis software allows one to select the optimum power by minimizing the Root Mean Square Prediction Error (RMSPE) (Brusilovskiy, 2009). The Root Mean Square Prediction Error is calculated based on the cross-validation statistics, where one known point is excluded from the analysis, and all other known points are used to predict the said point.

Sometimes localized IDW is applied to predict unknown points. In localized IDW, the number of measured points used in the prediction can be limited by using a radius of influence. The radius of influence is assumed to be the maximum distance after which the control points have no influence in the prediction.

Here, IDW was applied to 14 test days and the optimum power was calculated separately for each day. No radius of influence was used. IDW works well when there is good density of point measurements available such as the area around the Great lakes, but its performance in upper Quebec and Labrador is lacking. The advantage of using IDW is in its simplicity. The drawback is that the prediction is totally dependent on the surrounding measured points, and without sufficient control points it may produce garbage values.

4.2.2 Kriging

Kriging is a geostatistical approach to predict the value at an unsampled point from the sampled points around. Kriging is named after Danie Gerhardus Krige, who first presented the ideas in 1951. These ideas were later formalized by Georges Matheron. Kriging has mainly two parts: the first part is called the variogram which is the spatial structure of the data, and the second part involves a fitted variogram to predict unknown values. Details about these two parts are discussed below.

All interpolation algorithms use some kind of weighted sum of sample points to estimate the value at a prediction point. Most assigned weights follow functions that give a decreasing weight with increasing distance from the unknown point (Bohling, 2005). Kriging assigns weight according to a data driven weighted function, in addition to some arbitrary functions (Isaaks and Srivastava, 1989). Kriging produces similar results to other interpolation process if the data locations are dense and uniformly distributed (Bohling, 2005). Like many other interpolation techniques, kriging often underestimates high values and overestimate low values. The main advantage of kriging is that it uses a data derived weighted function. It is possible to generate an error field from the kriging prediction. Due to these advantages, kriging is widely used in environmental science, hydrology, hydrogeology, natural resources, remote sensing, mining, etc.

Variogram

The first part of kriging requires one to fit a variogram. The weighting function of kriging is derived using a fitted variogram. Before fitting a variogram, one needs to understand covariance, correlation and semivariance. Covariance and correlation are measures of spatial similarity between two variables (Bohling, 2005). If one plots inter point distance, h, on the x axis and covariance, c(h), on the y axis, the plot is called a co-variogram. Geostatistical methods incorporate this covariance-distance relationship into the interpolation models to calculate weights (Brusilovskiy, 2009). Semivariance is the measure of dissimilarity of the variables, and a semi-variogram is the plot of distance vs. semivariance (Bohling, 2005). In time series analysis, covariance and correlation functions are mainly used. In geo-statistics semivariance and semi-variogram are more important, as they averages squared differences of the variable, and tend to filter the influence of a spatially varying mean (Bohling, 2005). Semivariance may keep increasing proportionally with lag leading to an infinite global variance which is described more accurately in a semivariogram than a covariogram (Bohling, 2005). Figure 4.1 presents a sample plot of lag vs. covariance, lag vs. correlation, and lag vs. semivariance. The plot is taken from Bohling's (2005) lecture.

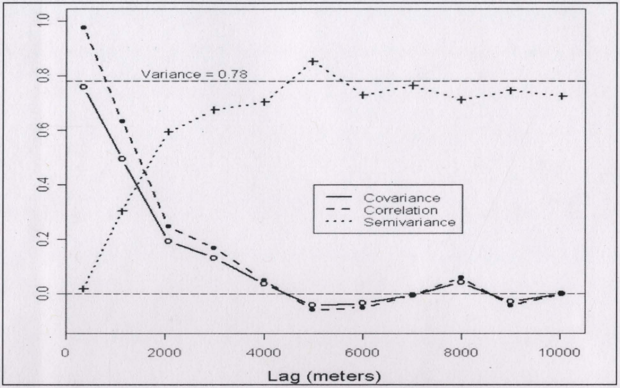


Figure 4.1 Sample Plot of Lag vs. Covariance, Lag vs. Correlation and Lag vs. Semivariance (Bohling, 2005)

If u denotes a vector of spatial co-ordinates, z(u) represents a spatial variable under consideration, h represents the lag vector and denotes separation distance between two spatial locations, z(u+h) represents the lagged version of variable under consideration, and N(h) represents the number of pairs separated by lag h. The semivariance $\gamma(h)$ statistics for lag h can be computed as shown in Eq. 4.1, derived from Bohling (2005).

[4.1] $$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(u_\alpha + h) - z(u_\alpha)]^2$$

<u>Fitting the Variogram Model</u>

Various curve types can be fitted to a variogram model. The curve fitting generally follows the method of least squares, which is often used for multivariate linear regression.

To fit a variogram one needs to understand the characteristics of a semivariogram. There are three parameters to fit a variogram model: sill, range and nugget (Bohling, 2005) as shown in Figure 4.2. The definition of sill, range and nugget is elegantly described by Bohling (2005) in his geostatistics and variogram analysis lecture. The value of the semivariance at which the variogram levels off is known as the sill. The lag distance at which the semivariogram levels off is denoted as the range. Autocorrelation is essentially zero beyond the range. Theoretically, the semivariance value at the origin should be zero. If the value is significantly different from zero when lag is close to zero, then this semivariogram value is known as the nugget. The nugget represents variability at smaller distances and also includes measurement error (Bohling, 2005).

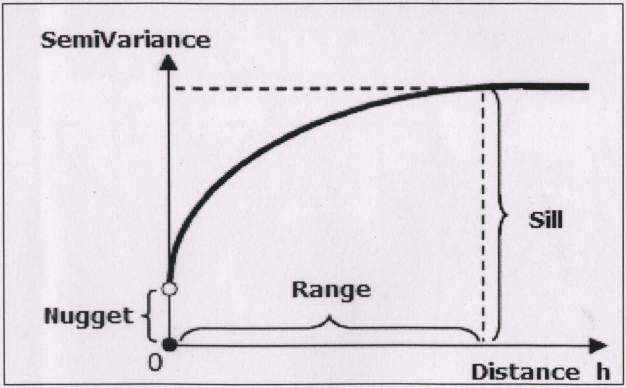


Figure 4.2 Parameters of a Sample Semivariogram

The ratio of the nugget to sill is known as nugget effect, which represents the percentage in variation of data that is not spatial (Brusilovskiy, 2009), while the difference between sill and nugget is known as partial sill. Sometimes, variogram occur with a pure nugget effect when there is absolutely no spatial autocorrelation in the data. This causes the interpolation to produce unreasonable predictions. Singular models may also be produced when there is an infinite number of possible combinations of sill and range (both very large) to fit to a straight line(Bohling, 2005).

For kriging, the empirical semivariogram model needs to be replaced with an acceptable semivariogram model; most common among them are the Spherical, Exponential, Gaussian, and Power models. Figure 4.3 presents the shape of curve fitting produced by Spherical, Exponential, and Gaussian models. If the spatial correlation structures are not the same in all directions, then directional fitting of the semivariogram model is necessary. This effect is known as anisotropy (Bohling, 2005).

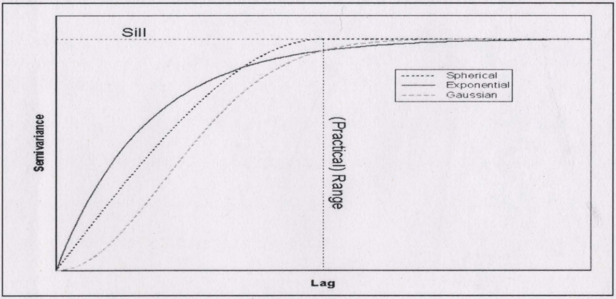


Figure 4.3 Curve Fitting in Different Variogram Models (Bohling, 2005)

Ordinary Kriging

As derived by Bohling (2005), the estimator $Z^*(u)$ is defined as in Eq. 4.2 follows:

$$[4.2] \qquad Z^*(u) - m(u) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha [Z(u_\alpha) - m(u_\alpha)]$$

Where $Z^*(u)$ is defined as the kriging estimator, u and u_α are the location vectors of the estimated point and one of the neighbor data points, respectively n(u) is the number of data points used in kriging, m(u) and $m(u_\alpha)$ are the expected means of Z(u) and $Z(u_\alpha)$ respectively, and λ_α is denoted as the kriging weight assigned to $Z(u_\alpha)$ for estimating at the location of u. Here Z(u) is treated as a random component.

The weights of Eq. 4.2 are calculated by using a fitted variogram. In the case of simple kriging, it is assumed that the trend component is a constant with a known mean, m(u)=m, while in ordinary kriging the mean is assumed to be a constant in the local neighborhood of each estimation point rather than the entire domain (Bohling, 2005).

The basic assumption of kriging is the normality of the dataset. If the dataset violate this assumption, the prediction may be inaccurate, and data transformation will be necessary. The end result of ordinary kriging should produce a continuous surface.

Universal Kriging

More recently, hybrid interpolation techniques have been widely used (Hengl et al., 2007). One of the most promising techniques is known as universal kriging which combines both the regression technique and simple kriging. It is also known as regression kriging or kriging with external drift. Details about regression kriging and its application in spatial analysis are described by Hengl et al. (2007).

According to Hengl et al. (2007), universal kriging (UK), regression kriging (RK), kriging with external drift (KED) all are mathematically equivalent, even though some authors use different terms depending on circumstances. All these methods basically follow the universal kriging model, which was first introduced by Matheron (1969). Originally, the UK term was used if the trend was modeled as a function of coordinates (Hengl et al., 2007). If the drift is defined externally as a function of some auxiliary variables rather than the coordinates, it is commonly referred to as kriging with external drift (KED). The term regression kriging is used in the literature when the drift and residuals are estimated separately and then summed (Hengl et al., 2007). There may be little difference in the definition of these three terms and each produces the same prediction and variance. Here the term universal kriging (UK) will be used.

Spatial predictions are generally made by calculating a weighted average of observations. Eq. 4.3 denotes basic spatial prediction (Webster and Oliver, 2001)

[4.3] $z(u) = \sum_{\alpha=1}^{k} \lambda_{\alpha} . z(u_{\alpha})$

where, z(u) is the estimated value at an unsampled location u, $z(u_{\alpha})$ is the sample data and their locations, and λ is the calculated weights. In ordinary kriging, the weights depend on the spatial autocorrelation structure of the variable and is calculated in such a manner that the prediction error variance is minimized (Hengl et al., 2007). As an alternative to kriging, multivariate regression can be used, which makes predictions by establishing a relationship between the response and auxiliary variables. Again, prediction by regression is based on the weighted average and can be denoted by the following Eq. 4.4. (Hengl et al., 2007).

[4.4] $z(u) = \sum_{i=0}^{n} \beta_i . q_i(u)$

In Eq. 4.4, $q_i(u)$ are values of the auxiliary variables at the target or unsampled location; β_i are the regression coefficients, which can be calculated by generalized or ordinary least square methods; and n is the number of predictors.

In regression kriging or universal kriging, these two approaches are combined. Regression is used to fit the explanatory variation while simple kriging is used to fit the residuals. The prediction by UK can be defined by Eq. 4.5 as offered by Hengl et al. (2007).

[4.5] $z(u) = m(u) + e(u) = \sum_{i=0}^{n} \beta_i . q_i(u) + \sum_{\alpha=1}^{k} \lambda_\alpha . z(u_\alpha)$

Here m(u) is the fitted drift from regression and e(u) is the interpolated residual by simple kriging technique.

Among various interpolation techniques IDW, spline, and OK are relatively simple, and often included in almost all software packages. That being said, when the background information or auxiliary variables are available these methods are less favorable. In such cases regression or the UK method are more suitable. There is also co-kriging, which is suitable when the measurement density of auxiliary variables is greater than that of the target variable. However, if the auxiliary variables are available in a map or grid format, then UK is the best choice.

Kriging in CRAN-R

CRAN-R is a software package that can be used in geostatistical analysis. A well known geostatistical analysis package gstat (Pebesma, 2004) can be incorporated in CRAN R. Gstat contains multiple options for variogram fitting and kriging prediction. There is also a package named automap (Hiemstra et al., 2012) that is based on gstat geostatistical routines, which can be used for automatic fitting of the variogram and automatic kriging. The "fit. variogram" function in gstat is used to fit the variogram. The initial sill is taken as the mean of the max and median of the semivariance. The initial range is defined as 0.10 times the diagonal of the bounding box of the data (Hiemstra et al., 2008). As stated in Hiemstra et al. (2008), the minimum semivariance is taken as the initial nugget. Five variogram models (Spherical, Exponential, Gaussian, Matern family, Matern and Stein's) are iterated to chose the best model based on the smallest residual sum of squares, and the fitted variogram is used to krige the data. The problem associated with automatic kriging is that sometimes negative sill values occur due to a wrong initial guess, and in such cases automatic kriging cannot be applied.

4.2.3 Multivariate Linear Regression

Multivariate Linear Regression (MVR) can also be used as a spatial interpolation technique. The best example of using Regression in SWE interpolation can be found in Fassnacht et al. (2003), where physiographical variables were used as auxiliary variables and SWE was the response. The developed regression equation (Fassnacht et al., 2003) was compared with other interpolation methods and found that the Multivariate Linear Regression worked better than other techniques, such as optimal interpolation.

For SWE prediction in Fassnacht et al. (2003), various physiographical variables were considered; and only those variables that had a good co-relation with SWE were used in regression. These included latitude, elevation, footprint slope, local slope, barrier height, distance from ocean, and forest density (Fassnacht et al., 2003).

4.2.4 Regression + IDW

Fassnacht et al. (2003) also used the regression detrended inverse distance weighting method. It can be called as the "Regression+IDW" method. The method contains two parts: regression and inverse distance weighted interpolation of residuals. Regression was conducted with SWE as the response and other correlated variables as predictors. The residuals were calculated and spatially interpolated, and the final prediction was done by adding the regression predicted values and the interpolated residuals (Fassnacht et al., 2003).

Regression techniques can play an important role in spatial analysis; however, the problem with regression, is its sensitivity to the residual assumptions (normality, constant variance and homogeneity). Automatic interpolation by regression is quite difficult. Regression + IDW seems a very promising method with the same difficulties as in regression. This method is quite similar to UK. In UK, the residuals are spatially interpolated by using Simple Kriging (SK), but in Regression+IDW, Inverse distance weighted (IDW) was used to interpolate the residuals.

4.2.5 Summary

From the basic theoretical review of various spatial interpolation methods, it seems that all the techniques are based on weighted averages, and the prediction accuracy depends on the number and density of control points. Inverse Distance Weighted (IDW) and Ordinary kriging (OK) are simple interpolation techniques that only depend on data points. These techniques may be accurate enough to predict at ranges close to the known points, however by going further the accuracy reduces significantly. Between these two methods, OK seems theoretically more sound as it considers the spatial structure of the known points rather than using some power functions. Regression is a promising interpolation technique, but theoretically Universal Kriging and Regression + IDW are more sound, as regression residuals are considered in these two methods. In UK and Regression + IDW, a number of variables or background fields in addition to the control points, can be used in the prediction.

4.3 Evaluation of Spatial Interpolation Techniques

Based on different SWE prediction studies, five spatial interpolation methods (Inverse distance weighting (IDW), Multivariate Linear Regression, Universal Kriging (UK), Ordinary Kriging (OK) , Regression+IDW) were considered for SWE map production. These methods were compared over 14 test days selected during the period 2008-2010. These test days are selected from each month of snow period between 2008-2010 and also based on data availability. The five spatial interpolation methods were evaluated calculating the RMSE at the independent SWE validation points throughout the transect, as well as with visual consistency checks. For multivariate linear regression, a number of physiographical variables together with CMC snow depth were added as predictors. Our analysis has shown that, the CMC snow depth analysis provided a very good background field for Universal Kriging and similarly as an important variable in Multivariate Linear Regression for SWE interpolation. Among these five spatial interpolation method trials, it was found that SWE prediction by Universal Kriging with CMC snow depth as the predictor produced the lowest RSME result. Table 4.1 details the calculated RMSE for 5 spatial interpolation techniques over the 14 trial days.

Table 4.1 Root Mean Square Error (RMSE) of swe (mm)

Date	IDW	OK	Regression	Regression+IDW	UK
17/03/2008	96.52	130.50	59.08	52.60	55.47
04/02/2008	32.36	25.70	47.45	36.74	30.47
15/01/2008	19.92	20.13	23.02	19.89	21.00
15/04/2008	33.36	37.57	93.48	80.70	65.15
01/12/2008	11.68	12.93	23.44	13.16	13.43
15/01/2009	28.39	24.85	31.06	25.87	25.14
02/02/2009	27.82	28.28	29.54	30.14	28.95
30/03/2009	66.05	65.64	84.53	66.48	66.44
01/04/2009	33.97	33.03	51.66	90.00	31.86
15/12/2009	12.29	15.10	26.26	11.18	13.20
04/01/2010	26.12	25.51	41.54	29.79	27.02
02/02/2010	19.98	19.95	17.10	16.72	20.03
01/03/2010	46.85	52.51	46.99	31.64	34.05
01/04/2010	2.88	2.70	13.42	2.74	3.03
AVERAGE	32.73	35.31	42.04	36.26	31.09

Multivariate regression revealed that in most cases CMC snow depth, vegetation mass, distance from ocean, latitude and longitude were significant to SWE prediction. Residual analysis checks also found that log transformations of data were required in cases with more than 20% zero values. Figure 4.4, presents graphical analysis checks on residuals. These show residuals maintain normality, constant variance, and randomization after the regression.

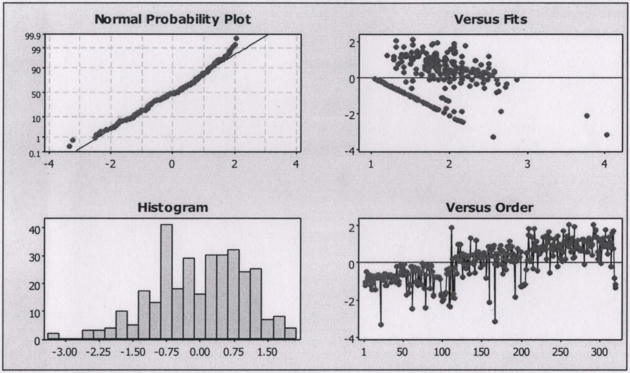


Figure 4.4Residuals Plot after the Regression for SWE Prediction

Variogram fitting plays an important role in kriging. Spatial distribution trends in SWE can be found by constructing the variogram based on the semi variance of the data. For OK, point data plotting is necessary for variogram construction. However, for UK residual variograms are generated via regression. Figure 4.5, shows sample variograms plotted for OK.

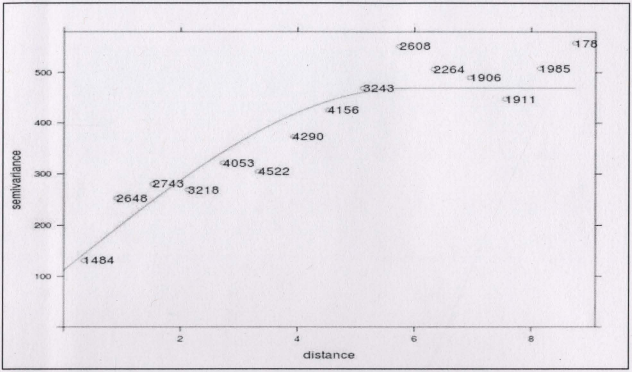
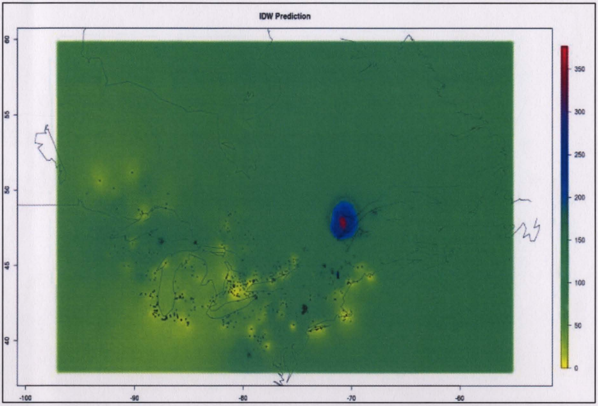
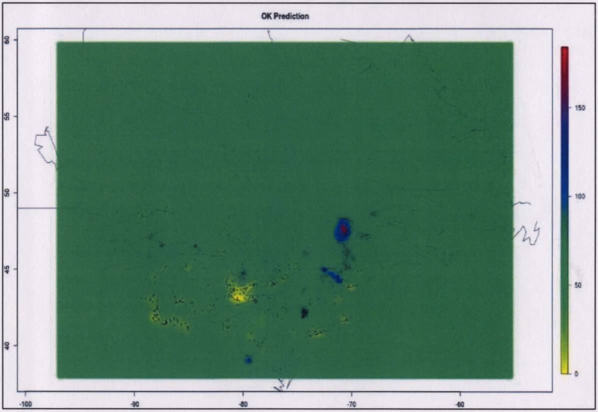
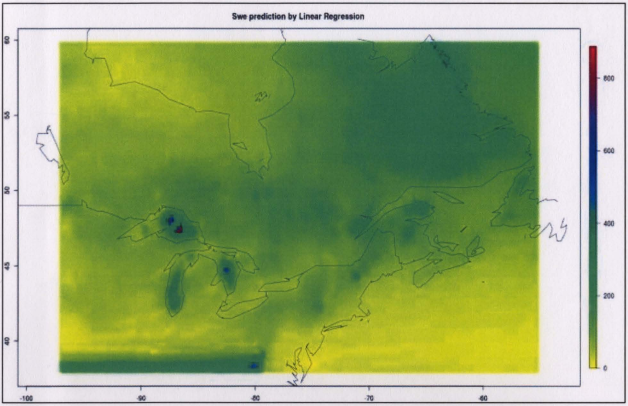


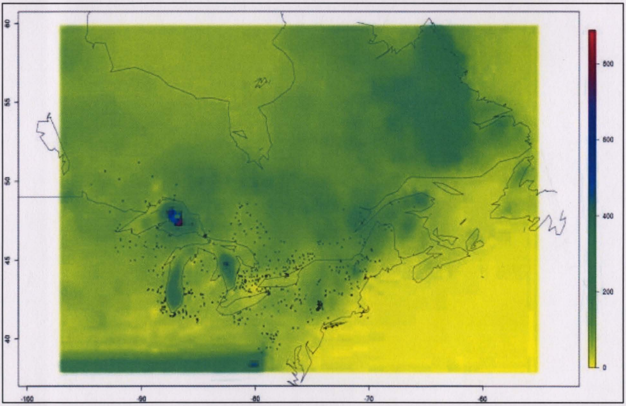
Figure 4.5 Variogram Fitted with Measured SWE at 15 December, 2009 for OK

Universal Kriging (UK) allows a number of background fields to be used as predictors and therefore it can guide the interpolation. Once relationships are established with predictors, the spatial distribution of residuals provide a guide to finalize the interpolation prediction. This method relies on the information extracted by predictor variables and on simple kriging of regression residuals (Hengl et. al 2007). This technique has not been widely used as an assimilation technique with preference given to the Kalman filter, Creesman interpolation, and Statistical Interpolation methods. For this study, UK allows both background fields and point data to be incorporated into the data assimilation process. In fact, any guess field can be selected as a background field provided it is well correlated with SWE.


IDW Prediction
60
55
50
45
40
-100
-90
-80
-70
-60
350
300
250
200
150
100
50
0


OK Prediction
60
55
50
45
40
-100
-90
-80
-70
-60
150
100
50
0


Swe prediction by Linear Regression
60
55
50
45
40
-100
-90
-80
-70
-60
800
600
400
200
0


60
55
50
45
40
-100
-90
-80
-70
-60
800
600
400
200
0

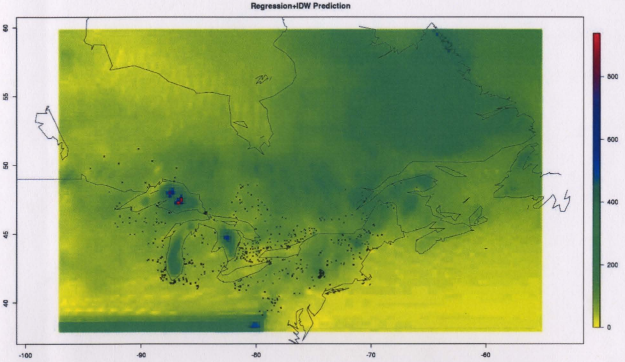


Figure 4.6 Sample SWE Maps Produced on 01 March, 2010 (from top IDW, OK, Regression, UK, Regression+IDW)

Figure 4.6 presents maps produced from different interpolation methods. This figure is important for checking the visual consistency, where it seems OK and IDW can predict well near the control points, but predict relatively poorly when the distance is greater. Regression, UK and Regression +IDW predict SWE adequately, but UK appears to be the most promising method based on the theoretical view as UK use regression and data derived kriging technique.

In this chapter, different spatial interpolation techniques were discussed and UK was selected as the most suitable spatial interpolation technique for further analysis in this research for its simplicity and accuracy. UK is theoretically sound and performs more accurately than other techniques investigated. Automatic interpolation and variogram fitting in UK can be accomplished by using different software packages. UK is better theoretically than Regression+IDW as it uses data derived function to interpolate residuals. For these reasons, UK was selected to develop statistical models in further SWE prediction, the results of which are discussed in Chapter 7.

Chapter 5 Snow Depth Adjustment

5.1 Introduction

In Canada, real time snow information is necessary for important services such as weather forecasting and water resources management. To meet this need, a real time operational snow depth analysis was established at the Canadian Meteorological Center (CMC): the global gridded daily snow depth analysis (Brasnett, 1999). An initial guess for snow depth was first produced from a simple snow pack model then combined with collected snow depth point measurements using the spatial interpolation technique (Daley, 1991). The final product is a 24 km resolution daily snow depth analysis (Brasnett, 1999).

According to Brasnett (1999), there is a bias in the CMC snow depth analysis product, which is caused primarily by inaccurate precipitation forecasting and a lot of representative snow depth measurements from open areas. Most of the snow depth stations are located in urban areas and in valleys, and it is assumed that the snow depth in mountainous terrain or other regions were not well represented by the CMC snow depth analysis (Brasnett, 1999).

In the current research, the CMC snow depth analysis was shown to be strongly correlated with the Snow Water Equivalent (SWE). One of the main reasons behind CMC snow depth bias, since most measurements of snow depth are drawn from open areas, such as airports (Brasnett, 1999). In the current study, the CMC snow depth data is altered to reflect snow depth conditions from the dominant vegetation of the region.

In the current research, the accuracy of the CMC snow depth was validated by calculating the Root Mean Square Error (RMSE) at the snow course points. As expected, it was found that the CMC snow depth contains a significant amount of error. Subsequently, a methodology was established from the current study to adjust the CMC snow depth product within a tolerable distance of snow course measurements. The goal of the current analysis was to improve the CMC snow depth product within a distance of 150 km of the measured snow course points. The tolerable distance is assumed based on localized influence from a previous SWE prediction study (Fassnacht et al., 2003) and the change of topography. It should be noted that, the CMC product was generated by using a standard data assimilation methodology (Brasnett, 1999) and it was not the goal to improve this product in the full spatial scale.

5.2 Methodology

The current research aimed to adjust the CMC snow depth by incorporating snow course measurements. The plan was to adjust the CMC snow depth within a close range to the snow course stations and only on those days when Canadian snow course data were available. Initially, the adjustment process was conducted using 14 days spanning 2008 to 2010, which were selected to be representative of differing conditions associated with the snow course data set. The snow depth was estimated at snow course stations using nearest neighbor interpolation from the gridded CMC snow depth. Then, the RMSE was calculated for each of the test days from the estimated and measured snow depth values, resulting in an estimation of the error associated with CMC snow depth. The objective was to reduce this error around snow course stations. Two methods were attempted for the adjustment process. The first method was based on multivariate linear regression and the second one was based on Universal Kriging (UK) interpolation. It seemed that the Universal Kriging method was easy to use and more promising. However, the conclusion drawn from the regression helped to establish the UK models. Details about these two methods and the final adjustment methodology is described in the following sections.

5.2.1 Multivariate Linear Regression

A multivariate linear regression was developed using daily snow depth data from snow course sites as the response variable, and estimated CMC snow depth as the main predictor or auxiliary variable. Other physiographic variables, including vegetation mass, elevation, barrier height, local slope, aspect distance to ocean and distance to water were also used as predictors. The regression equation was developed separately for each test day. Following the regression, the RMSE was calculated daily based on measured and newly predicted snow depth values and compared with the previously calculated RMSE value for CMC snow depth. During most of the test days, the RSME values were reduced in the newly predicted snow depth.

In order to fulfill the assumptions regarding residuals in regression, it was necessary to transform the snow depth data (both CMC and measured snow course) using the Box-Cox or log transformation. Both transformations produce singularities with zero values requiring that 1 be added to each snow depth before transformation.

The regression showed that in most cases vegetation mass and distance from the ocean have a significant relationship with snow depth. Log transformation was necessary in cases of more than 20% zero snow depth values, and if there was more than 50% zero values then the adjustment did not produce any improvements. Some sample regression equations and the analysis are provided in Appendix-B.

5.2.2 Universal Kriging

Universal Kriging (UK) was evaluated based on the effectiveness and benefit associated with UK in spatial analysis. The benefit of applying UK was described in Chapter 5. The effectiveness of applying UK for the CMC snow depth adjustment was evaluated on the selected 14 day samples. Two UK-based strategies were developed to incorporate snow depth from the snow course sites. The first model simply used the relationship between the measured snow depth and the CMC snow depth analysis. The second model added other predictors, such as distance from ocean, and vegetation mass to improve results. It need to be noted that all the physiographical variables including elevation, aspect, slope, vegetation mass, distance from ocean, distance from water body, vegetation type are evaluated but only vegetation mass and distance from ocean was found significant in test days. The selection between the two models was based on the calculated RMSE of snow depth in validation and cross validation points. It should be noted that the CMC snow depth adjustment was implemented only on those grid points within 150 km of a snow course station. Outside the 150 km radius, CMC snow depth was used directly. Thus, the new snow depth product represented a blend of both data sources and provided an improved background field for SWE assimilation.

By comparing the RMSE of the regression and UK methods, it was found that UK produced the best prediction of snow depth. Theoretically, UK is also more sound and easy to use as it is regression with simple kriging of residuals. A single month March, 2009; was selected to test the UK method developed above. A cut-off distance of 150 km around the snow course data points was applied. If at any grid points, there was no measured data within 150 km, then the CMC snow depth product was used directly rather than adjusting at those grids. Based on these criteria, an R script was developed to automatically adjust the CMC snow depth. The code automatically adjusted the snow depth and calculated RMSE on cross-validation and validation points, and compared it with the RMSE associated with the CMC snow depth data.

5.3 Results

In this section, the effectiveness of the CMC snow depth adjustment is discussed. It was found that UK could reduce the bias significantly compared to the Multivariate Linear Regression (see Table 5.1). In regression separate regression equations need to be developed manually for each day, but automation with UK is easier for longer simulation period. In this section, the results generated from the UK methods are mainly discussed. First, the adjustment process was evaluated within 14 trial dates. Table 5.1 presents the RMSE error in CMC snow depth before and after adjustment. In this table, the error associated with adjusted CMC snow depth was obtained by two separate UK models and also by regression. In most of the cases, UK can adjust the CMC snow depth quite significantly especially around the snow course points. It was also verified that if one goes further from the snow course points, the adjustment is not significant. Therefore, a cut-off distance of 150 km was assumed in the adjustment process. An R script was developed to automate the adjustment process. The script evaluates the two UK models based on the calculated RMSE at validation and cross validation points, and the best model was selected for the adjustment. In the case of an increasing RMSE error after the adjustment, the CMC snow depth was used directly for future analysis. This methodology was applied in a trial month period. Figure 5.1 presents the RMSE error of CMC snow depth and the revised product during March, 2009 which was calculated based on snow course sites only. 11 days were provided among 14 days as adjustment was not possible on the rest three days due to number of zero values resulting in increasing the error. .

Table 5.1 RMSE of CMC snow depth (cm) before and after Adjustment

Date		RMSE Error of Snow depth (cm)			
		CMC Snow depth	Linear Regression	UK1	UK2 (with physiographical variables)
01/03/2010	validation	27.00	18.08	18.14	18.39
	Cross-Validation	21.48	18.80	13.47	13.90
01/12/2008	validation	9.08	11.34	4.53	5.23
	Cross-Validation	8.63	7.90	1.04	1.11
02/02/2009	validation	14.89	7.80	9.68	9.50
	Cross-Validation	18.96	11.48	10.00	9.40
02/02/2010	validation	20.78	7.00	7.21	6.98
	Cross-Validation	15.07	11.70	0.76	0.78
04/01/2008	validation	10.56	10.73	7.19	7.67
	Cross-Validation	16.55	13.03	12.81	12.97
04/02/2008	validation	23.25	15.61	14.28	20.00
	Cross-Validation	20.94	16.34	16.23	14.00
15/01/2008	validation	7.018	6.84	5.83	5.57
	Cross-Validation	17.48	16.84	1.50	1.36
15/01/2009	validation	13.80	8.66	8.34	7.66
	Cross-Validation	13.10	8.28	8.51	8.36
15/12/2009	validation	8.43	8.65	5.21	5.82
	Cross-Validation	11.80	14.56	1.48	1.34
17/03/2008	validation	37.90	22.12	36.19	25.09
	Cross-Validation	29.97	16.32	14.36	13.42
30/03/209	validation	40.57	26.34	27.18	21.12
	Cross-Validation	24.66	23.22	2.03	1.72

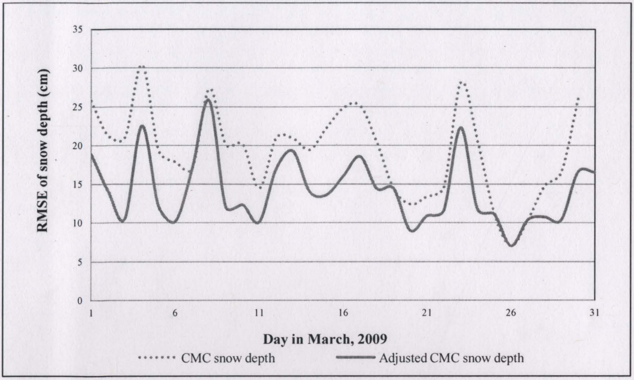


Figure 5.1 RMSE Error of CMC Snow Depth (cm) before and after Adjustment

5.4 Conclusion

The current research project was mostly dependent on the accuracy of the CMC snow depth product as it was used as the main predictor of SWE. Brasnett (1999) used a statistical interpolation technique to incorporate available snow depth measurements. The CMC analysis was produced with most of the available snow depth measurements, but these snow depths were mostly measured in open areas and valleys, resulting in a bias. In the current analysis, it was attempted to adjust the CMC snow depth by incorporating unused snow course points. It should be noted that it was not possible to reduce the CMC product bias in the full spatial and temporal scale based on only snow course measurements as snow course measurements are not spatially well distributed and only available twice per month. The goal was to reduce the bias only around the snow course points, not to reduce the bias in the total CMC product. Although, the conducted adjustment reduced bias over a small spatial and temporal scale, it was still beneficiary for the current research.

Chapter 6 SWE Prediction

6.1 Introduction

In this chapter, a snow water equivalent (SWE) prediction methodology is presented based on a review of existing literature (Chapter 2) and evaluation of different spatial interpolation techniques (Chapter 4). Existing deterministic models are combined with developed statistical models. Based on the work in Chapter 4 Universal Kriging (UK) was selected for spatial interpolation with the statistical models. Two different scenarios were developed with the expectation that the final SWE output should be a smooth surface. An R script was developed from the current research to combine all the data, adjust CMC snow depth, and apply the SWE prediction methodology described in this chapter. The approach can produce daily gridded SWE maps, predicted SWE, snow density and corresponding snow depth files in text format, and Root Mean Square Error (RMSE) values at validation and cross-validation points. In this chapter, the selected deterministic models, developed statistical models, scenario development, and the final SWE prediction are discussed.

6.2 Deterministic Models

Two deterministic models were selected to predict snow density. Inspite of using complex snowpack models (Brasnett, 1999; Brown et al., 2003), simple snow density prediction models are used. The model based on the snow climate class (Sturm and Holmgren, 1998), derived by Sturm et al. (2010) was selected as it was simple and could predict the SWE with acceptable accuracy. This model was developed by considering adequate field measurements from the US and Canada, and it was expected that the model prediction would represent the actual conditions. Snow course data was not used in the Sturm et al. (2010) model development, but it was used in validation. The second deterministic model applied in this research is based on the snow aging formula used in the Canadian Land Surface Scheme (CLASS) (Verseghy et al, 1993). These two models are discussed in the following sections.

6.2.1 Snow Climate Class Model

Two deterministic models have been used to predict the spatial distribution of SWE. The first is based on Snow Climate Class from Sturm et al. (1995). Sturm et al. (2010) present a bulk snow density estimator based on the Snow Climate Class (Sturm et al. 1995), day of year (DOY), and snow depth. From this predicted density and the adjusted CMC snow depth from Chapter 5, the SWE was spatially predicted. These SWE predictions can be used directly to produce SWE maps or can be used as background fields in statistical UK models. The equation presented by Sturm et al. (2010) is as follows:

[6.1] $$\rho_{h_i,DOY_i} = (\rho_{max} - \rho_0)[1 - \exp(-k_1 \times h_i - k_2 \times DOY_i)] + \rho_0$$

In this equation, ρ is the predicted density in gcm^{-3}, h is the snow depth in cm, ρ_{max}, ρ_o, k_1 and k_2 are model parameters selected based on snow climate classes and assigned values as noted in table 6.1 from Sturm et al. (2010). Because the winter season spanned two calendar years, the DOY runs from -92 (1 October) to +181 (30 June) excluding the 0 value (Sturm et al. 2010).

From the predicted density and CMC snow depth, the SWE at each prediction grid was easily predicted. This method is very simple and can predict SWE quite accurately. Details about the prediction accuracy are discussed in the literature review (Chapter 2).

Table 6.1 Parameter Table for Snow Density prediction (Sturm et al., 2010)

Snow class	ρ_{max}	ρ_0	k_1	k_2
Alpine	0.5975	0.2237	0.0012	0.0038
Maritime	0.5979	.2578	0.001	0.0038
Prairie	0.5940	.2332	0.0016	0.0031
Tundra	0.3630	.2425	0.0029	0.0049
Taiga	0.2170	0.2170	0.0000	0.0000

6.2.2 Snow Aging Model

An alternative deterministic model is the snow aging model approach. The snow aging used in CLASS (Verseghy et al., 1993) was used here to predict the change in density with time. Adjusted CMC snow depths, were differenced to yield daily increases or decreases in snow depth. Thus new snow was separated from old snow. The density of aged snow was calculated based on snow aging and the density of new snow was assumed as to be 100 kg/m^3.

The maximum snow density $\rho_{s,max}$ was estimated as a function of snow depth z_s according to Tabler et al. (1990) and represented in equation [6.2]:

[6.2] $$\rho_{s,max} = A_s - \left[\frac{204.70}{z_s}\right]\left[1.0 - \exp\left(-\frac{z_s}{0.673}\right)\right]$$

The empirical constant A_s was assigned a value of 450.0 for cold snow packs, and 700.0 for melting snow packs (Brown et al. 2006). In Eq. [6.2] z_s is the snow depth and ρ_{max} is the maximum density. The density of snow ρ_s for snow aging is assumed to increases exponentially with time (Longley et al., 1960; Gold et al., 1958) and is reflected in Eq. [6.3] with t as the time step and Δt used as 24hr = 86400 sec for the daily prediction.

[6.3] $$\rho_{s(t+1)} = [\rho_{s(t)} - \rho_{s,max}]\, exp[-0.01\, \Delta t/3600] + \rho_{s,max}$$

These two deterministic methods were selected based on their simplicity and adequacy in snow density prediction.

6.3 Statistical Models

The statistical models developed in the current research are based on Universal Kriging spatial interpolation technique. Background variables, possessing good correlation with SWE, are used to develop the models. Among these background fields, CMC snow depth is the main predictor. Other background fields are selected from various physiographical variables, and the significant physiographical variables are selected by using a multivariate linear regression technique on 14 test days. It was found that latitude, longitude, CMC snow depth, vegetation mass, and distance from ocean were good predictors of the SWE. The correlation between the predictor and response varied on a daily basis (see Appendix-B), however from the regression analysis it seemed that the selected auxiliary variables had significant relation with SWE in most of the test days. Some sample regressions on test days were provided in Appendix-B. Statistical models are developed in the current analysis using different combinations of predictors which is provided in the scenario development section 6.4. Predicted SWE from deterministic models were also used as predictors or backgrounds field in the statistical analysis.

6.4 Scenario Development

Two scenarios were developed to predict the SWE that depended on the number of snow course data points available for interpolation. Recall that snow courses in Canada are typically measured twice per month during the snow accumulation period. The first scenario (Scenario-I) was applied on those days when more than 50 Canadian snow course measurements were available. The number 50 was chosen just to differentiate between two scenarios in coding. For Scenario-I, only the snow climate class deterministic model was employed. Four UK models were evaluated within this scenario. The background fields used in these UK models include of the predicted SWE from snow climate class model, the adjusted CMC snow depth, the CMC snow depth and physiographical variables, and the predicted SWE from the snow climate class model and physiographical variables. The second scenario (Scenario-II) was applied when fewer than 50 Canadian snow course data are available. In this scenario two deterministic models: the snow climate class and snow aging models were used. The snow aging model calculated snow density based on the increasing or decreasing snow depth and required a spatial SWE field from the previous day. In the second scenario, only the generated SWE from the deterministic models were used as predictors, rather than using the CMC snow depth or other physiographic variables . It was found that as there were relatively few or no Canadian data points in second scenario, statistical models would have been difficult to develop and deterministic models would have worked more accurately. For this reason, scenario-II statistical models were developed by incorporating point measurements into the deterministic model prediction, rather than using CMC snow depth or other physiographical variables as predictors due to the less snow course points. Therefore, in scenario-II two statistical UK models were developed considering background fields as generated SWE from two deterministic models. In both scenarios, the best deterministic model and best statistical model were selected based on calculated the RMSE values and then later combined based on the Cressman interpolation algorithm (Cressman et al. 1959). A summary of developed scenarios are provided in Table 6.2 below.

Table 6.2 Summary of Scenario Development

Scenario	Criteria	Deterministic Models	Statistical Models	Predictors/Background Fields
Scenario-I	50 snow course measurements Or more	Snow Climate Class Model	UK 1	Predicted SWE from Snow Climate Class
			UK 2	Adjusted CMC snow depth
			UK 3	Adjusted CMC snow depth, latitude, longitude, vegetation mass and distance from ocean
			UK 4	Predicted SWE from Snow Climate Class, latitude, longitude, vegetation mass and distance from ocean
Scenario-II	less than 50 snow course measurements	Snow Climate Class Model	UK1	Predicted SWE from Snow Climate Class
		Snow Aging Model	UK2	Predicted SWE from Snow Aging Model

6.5 Cressman Data Assimilation

The final SWE product was produced from the combination of statistical and deterministic model output. Finally, the best statistical and deterministic models were selected based on calculated RMSE values at validation and cross validation points. Then, Cressman data assimilation was applied to combine both models prediction. In scenario-I only the snow climate class model (Sturm et al. 2010) was used for direct density prediction, while in scenario-II both the snow climate class and snow aging model (Verseghy et al, 1993) were used. As stated above, in scenario-I the best statistical model was selected from four statistical UK models and in scenario-II from two models. Cressman analysis technique (Cressman et al. 1959) is based on weighted distances. The weights are calculated for the point measurements by $w= (R^2-r^2)/(R^2+r^2)$, where w is the weight to the point, R is the cut-off distance after which the point has no influence at all, and r is the minimum distance between the measured points and the prediction grid. In the current research, the basic concept of the Cressman analysis technique was applied in a modified way. The minimum distance between the grid and all other snow course measurements were calculated and denoted as r. This needed to be calculated at each prediction grid and on a daily basis. The value of cut-off distance R was assumed separately for each scenario beyond which the measured points had no influence on prediction grids. The influence of snow course at grid points was incorporated by using the predictions from UK models. Thus weights are calculated at each prediction grid and these weights were assigned to the predicted SWE from the statistical model. After subtracting the calculated weight from the total weight, a subtracted weight was assigned to the deterministic model prediction. The final SWE at each grid point was calculated by combining statistical model and deterministic model results.

The cut-off distance used in Cressman analysis was applied to construct the range of the effect of the statistical models or influence of measurement points. In the grid points beyond the cut-off distance, only one prediction from the best deterministic model was used as the final SWE. The importance of cut-off distance was obtained by conducting a simple analysis. In this study, most of the snow course points were around the North American Great Lakes region. Great Lakes data was used to predict the SWE in Labrador and it was found that the SWE predicted from the deterministic model was more accurate than the results from the statistical model. From this, it was evident that data points only had influence within a limited area, and that influence decreases with increasing distance. As such, a cut-off distance needs to be fixed after which the snow course points will have no influence in prediction. Two cut-off distances were assumed for two different scenarios. In scenario-I the cut-off distance was selected as 500 km and in scenario-II, it was assumed as 1000 km. The cut-off distance in scenario-II was assumed greater than scenario-I due to the unavailability of Canadian snow course points for scenario-II. This cut-off distance was assumed based on a previous SWE prediction by Fassnacht et al. (2003).

6.6 Final SWE Prediction

In this chapter the main methodology of predicting SWE from adjusted CMC snow depth is described. However, the total SWE prediction methodology contains both the data collection part in Chapter 3 and CMC snow depth adjustment in Chapter 5. This complex methodology is compiled in a prediction tool which is a combination of FORTRAN codes and R scripts. Details about this prediction tool is described in Appendix C but the full SWE prediction methodology used in the prediction tool is described in the flow chart below (Figure 6.1). The tool is solely developed by the author and the copyright of the tool is belonged to the author only. This SWE prediction tool is named as SWE_MAP_V1.0 and is provided in a DVD with the thesis.

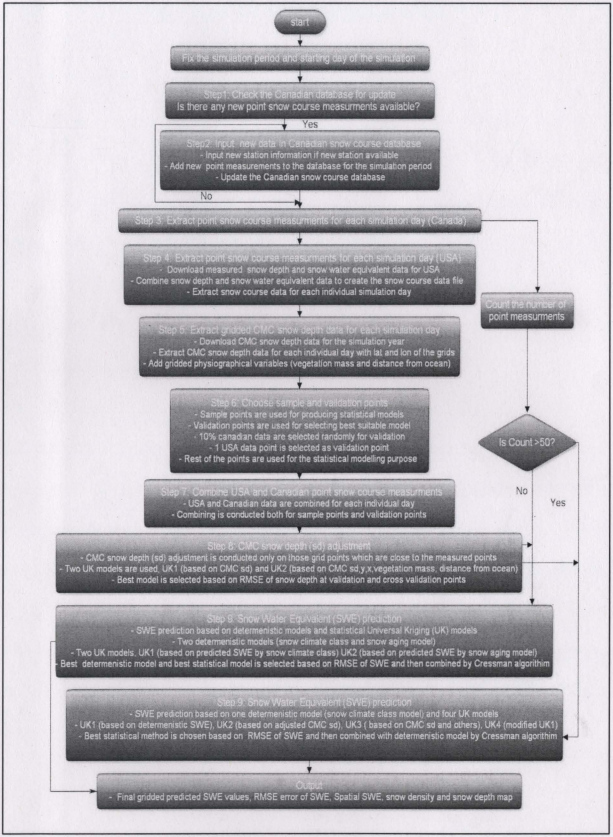


Figure 6.1 Flow Chart Using in SWE Prediction Tool

Chapter 7 Results and Discussions

7.1 Introduction

In this chapter, the output of the current research and its effectiveness is discussed, including a critical analysis of achievements and drawbacks. The primary objective of this work was to develop a Snow Water Equivalent (SWE) prediction tool, which could produce spatial SWE analysis on a daily operational basis. In addition to the SWE analysis, other outputs were also generated including:

- An adjusted Canadian Meteorological Center (CMC) snow depth analysis product
- Corresponding snow density
- Spatial SWE, snow depth and snow density maps from the prediction

All the outputs were generated on a daily basis within a selected time period and within the selected transect from the Great Lakes to Labrador through Quebec. Adjusted snow depth was predicted in cm, SWE and RMSE in mm, and snow density in g/cm^3. All these outputs were generated in a text file format and the maps were generated in jpeg format. The SWE analysis output is evaluated and validated on the basis of a 3-year (2008-2010) model run. From the critical analysis of the outputs, the research showed quite accurate results in spatial daily prediction of SWE, however, there were some limitations in finer scale, as discussed below.

7.2 Accuracy of the SWE Analysis

The accuracy of the SWE analysis was determined by calculating the RMSE values of the predicted SWE. The RMSE is widely used in spatial analysis to find out the accuracy of the spatial prediction (Fassnacht et al., 2003). In the current research, the difference between predicted and measured values were calculated in both validation and cross-validation points on a daily basis, and the final RMSE values were calculated by combining errors at all validation and cross-validation points. The validation points were selected only in scenario-I, where a significant amount of Canadian snow course measurements were available. These validation points were not used in the development of the predictive tool and were selected at random. Additionally, cross-validation, where a point is isolated and other points are used to predict, was also applied. The average RMSE values obtained for the snow period (December-April) was 37.22 mm for 2008, 33.31 mm for 2009, and 31.93 mm for 2010. All the errors were calculated at validation and cross-validation points and tended to increase with greater distance from snow course stations.

As shown below, Figure 7.1 represented RMSE values for the SWE prediction. January, February and March of the years 2008, 2009 and 2010 were selected for the plotting as they were the main periods when significant amounts of snow are on the ground. While there were no significant patterns in the RMSE values with time, there appeared to be increasing error as the snow season advanced. This was consistent with greater complexity of the snow pack. The RMSE depends primarily on the number of control points and their spatial distributions. As such, the SWE prediction accuracy also depended on the number and densities of snow course points. The highest RMSE found in each of the three years was around 90 mm. That being said, if one used the measured points to predict SWE at distances greater than 500-1000 km from the snow course stations then the RMSE may have increased to roughly 100 mm or higher on the peak snow period. January to March is considered for plotting as there is zero RMSE values at December, April and May.

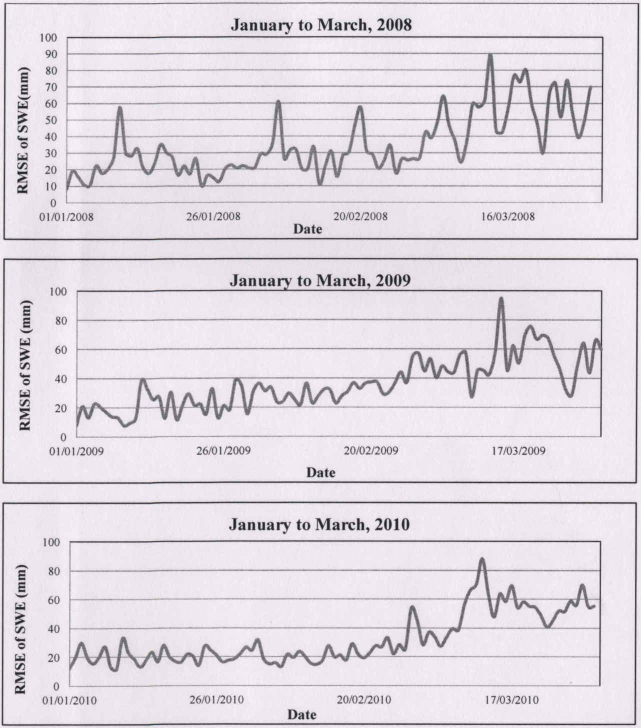


Figure 7.1 RMSE Values for SWE Prediction at First 3 Months of 2008,2009,2010

7.3 Point Validation

Three independent stations were selected to compare observed and simulated values. These stations had not been used in the analysis. The SWE was predicted from the final gridded SWE output by using a nearest neighbor interpolation technique. The locations of these stations are shown in Figure 7.2 in three parts of Canada: Ontario, Quebec, and Labrador. This was done to validate the SWE in larger spatial scales. The period from 2008 to 2010 was selected for this point validation and three snow months (January, February, and March) were primarily considered. April was also considered for 2008 and 2010.

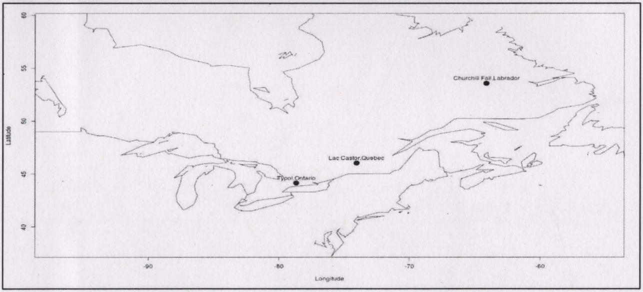


Figure 7.2 Independent Validation Station's Location

In Figure 7.3 the comparison was made between the simulated and observed values for the period of January-April, 2008. It seemed that the prediction was generally satisfactory based on the graphical plot for all three stations, while SWE was slightly underestimated in Churchill Falls and Lac Castor. In both Churchill Falls and Lac Castor the simulation showed sudden melting of snow, but the observed value suggested slow snow melt. In Typol the simulation slightly over-predicted at certain locations, as shown in Figure 7.3.

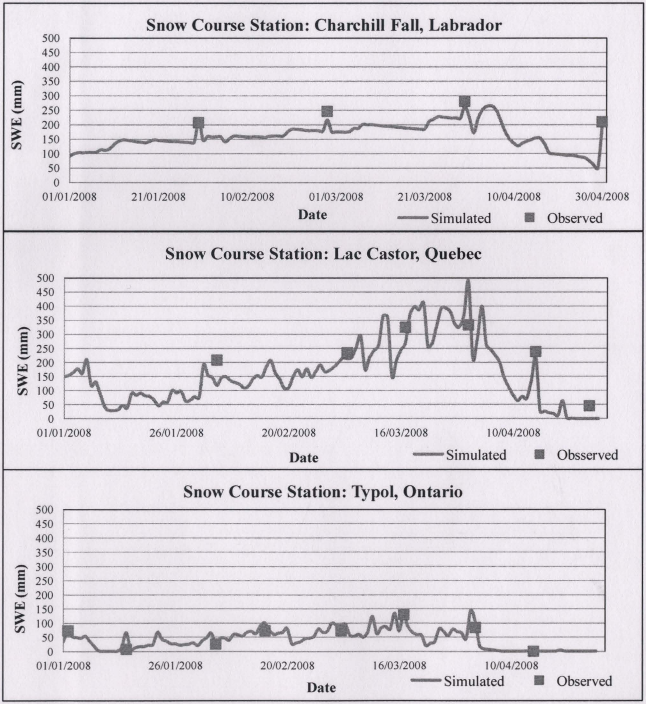


Figure 7.3 Comparison Between Simulated and Measured SWE at 2008

Figure 7.3: Similarly, Figure 7.4 represents the comparison for 2009 and Figure 7.5 for 2010. Results were consistently good for Churchill Falls and the Typol station in 2009, however, there appeared to be a large under- prediction of SWE for Lac Castor.

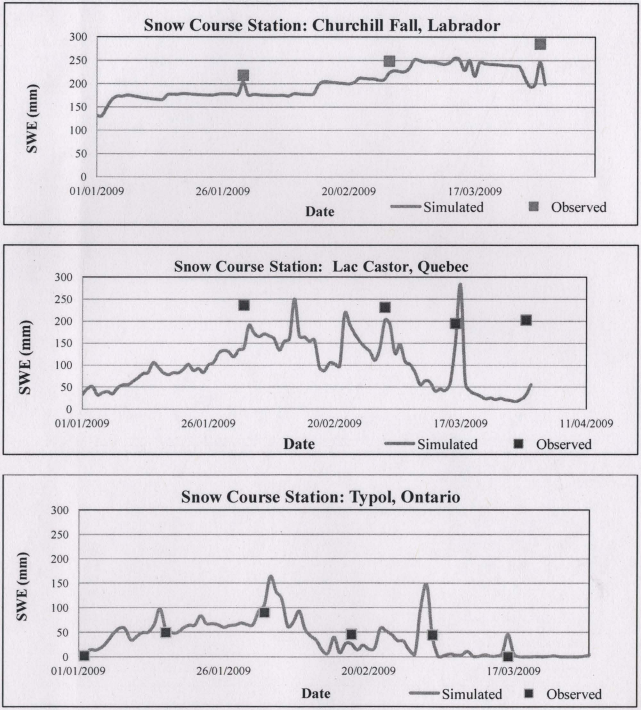


Figure 7.4 Comparison Between Simulated and Measured SWE at 2009

The SWE prediction was less accurate in 2010 than in 2008 and 2009. That being said, results were typically good for Churchill Falls and Typol except on one day. There does appears to be a consistently large under prediction of SWE for Lac Castor towards the end of season.

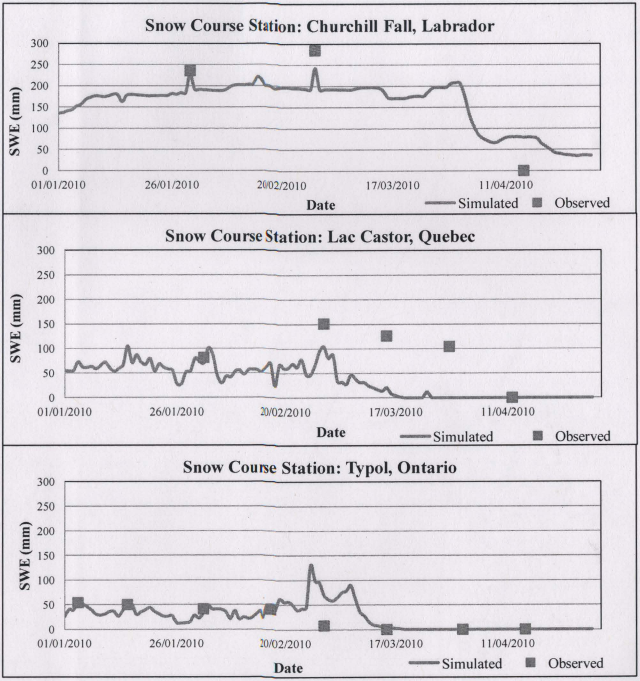


Figure 7.5 Comparison Between Simulated and Measured SWE at 2010

From the above discussion and graphs, it is evident that there are some limitations or drawbacks to the SWE prediction. SWE was poorly predicted in 2010 compared to 2008 and 2009. One possible explanation may be that in 2010 there was less snowfall and the background fields generated from deterministic models might have appeared to be a large under prediction of SWE due to smaller snow depth values than usual, which would affect the final prediction. The SWE prediction seems inconsistent in Lac Castor. The main reason behind this is likely the complex topography of Quebec. More detailed topographic analysis might be necessary in this case.

7.4 Comparison with CMC Monthly SWE Estimation

The CMC has developed a monthly SWE estimation product, based on the snow depth analysis (Brasnett, 1999) and a density lookup table (Brown et al., 2003), though no snow course data has been used in their SWE estimation. In this section, a comparison is made between the CMC monthly SWE and the daily SWE analysis resulting from this work. In order to perform the comparison, the daily SWE was constructed into a monthly scale for 2008. Both large spatial and snow course station comparisons were conducted, the result of which are presented in Figures 7.6 and 7.7. The mean monthly observed value at the snow course station was also included as a point of reference. As snow course measurements were conducted twice per month, these monthly averaged values are not the representative mean of daily values for the whole month. From the graphs below, it shows that the monthly average of the SWE analysis is very much similar to the CMC monthly SWE estimation.

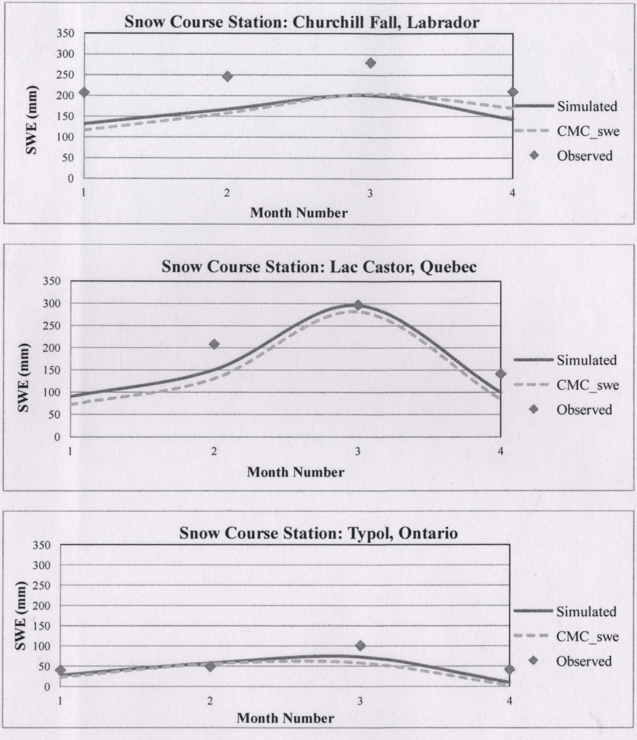


Figure 7.6 Comparison Between Monthly Average Simulated and CMC SWE

Figure 7.7 presents a comparison of the spatial monthly average SWE maps from the current simulation and monthly CMC SWE estimations for January, 2008.

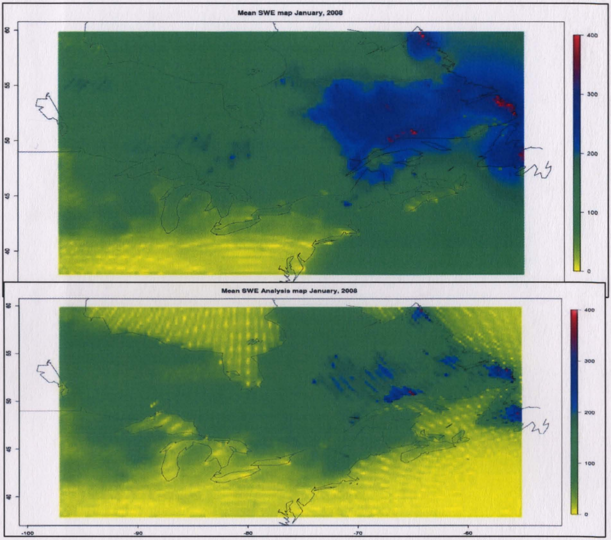


Figure 7.7 Monthly Average SWE Map (Top- CMC SWE map, Bottom- Current Simulated SWE MAP)

From the above figure, it seems that the simulated SWE is slightly less than that of CMC, though the spatial variation is similar. It is difficult to say which monthly average estimation is more accurate, as snow course measurements were only available twice per month.

7.5 Validation with Respect to Snow Cover

Satellite measurements of snow cover are currently available at a relatively high resolution. However the current research is primarily based on the CMC snow depth analysis while the CMC product was independently validated with available snow cover data (Brasnett, 1999). Therefore, a random selection of daily data was chosen to validate results from this work. The National Snow and Ice Data Center (NSIDC) snow extent cover data was useful for this purpose. Figure 7.8 represents a validity check for 20 December, 2009. The map on the left was generated from the NSIDC snow extent data where the value of 4 denotes that there is snow on the ground. On the right side, the SWE map is generated from the current simulation. Both maps show that only the US portion below the Great Lakes has no snow.

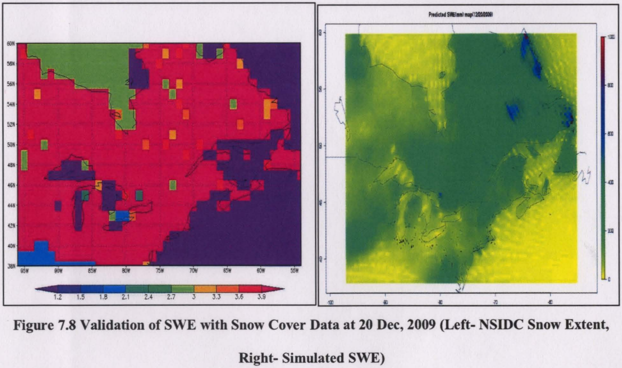


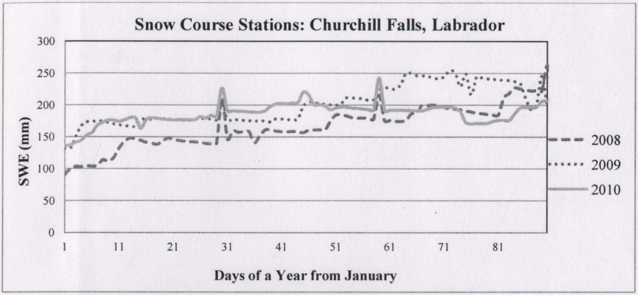
Figure 7.8 Validation of SWE with Snow Cover Data at 20 Dec, 2009 (Left- NSIDC Snow Extent, Right- Simulated SWE)

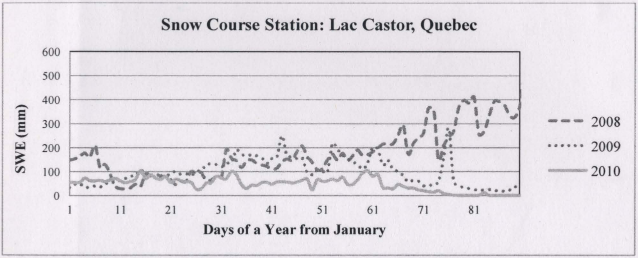
7.6 SWE Trend Analysis

The developed daily SWE analysis can be useful for the SWE trend analysis, which can be conducted from the generated SWE maps. Additionally, graphs can be plotted for individual stations, to analyze time series trends on a finer scale. From this kind of time series trend analysis, the change of SWE over years can be visualized, and can be used in decision making for the weather forecasting and climate change study. Graphs can be plotted for multiple stations, to identify temporal trends as well as location dependency.

In Figure 7.9, time series plots of the SWE for January to March from 2008-2010 for three locations are provided. While, no patterns were apparent from the graphs, they were still helpful in identifying SWE peaks and other important decision making inputs.

In Figure 7.10, time series of the SWE is plotted for three different stations. From these plots, one can understand how the SWE varied over time and the difference between stations. The graphs were plotted by taking the SWE analysis for the period of January-March, 2008. It appears that the peak SWE for all three stations was reached coincidently. Sudden variations in the SWE was observed for Lac Castor, Quebec while, the other two stations showed less extreme changes in SWE with time.





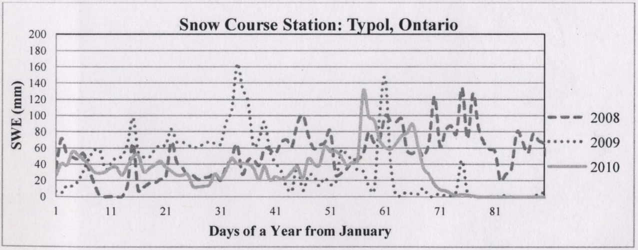


Figure 7.9 Variation of SWE from January-March for 2008, 2009, 2010

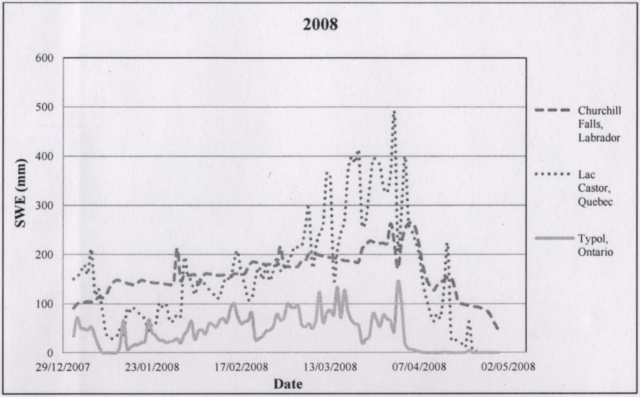


Figure 7.10 Time Series of SWE at Three Stations for 2008

7.7 SWE Map

The final output of the current research was daily SWE maps. Visual consistency checks of the SWE are necessary to validate the SWE analysis. These SWE maps can also be used to identify spatial and temporal changes in the SWE and can play an important role in decision making.

In Figure 7.11 the SWE maps are provided for the first 6 days of January, 2008. It seems that the variation of SWE was quite consistent over time and space. The main drawback from the generated maps is that, the SWE can vary unexpectedly and suddenly. It seemed that sometimes the melting of snow was very sudden as snow appeared and disappeared very suddenly in certain regions. This inconsistency is the major concern of the current research.

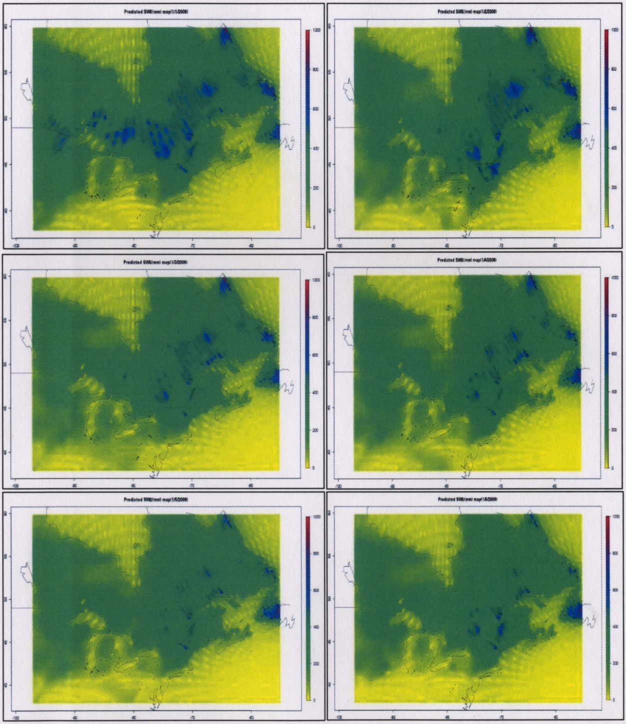


Figure 7.11 SWE Maps from 01 Januray-06 January, 2008 (Start from top left Corner)

7.8 Achievements and Limitations

In this section the achievements and drawbacks or limitations of the current research are discussed.

The achievements of the current research are as follows:

- A daily gridded spatial SWE and snow density product was created.
- The point SWE measurements were successfully transferred into a spatially consistent dataset.
- The time series gaps between SWE measurements were constructed.
- An automated SWE prediction tool was developed, which could produce a daily spatial SWE product. This tool could be used in real time SWE analysis, if real time SWE measurements were available.
- The calculated RMSE values were quite reasonable compared to other studies (Fassnacht et al. 2003), and implied that the accuracy of the product was quite adequate around the stations.
- The prediction of the SWE around Great Lakes tended to have relatively higher accuracy due to the availability of a large number of point SWE measurements in the area.
- Snow course data were incorporated into the SWE prediction and the CMC snow depth adjustment for the first time.
- Universal Kriging (UK) was successfully applied for data assimilation and spatial prediction. UK is a relatively new approach in data assimilation studies.

- It was found that vegetation mass and distance from ocean had a significant relationship with the SWE.

The limitations of the current research are as follows:

- As there were very few SWE measurements in northern Quebec and Labrador, the prediction was more difficult in these regions. The details summary of data collection are provided in appendix A. In most simulation days, one needed to rely on deterministic models in this region, rather than using statistical models. The RMSE values were slightly higher in this regions and can be more than 100 mm of the SWE.
- On some days, the SWE suddenly disappeared and the next day it appeared again. This sudden increase and decrease in SWE is a major concern of the current research.
- Prediction of SWE became more difficult in complex topography. That's why during validation, prediction of the SWE in Lac Castor, Quebec was less accurate.
- The snow climate class model is derived from primary data collection. As such, if the precipitation rate of snowfall changes significantly in future, the results of this model may not be as accurate.
- The RMSE value increased during the melting period of snow. The SWE prediction at the melting period needed to be more accurate.

7.9 Discussion

A gridded daily SWE prediction tool was developed from the current research. A simulation was also run for the period of 2008-2010 to generate the SWE analysis. The outputs from the research will help water resources engineer, hydrologists, water managers, hydrological modelers, and climate change researchers. Environment Canada is interested in this research, as the gridded SWE product could be used for numerical weather prediction. The results and the achievements and drawbacks of the study are already discussed at the above sections. The goal of the research was attained for a coarser resolution purpose, as the results generated from the SWE prediction tool seemed accurate. However, for finer scale analysis, more accurate SWE prediction may be necessary.

In the current study, the SWE analysis was developed on the basis of the CMC product by incorporating snow course data sets and physiographical variables. The current research did not develop a new data product based on primary data collection or by establishing a forecast model. It just used the CMC snow depth and incorporated this in a prediction tool developed from the current research. Mainly an approach was developed which could predict SWE from the CMC snow depth. The accuracy of the SWE analysis is dependent on the of the CMC analysis. Without the CMC snow depth product, it will be very difficult to develop the SWE analysis. So, the methodology developed from the current research is dependent on the availability of gridded snow depth.

Snow course data is an important source of information, which has not been used previously in the CMC snow depth or SWE prediction. A snow course database was built as a part of current research and was used in the SWE prediction. Continuous updating of this snow course database will be very helpful for future analysis.

It would be difficult to validate and determine the actual accuracy of the developed SWE analysis product, as there is no independent database. Validation and Cross-validation was conducted on a small number of points, and mostly the accuracy of the product was determined from the visual consistency check.

UK was applied here to convert point snow measurements into a spatial consistent database. UK is a very robust method (Hengl et al., 2007) and it is expected that it can be used widely for data assimilation.

From the results of the daily gridded SWE analysis, it seems that the research showed significant progress in SWE prediction. The limitations associated with the current research may be overcome by future studies. Some recommendations about the future studies are provided in the concluding chapter.

Chapter 8 Conclusion and Recommendation

8.1 Conclusion

In this chapter, a summary of the current research and the findings are discussed.

A methodology was developed as part of the current research to predict Snow Water Equivalent (SWE) from the CMC snow depth analysis product (Brasnett, 1999), physiographical variables, and by incorporating snow course measurements. In Canada, there is no Snow Water Equivalent (SWE) data assimilation system and this research aimed to fill this gap. It is also important that snow course measurements were incorporated for the first time in the daily SWE analysis. In the established methodology, a combination of deterministic snow density prediction models and developed statistical models are used to generate the final analysis. Universal Kriging (UK) a relatively new approach in spatial prediction was applied here to convert the point measurements into a spatially consistent data set. The methodology developed here could be used, with modification, for other future spatial analysis.

A SWE prediction tool was developed from the current research, which could automatically generate daily spatial SWE analysis within the transect of the Great Lakes through Quebec and into Labrador. The outputs from the prediction tool were the SWE analysis product, the adjusted CMC snow depth product, and the corresponding snow density. All these outputs were generated in a 24-hr format and using the same grid locations of the CMC snow depth. The grid resolution was 24 km. The SWE analysis, adjusted CMC snow depth, and corresponding snow density maps were also produced using the prediction tool. For each day within the simulation period, the RMSE values were also generated from validation and cross-validation points.

The SWE analysis outputs were generated for the period of 2008-2010. The generated outputs were used to validate the prediction tool, and to determine the accuracy of the analysis. The significance of the generated results were discussed, and the limitations of the current research were also identified in Chapter 7. From the prediction, it seemed that the results were accurate at a coarser scale. At single station validation points, one found quite accurate results at Churchill Falls, Labrador and Typol, Ontario. The results at Lac Castor, Quebec were less accurate. The generated RMSE values were quite reasonable, but they only denoted accuracy around the snow course points. The calculated mean from the SWE analysis product and CMC monthly average SWE values (Brown et al., 2003) were similar to the selected station points, but varied spatially. The SWE analysis product was also checked with the snow cover dataset. Though there were some limitations, but the spatial and temporal variation of the SWE seemed reasonable from visual consistency checks though the tool had a tendency to underestimate the SWE.

The prediction of SWE was relatively accurate around the Great Lakes area, however results were less reliable in Quebec. The reason behind this is primarily due to the complex topography of Quebec.

The SWE evolved quite naturally in both the spatial and temporal scales. However, sometimes the SWE suddenly appeared and disappeared. These sudden increases and decreases in the SWE are a major concern in the current research.

In the analysis, the RMSE values increased during the melting period, due to the associated complexity of the snow pack at the time of snowmelt.

A snow course database from the Great Lakes to Ontario was constructed as part of the current research, as an extension of Ross Brown`s Snow CD (MSC, 2000) until 2011and was used in the prediction of the SWE. The snow course data were collected from a variety of agencies and organizations around Ontario, Labrador and Quebec. It needs to be noted that snow course data for upper Quebec was not incorporated in the current research due to availability of the data. These data are available at Quebec-Hydro, but within the research period it was not possible to collect this data-set. It is expected that incorporating this data set would increase the prediction accuracy in Quebec regions.

The developed SWE analysis can be very helpful in water resources modelling, Numerical Weather Prediction (NWP), hydrological modelling, climate change studies, and water management. The SWE analysis could be used for SWE trend analysis and to determine water availability for hydro-electric development.

8.2 Recommendations

The current research aimed to develop an approach to generate a daily spatial SWE analysis product. It is the first step of developing a snow data assimilation system for the whole of Canada. In the current study, a methodology was developed and by applying the methodology, a gridded daily SWE product was generated for the period of 2008-2010. Environment Canada was interested in the project, as the gridded SWE product could be very much helpful in water resources modelling especially in hydrological modelling. From the findings of the research, it appeared that the results were accurate enough to be useful, but still have some limitations at the finer scale. These limitations could be improved by future studies. Some recommendations are mentioned in this chapter, which may guide future researchers. The recommendations below may also help Environment Canada, develop gridded SWE products for further periods and in real time. The recommendations for future improvements are as follows.

- The prediction can be improved especially in the Quebec region by incorporating data from Quebec-Hydro.
- The sudden increase and decrease in the SWE may be due to the use of different methods and the best selected method can be different from day to day. This problem can be solved by introducing variables such as temperature. Based on temperature, one can fix the threshold value of maximum snow melting within 24 hours and thus reduce the sudden snow melting problem.
- In future studies, more focus is necessary regarding the SWE prediction in complex topography. Details regarding land cover type and more accurate forest density would be useful in this endeavour. Different methodologies and regression equations for different land cover types can also be developed for more accurate SWE prediction.
- More snow course sites can be established for the SWE measurements and these primary data would be a very good step for future SWE prediction. Representative snow course sites of different topographies can be established and data collected from the newly established snow course sites to improve the SWE prediction.

- There are also some approaches that can be used in the near future to improve the prediction tool. One could incorporate snow density predictions from other deterministic models which have not been used here, especially complex snowpack models output. One can also incorporate historical climatologically prediction in the prediction tool. If the prediction tool frequently underestimates the SWE then a multiplication factor to increase the predicted SWE could be applied.
- The current research has compared five spatial interpolation techniques to choose the best one. Theoretically, the UK is sound and suitable in the current work, but other methods, such as statistical interpolation, or the kalman filter, still could be explored.

By following the recommendations, the SWE prediction can be improved in finer scale. The actual success of the current research mainly depends upon its contribution in the field of hydrological and water resources modelling. The developed SWE product need to be integrated in hydrological modelling to find out the improvement in hydrological prediction. An improvement in flood forecasting or water availability prediction can suggest a great success of the current research.

Bibliography

Arsenault, R. K. "Snow Hydrology: Guest Lecture." 2010.Web.

Barrett, A. National Operational Hydrologic Remote Sensing Center Snow Data Assimilation System (SNODAS) Products at NSIDC. NSIDC Special Report 11. Vol. . Boulder, CO ,USA: National Snow and Ice Data Center, 2003. Print.

Bliss, N. B., L. M. Olsen. "Development of a 30-Arc-Second Digital Elevation Model of South America". Thirteen, Human Interactions with the Environment - Perspectives from Space, Sioux Falls, South Dakota, USA 1996. Print.

Bohling, G. "INTRODUCTION TO GEOSTATISTICS AND VARIOGRAM ANALYSIS." 17 October 2005. Web. 02 January 2012 <http://people.ku.edu/~gbohling/cpe940>.

---. "KRIGING." 19 October 2005. Web. 02 January 2012 <http://people.ku.edu/~gbohling/cpe940>.

Brasnett, B. "A Global Analysis of Snow Depth for Numerical Weather Prediction." J. Appl. Meteorol. 38 (1999): 726-740. Print.

Brown, D. R., and P. Mote. "The Response of Northern Hemisphere Snow Cover to a Changing Climate." Journal of Climate 22 (2009): 2124-2145. Print.

Brown, D. R., Walker A., and Goodison B. E. "Seasonal Snow Cover Monitoring in Canada-an Assessment of Canadian Contributions for Global Climate Monitoring". 57th Eastern Snow Conference, Syracuse, New York, USA, 17-19 May, 2000. 131-141. Print.

Brown, D. R. "Analysis of Snow Cover Variability and Change in Quebec 1948-2005." J. Hydrological Processes 24 (2010): 1929-1954. Print.

---. "Northern Hemisphere Snow Cover Variability and Change, 1915-1997." Journal of Climate 13 (2000): 2339-2355. Print.

Brown, D., R.; BRASNETT, B.; ROBINSON, D. " Gridded North American Monthly Snow Depth and Snow Water Equivalent for GCM Evaluation." J. Atmosphere-Ocean 41 (2003): 1-14. Print.

---. "Development of a Gridded North American Daily Snow Depth and Snow Water Equivalent Dataset for GCM Validation". 58th Eastern Snow Conference, Ottawa, Ontario, Canada. 14-17 May, 2001. 333–340. Print.

Brown, R., et al. " Estimation of Snow Cover in CLASS for SnowMIP." J. Atmosphere-Ocean 44 (2006): 223-238. Print.

Brown, Ross D., and Brasnett B. Canadian Meteorological Centre (CMC) Daily Snow Depth Analysis Data. Boulder, Colorado, USA: National Snow and Ice Data Center, 2010. Print.

Cressman, G. " An Operational Objective Analysis System." Mon. Weather Rev 87 (1959): 367-374. Print.

Daley, R. Atmospheric Data Analysis. Cambridge University Press, 1991. Print.

Dingman, L. S. "Snow and Snowmelt." Physical Hydrology. New Jersey, USA: Prentice Hall, 2002. 167-219. Print.

Elder, K., W. Rosenthal, and R. E. Davis. " Estimating the Spatial Distribution of Snow Water Equivalence in a Montane Watershed." J. Hydrological Processes 12 (1998): 1793-1808. Print.

Fassnacht, S. R., K. A. Dressler, and R. C. Bales. " Snow Water Equivalent Interpolation for the Colorado River Basin from Snow Telemetry (SNOTEL) Data." J. water resources research 39 (2003): 1-10. Print.

Finkelstein, P. L. "The spatial analysis of acid precipitation data." J. Clim. Appl. Meterolo. 23 (1984): 52-62. Print.

Foster, J. L., D. K. Hall, and A. T. C. Chang. " Remote Sensing of Snow." EOS 11 Aug 1987: 681-684. Print.

Garstka, W. U. "Snow and Snow Survey." Handbook of Applied Hydrology. New York: NY: McGraw-Hill, 1964. Print.

Gelman, A., et al. Bayesian Data Analysis. 2 nd ed. CRC Press, 2004. Print.

Brusilovskiy, E. "Spatial Interpolation: A Brief Introduction." 2009.Web. 10 October 2012 <http://www.bisolutions.us/A-Brief-Introduction-to-Spatial-Interpolation.php>.

Bussieres, N., and W. Hogg. " The Objective Analysis of Daily Rainfall by Distance Weighting Schemes on a Mesoscale Grid." J. Atmosphere-Ocean 27 (1989): 521-541. Print.

Carroll, S. S. "Modeling Measurement Errors when Estimating Snow-Water Equivalent." J. Hydrol. 172 (1995): 247-260. Print.

Carroll, S. S., T. R. Carroll, and R. W. Poston. " Spatial Modeling and Prediction of Snow-Water Equivalent using Ground-Based, Airborne, and Satellite Snow Data." J. Geophys. Res. 104 (1997): 19,623– 19,629. Print.

Carroll, S. S., and N. Cressie. " A Comparison of Geostatistical Methodologies used to Estimate Snow-Water Equivalent." Water Resour. Bull. 32 (1996): 267–278. Print.

---. "Spatial Modeling of Snow-Water Equivalent using Covariances Estimated from Spatial and Geomorphic Attributes." J. Hydrol. 190 (1997): 42-59. Print.

Carroll, T., et al. " NOHRSC Operations and the Simulation of Snow Cover Properties for the Conterminous." 69th Annual Meeting of the Western Snow Conference (2001): 1-14. Print.

Chow, V. T. Handbook of Applied Hydrology. New York, USA: NY: McGraw-Hill, 1964. Print.

Gold, L. W. "Changes in a Shallow Snow Cover Subject to a Temperate Climate." J. Glaciol. 3 (1958): 218-222. Print.

Goodchild, M. F., and K. K. Kemp. "NCGIA Core Curriculum in GIS." 1990.Web. <http://www.ncgia.ucsb.edu/giscc/>.

Hengal, T., G. B. M. Heuvelink, and G. D. Rossiter. "About Regression-Kriging: From Equations to Case Studies." J. Computers and Geosciences 33 (2007): 1301-1355. Print.

Hengal, T., G. Heuvelink, and A. Stein. "A Generic Framework for Spatial Prediction of Soil Variables Based on Regression-Kriging." Geoderma 120 (2004): 75-93. Print.

Hiemstra, P. H., et al. "Real-Time Automatic Interpolation of Ambient Gamma Dose Rates from the Dutch Radioactivity Monitoring Network." J. Computers and Geosciences (2008)Print.

Isaaks, H. E., and M. R. Srivastava. Applied Geostatistics. New York: Oxford University Press, 1989. Print.

Longley, R. W. "Snow Depth and Snow Density at Resolute, Northwest Territories." J. Glaciol. 3 (1960): 733-738. Print.

Magono, C., and W. C. Lee. "Metrological Classification of Natural Snow Crystals." J. Faculty Sci., Hokkadio Uni. 2 (1996): 321-335. Print.

Mahfouf, J-F, B. Brasnett, and S. Gagon. "A Canadian Precipitation Analysis (CaPA) Project: Description and Preliminary Results." J. Atmosphere-Ocean 45 (2007): 1-17. Print.

Matheron, G. "Part 1 of Cahiers Du Centre De Morphologie Mathématique De Fontainebleau." Le krigeage universel. École nationale supérieure des mines de Paris (1969) Print.

McKay, G. A. Handbook on the Principle of Hydrology. Precipitation. In D. M. Gray, ed. ed. Port Washington, NY: Water Information Center, Inc., 1970. Print.

Mellor, M. Snow and Ice at the Earth's Surface. Hanover, NH:U.S. Army Cold Regions Research and Engineering Laboratory: Cold Regions Science and Engineering Monograph II-C1, 1964.

MSC. Canadian Snow Data CD-ROM. CRYSIS Project. Downsview, Ontario, Canada: Climate Processes and Earth Observation Division, Meteorological Service of Canada, 2000. Print.

NOAA/NESDIS/OSDPD/SSD. IMS Daily Northern Hemisphere Snow and Ice Analysis at 4 km and 24 km Resolution. Boulder, Colorado, USA: National Snow and Ice Data Center., 2004. Print.

Pebesma, E. J. "Multivariable Geostatistics in S: The Gstat Package." J. Computers and Geosciences 30 (2004): 683-691. Print.

Ramsay, B. H. "The Interactive Multisensor Snow and Ice Mapping System." J. Hydrological Processes 12 (1998): 1537-1546. Print.

Rodell, M., et al. "The Global Land Data Assimilation System." Bull. Amer. Meteor. Soc 85.3 (2004): 381-394. Print.

Solomon, I. S., and P. J. Denouvilliez. " the use of a Square Grid System for Computer Estimation of Precipitation, Temperature, and Runoff." J. Water Resources Research 4 (1968): 919-929. Print.

Sommerfeld, R. A. Classification Outline for Snow on the Ground. Research paper RM-48 Vol. Fort Collins, CO, USA: U.S. Forest Service-Rocky Mountain Forest and Range Experiment Station, 1969. Print.

Spiegelhalter, D. J., et al. "Bayesian Measures of Model Complexity and Fit." J. Stat. Soc. 64B (2002): 583-639. Print.

Sturm, M., J. Holmgren, and G. Liston. "A Seasonal Snow Cover Classification System for Local to Global Applications." J. Climate 8 (1995): 1261-1283. Print.

Sturm, M., et al. " Estimating snow water equivalent using snow depth data and climate classes." J. of Hydrometeorology 11 (2010): 1380-1394. Print.

Tabler, R. D., et al. "Estimating Snow Transport from Wind Speed Records: Estimates Versus Measurements at Prudhoe Bay, Alaska".17-19 April, Sacramento, CA., USA. 58th Western Snow Conf. , 1990. 61-78. Print.

UNESCO/IASH/WMO. Seasonal Snow Cove. Paris, French.: UNESCO/Int. Assoc. Sci. Hydrol./World Meteorological Organization, 1970. Print.

Verseghy, D. L., N. A. McFarlane, and M. Lazare. "A Canadian Land Surface Scheme for GCMS, II. Vegetation Model and Coupled Runs." Int. J. Climatol 13.4 (1993): 347-370. Print.

Wessel, P., and W. H. F. Smith. " A Global Self-Consistent, Hierarchical, High-Resolution Shoreline Database." ." J. Geophys. Res. 101 (1996): 8741-8743. Print.

Appendix-A Snow Course Data

Snow course data are collected from different sources of Canada. Snow course data mainly means snow depth and snow water equivalent measured by standard snow tube survey method. In the current research, Ross Brown snow course data (MSC, 2000) is extended for a period of 2011, within the research are. Our research area starts from the Great lakes to Labrador through Quebec.

a. Snow Course Database (MSC, 2000)

The main snow course database was created by Ross Brown (MSC,2000). It contains both snow depth and snow water equivalent data. The database was created by collecting data from different agencies and some snow water equivalent were estimated. The database has data throughout Canada for the period of 1935 to 2004. This is the best collection of data so far. For Quality Control of this data set Flag was also provided.

The summary of the dataset (MSC, 2000) is given below

The database consists of the following file types:

- Observed SWE Data
  - Observed SWE Files - Snow depth in cm and SWE in mm, collected from six agencies across Canada.
  - Station Cataloger File - Station information including latitude, longitude, name, elevation for all stations with observed SWE data.
- Estimated SWE Data
  - Estimated SWE Files - Biweekly snow depth measurements and SWE estimated at stations from the Canadian Daily Snow Depth Database.

SWE estimates were based on snow depth and interpolated snow densities from the Observed SWE Dataset.

- Station Cataloger File - Station information including latitude, longitude, name, elevation for all stations with observed SWE data.

- Gridded Snow Density Normals Files - Gridded snow density normals were computed from the Observed SWE Dataset. The frequency of these data were bi-weekly.
- FORTRAN Files - Sample FORTRAN extraction software and sample FORTRAN code for reading data files.

The database was constructed by collecting data from AES snow cover data books, British Columbia Environment, Ontario Ministry of Natural Resources, Atmospheric Environment Service, Environment New Brunswick, Alberta Environment, Indian and Northern Affairs Canada (MSC, 2000).

The Contact address regarding this data set is

Climate Processes and Earth Observation Division

Atmospheric Environment Service

Environment Canada

2121 Route Trans-Canadienne

Dorval, Quebec, H9P 1J3

CANADA

b. Labrador Precipitation data

Churchill Falls precipitation data was collected from NALCOR Energy for Labrador region. The data set contains snow depth and snow water equivalent. The data was collected from field by using a precipitation can at each site. The data was collected for four months January, February, March, April. The data is from 1984 to 2010.

Contact Information

GHynes@nalcorenergy.com

c. Ontario Provence Data

Ontario Province Data was collected from Ontario Power Generation (OPG) authority and Ontario Ministry.

The contact information for OPG is

http://www.opg.com/safety/water/snow_survey.asp

For more information on water use please contact:

Margaret McMahon

(905) 357-0322 ext.2911

margaret.mcmahon@opg.com

The contact information for Ontario Ministry is

Gordon Gallant

Water Level Management Specialist

Surface Water Monitoring Centre

Integration Branch, Regional Operations

Ministry of Natural Resources

300 Water St, 5th Floor South Tower

Peterborough, ON

K9J 8M5

(705)755 5200 voice

(705)755 5038 fax

(705)761 3700 cell

gordon.gallant@ontario.ca

d. Quebec Data

Quebec province data is collected from Quebec Ministry. We have tried to collect data for upper Quebec from Quebec Hydro, but unable to collect within the research period.

The contact information is

Pierre-Yves St-Louis

Info-Climat

Service de l'information sur le milieu atmosphérique (SIMAT)

Direction du suivi de l'état de l'environnement (DSEE)

Ministère du Développement durable, de l'Environnement et des Parcs

675, boul. René-Lévesque Est, 7e étage

Québec (Québec)   G1R 5V7

Appendix-B Regression

In this appendix, details regression equations and assumptions related with the regression for 14 test days are provided. Details about regression on CMC snow depth adjustment and SWE prediction on 14 test days are provided. From these regression equations and assumptions regarding residuals, one can easily identify predictors which has significant relationship with SWE, and the results obtained from the regression is very helpful to develop Universal Kriging model in future. In some cases, the assumptions are not satisfied quite well, these can happen in real life data. Different transformations have also tried to satisfy the assumptions.

CMC snow depth adjustment

01-03-2010

General Regression Analysis: sd+1 versus y, CMC_sd, slope, lslope, bar_NE, ...

```
Box-Cox transformation of the response with rounded lambda = 0.384185
The 95% CI for lambda is (0.325, 0.435)

Regression Equation

sd+1^0.384185  =  -16.0551 + 0.0731812 y + 0.0304689 CMC_sd - 0.219557 slope +
                   0.312011 lslope + 0.0350206 bar_NE + 0.00118094 bar_NW +
                   0.000382895 bar_SE + 0.0372048 elev + 8.40639e-005 vm

648 cases used, 1 cases contain missing values

Coefficients
```

Term	Coef	SE Coef	T	P
Constant	-16.0551	8.70370	-1.8446	0.066
y	0.0732	0.01919	3.8144	0.000
CMC_sd	0.0305	0.00257	11.8747	0.000
slope	-0.2196	0.05895	-3.7247	0.000
lslope	0.3120	0.09999	3.1204	0.002
bar_NE	0.0350	0.00722	4.8522	0.000
bar_NW	0.0012	0.00051	2.3102	0.021
bar_SE	0.0004	0.00017	2.2169	0.027
elev	0.0372	0.00716	5.1941	0.000
vm	0.0001	0.00002	5.5358	0.000

```
Summary of Model
```

S = 0.796258 R-Sq = 55.67% R-Sq(adj) = 55.05%
PRESS = 415.782 R-Sq(pred) = 54.44%

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	9	508.006	508.006	56.4451	89.026	0.0000000
y	1	60.253	9.225	9.2248	14.550	0.0001498
CMC_sd	1	372.174	89.404	89.4038	141.009	0.0000000
slope	1	1.037	8.796	8.7963	13.874	0.0002128
lslope	1	5.486	6.173	6.1734	9.737	0.0018877
bar_NE	1	30.669	14.928	14.9276	23.544	0.0000015
bar_NW	1	1.334	3.384	3.3838	5.337	0.0211957
bar_SE	1	0.000	3.116	3.1159	4.915	0.0269836
elev	1	17.624	17.105	17.1051	26.979	0.0000003
vm	1	19.430	19.430	19.4296	30.645	0.0000000
Error	638	404.510	404.510	0.6340		
Lack-of-Fit	533	404.279	404.279	0.7585	345.870	0.0000000
Pure Error	105	0.230	0.230	0.0022		
Total	647	912.515				

Fits and Diagnostics for Unusual Observations for Transformed Response

Obs	sd+1^0.384185	Fit	SE Fit	Residual	St Resid		
1	1.00000	2.79537	0.085551	-1.79537	-2.26788	R	
6	1.00000	3.86154	0.082234	-2.86154	-3.61306	R	
9	1.00000	3.86154	0.082234	-2.86154	-3.61306	R	
25	1.17095	2.80874	0.109316	-1.63779	-2.07652	R	
44	2.52637	4.31281	0.070892	-1.78644	-2.25249	R	
45	2.41179	4.30802	0.071305	-1.89623	-2.39104	R	
51	7.41309	4.39064	0.067829	3.02245	3.80966	R	
76	7.41309	3.78820	0.071300	3.62489	4.57076	R	
78	7.41309	3.78820	0.071300	3.62489	4.57076	R	
81	4.43201	3.75530	0.176708	0.67671	0.87160		X
101	7.12905	5.34897	0.098398	1.78008	2.25282	R	
108	5.98043	4.38438	0.068257	1.59605	2.01184	R	
109	5.44285	3.27964	0.097711	2.16321	2.73741	R	
113	5.98043	5.98043	0.398129	-0.00000	-0.00001		X
114	5.98043	5.98042	0.398129	0.00000	0.00000		X
131	4.55637	4.18584	0.171455	0.37053	0.47652		X
134	1.85581	4.06908	0.066774	-2.21327	-2.78941	R	
175	6.08983	4.38511	0.068205	1.70472	2.14881	R	
180	5.29797	3.26568	0.070278	2.03229	2.56230	R	
216	5.98043	5.98043	0.398129	-0.00000	-0.00001		X
217	5.98043	5.98042	0.398129	0.00000	0.00000		X
284	1.70334	3.50972	0.064477	-1.80638	-2.27606	R	
286	1.70334	3.30735	0.086751	-1.60401	-2.02649	R	
288	1.70334	3.75290	0.052535	-2.04955	-2.57960	R	
290	2.91802	3.64805	0.243517	-0.73003	-0.96297		X
295	2.91802	3.64805	0.243517	-0.73003	-0.96297		X
300	5.45604	3.52362	0.072680	1.93242	2.43705	R	
312	5.86772	4.18477	0.171419	1.68295	2.16433	R	X
313	5.86772	4.18493	0.171425	1.68279	2.16412	R	X
322	5.86772	4.18477	0.171419	1.68295	2.16433	R	X
323	5.86772	4.18493	0.171425	1.68279	2.16412	R	X
351	1.85581	3.99977	0.102202	-2.14396	-2.71500	R	
356	1.52512	3.17282	0.094309	-1.64770	-2.08397	R	
370	6.27789	6.64736	0.212892	-0.36947	-0.48154		X
376	7.26226	6.35843	0.173742	0.90383	1.16312		X
392	3.23966	3.50403	0.352190	-0.26437	-0.37019		X
395	5.92450	4.13993	0.075076	1.78458	2.25123	R	
406	5.92450	3.84540	0.077014	2.07911	2.62339	R	
407	5.92450	3.84540	0.077014	2.07911	2.62339	R	
439	7.10093	7.13717	0.172408	-0.03623	-0.04661		X
456	3.23966	3.50403	0.352190	-0.26437	-0.37019		X
459	5.56958	3.68934	0.059684	1.88024	2.36801	R	

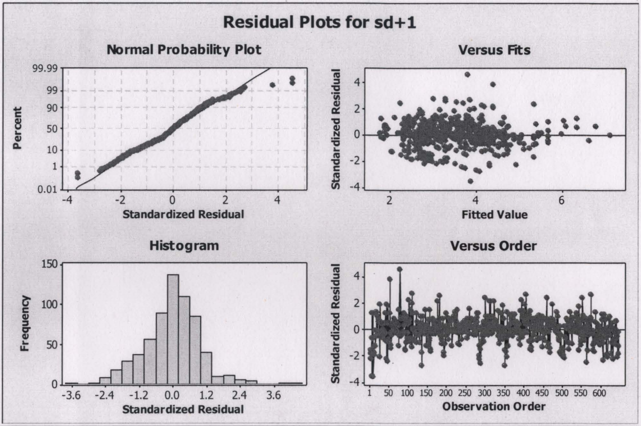
505	1.30512	2.96270	0.083859	-1.65757	-2.09335	R	
507	1.52512	3.22489	0.074416	-1.69977	-2.14408	R	
532	4.86697	4.87909	0.174640	-0.01212	-0.01561		X
553	4.08969	2.21686	0.094060	1.87283	2.36862	R	
565	1.70334	3.65626	0.078859	-1.95291	-2.46473	R	
576	5.09261	2.99472	0.137523	2.09789	2.67488	R	
606	3.08682	4.93392	0.137557	-1.84710	-2.35513	R	
615	1.30512	2.89071	0.079291	-1.58559	-2.00124	R	
619	1.30512	2.97161	0.077911	-1.66648	-2.10299	R	
632	1.30512	3.06344	0.084108	-1.75831	-2.22064	R	
645	1.17095	2.80874	0.109316	-1.63779	-2.07652	R	

Fits for Unusual Observations for Original Response

Obs	sd+1	Fit		
1	1.000	14.523	R	
6	1.000	33.674	R	
9	1.000	33.674	R	
25	1.508	14.704	R	
44	11.160	44.898	R	
45	9.890	44.768	R	
51	183.880	47.038	R	
76	183.880	32.034	R	
78	183.880	32.034	R	
81	48.200	31.315		X
101	166.100	78.637	R	
108	105.140	46.863	R	
109	82.280	22.012	R	
113	105.140	105.140		X
114	105.140	105.140		X
131	51.800	41.538		X
134	5.000	38.589	R	
175	110.220	46.884	R	
180	76.700	21.769	R	
216	105.140	105.140		X
217	105.140	105.140		X
284	4.000	26.260	R	
286	4.000	22.499	R	
288	4.000	31.263	R	
290	16.240	29.040		X
295	16.240	29.040		X
300	82.800	26.532	R	
312	100.060	41.510	R	X
313	100.060	41.515	R	X
322	100.060	41.510	R	X
323	100.060	41.515	R	X
351	5.000	36.902	R	
356	3.000	20.194	R	
370	119.300	138.448		X
376	174.300	123.325		X
392	21.320	26.150		X
395	102.600	40.363	R	
406	102.600	33.308	R	
407	102.600	33.308	R	
439	164.400	166.592		X
456	21.320	26.150		X
459	87.360	29.903	R	
505	2.000	16.895	R	
507	3.000	21.068	R	
532	61.500	61.900		X
553	39.100	7.942	R	
565	4.000	29.210	R	
576	69.200	17.375	R	
606	18.800	63.727	R	
615	2.000	15.847	R	
619	2.000	17.028	R	
632	2.000	18.431	R	

```
645     1.508    14.704  R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.
```

Residual Plots for sd+1



01-12-2008

Regression Analysis: ln(sd+1) versus ln(cmc_sd+1), dist_ocean, ...

```
The regression equation is
ln(sd+1) = 28.7 + 0.414 ln(cmc_sd+1) + 0.000003 dist_ocean + 0.000055 vm
           - 0.00176 bar_NW - 0.000955 bar_SE - 0.218 slope - 0.511 lslope
           + 0.110 x
```

Predictor	Coef	SE Coef	T	P
Constant	28.679	8.662	3.31	0.001
ln(cmc_sd+1)	0.41416	0.08102	5.11	0.000
dist_ocean	0.00000283	0.00000072	3.91	0.000
vm	0.00005463	0.00002326	2.35	0.020
bar_NW	-0.0017598	0.0005921	-2.97	0.003
bar_SE	-0.0009546	0.0004589	-2.08	0.039
slope	-0.21797	0.09201	-2.37	0.019
lslope	-0.5108	0.1605	-3.18	0.002
x	0.10974	0.03421	3.21	0.002

```
S = 0.849183    R-Sq = 46.0%    R-Sq(adj) = 44.0%

PRESS = 169.946    R-Sq(pred) = 40.80%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	132.029	16.504	22.89	0.000
Residual Error	215	155.039	0.721		
Total	223	287.068			

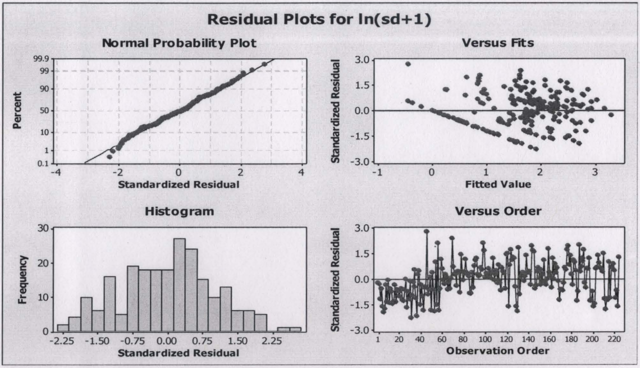
Source	DF	Seq SS
ln(cmc_sd+1)	1	42.735
dist_ocean	1	44.359
vm	1	9.466
bar_NW	1	2.607
bar_SE	1	4.764
slope	1	14.765
lslope	1	5.915
x	1	7.418

Unusual Observations

Obs	ln(cmc_sd+1)	ln(sd+1)	Fit	SE Fit	Residual	St Resid
31	2.86	0.0000	1.9225	0.1374	-1.9225	-2.29R
37	2.27	0.0030	1.8134	0.1898	-1.8104	-2.19R
38	0.00	0.0000	-0.1990	0.3021	0.1990	0.25 X
39	0.00	0.0000	-0.3629	0.3277	0.3629	0.46 X
46	0.00	1.8050	-0.4457	0.2464	2.2507	2.77R
56	0.26	3.3652	1.6213	0.1825	1.7439	2.10R
67	0.92	1.2641	1.7186	0.3528	-0.4545	-0.59 X
69	3.33	3.5553	1.6358	0.2703	1.9195	2.38R
76	0.41	2.2915	1.9906	0.5014	0.3009	0.44 X
98	0.00	2.5602	0.8804	0.2944	1.6798	2.11R
184	0.92	2.0615	1.7079	0.3525	0.3536	0.46 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for ln(sd+1)



02-02-2009

Regression Analysis: sd versus y, x, ...

```
The regression equation is
sd = - 2.1 + 4.23 y + 1.67 x + 0.487 CMC_sd - 5.79 lslope + 0.0161 bar_N
     - 0.0588 bar_NW - 1.79 north
```

Predictor	Coef	SE Coef	T	P
Constant	-2.13	21.07	-0.10	0.920
y	4.2268	0.4741	8.91	0.000
x	1.6707	0.1757	9.51	0.000
CMC_sd	0.48707	0.04614	10.56	0.000
lslope	-5.788	1.451	-3.99	0.000
bar_N	0.016101	0.005859	2.75	0.006
bar_NW	-0.058830	0.006263	-9.39	0.000
north	-1.7855	0.7900	-2.26	0.024

S = 11.5851   R-Sq = 77.9%   R-Sq(adj) = 77.6%

PRESS = 61645.1   R-Sq(pred) = 76.98%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	208646	29807	222.08	0.000
Residual Error	441	59188	134		
Total	448	267834			

Source	DF	Seq SS
y	1	109872
x	1	54519
CMC_sd	1	24691

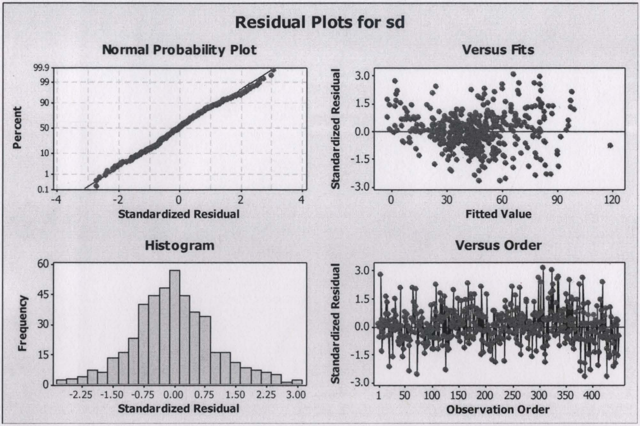
lslope	1	219
bar_N	1	6944
bar_NW	1	11715
north	1	686

Unusual Observations

Obs	y	sd	Fit	SE Fit	Residual	St Resid
5	39.8	33.020	1.675	1.676	31.345	2.73R
20	46.5	22.860	47.769	1.558	-24.909	-2.17R
41	42.3	45.720	22.140	1.812	23.580	2.06R
58	44.5	17.780	45.284	1.693	-27.504	-2.40R
65	47.2	73.660	50.782	2.095	22.878	2.01R
73	44.6	32.000	61.756	1.471	-29.756	-2.59R
108	43.5	106.680	81.476	2.135	25.204	2.21R
126	44.2	57.150	83.958	1.766	-26.808	-2.34R
168	43.3	69.596	43.624	1.351	25.972	2.26R
193	46.9	122.300	98.270	2.432	24.030	2.12R
249	45.2	80.000	55.848	1.379	24.152	2.10R
264	43.4	92.456	69.423	1.569	23.033	2.01R
281	48.6	56.600	70.421	3.102	-13.821	-1.24 X
299	46.0	107.400	77.613	1.303	29.787	2.59R
303	43.2	86.868	60.074	1.209	26.794	2.33R
308	44.5	101.600	66.059	1.445	35.541	3.09R
323	43.4	114.808	80.804	2.153	34.004	2.99R
327	44.5	93.980	68.485	1.459	25.495	2.22R
328	39.1	25.400	22.315	3.185	3.085	0.28 X
335	43.2	89.408	60.223	1.194	29.185	2.53R
337	43.4	99.060	69.090	1.573	29.970	2.61R
356	44.0	52.000	23.980	1.492	28.020	2.44R
376	41.6	27.940	2.927	1.775	25.013	2.18R
377	43.8	50.546	78.133	1.988	-27.587	-2.42R
382	43.4	109.982	80.888	2.152	29.094	2.56R
385	45.4	43.000	65.973	1.876	-22.973	-2.01R
386	44.5	16.000	47.405	0.920	-31.405	-2.72R
396	46.1	60.100	86.831	1.925	-26.731	-2.34R
410	43.2	7.620	31.746	1.901	-24.126	-2.11R
411	43.2	7.620	31.745	1.901	-24.125	-2.11R
418	44.5	77.000	50.310	1.348	26.690	2.32R
427	45.2	28.000	59.030	1.384	-31.030	-2.70R
440	41.9	7.620	32.515	1.741	-24.895	-2.17R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for sd



SWE Prediction

01-12-2008

Regression Analysis: ln(swe+1) versus x, elev, ...

```
The regression equation is
ln(swe+1) = 47.7 + 0.246 x + 0.00341 elev - 0.341 slope + 0.000005 dist_ocean
            + 0.479 ln(sdc+1)


Predictor          Coef      SE Coef      T      P
Constant         47.665        9.978   4.78  0.000
x               0.24634      0.04690   5.25  0.000
elev          0.0034058    0.0007753   4.39  0.000
slope           -0.3411       0.1003  -3.40  0.001
dist_ocean   0.00000456   0.00000094   4.87  0.000
ln(sdc+1)       0.47878      0.09958   4.81  0.000


S = 1.12622   R-Sq = 33.0%   R-Sq(adj) = 31.5%

PRESS = 294.746   R-Sq(pred) = 28.61%


Analysis of Variance

Source           DF        SS      MS      F      P
Regression        5   136.365  27.273  21.50  0.000
Residual Error  218   276.504   1.268
Total           223   412.869
```

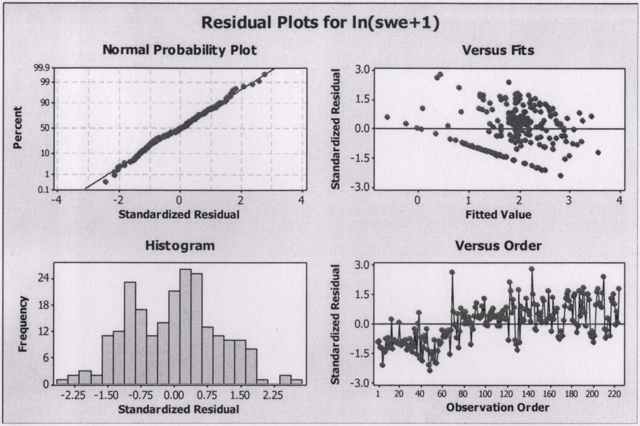
Source	DF	Seq SS
x	1	2.931
elev	1	21.208
slope	1	35.614
dist_ocean	1	47.292
ln(sdc+1)	1	29.319

Unusual Observations

Obs	x	ln(swe+1)	Fit	SE Fit	Residual	St Resid
5	-80.8	0.0000	2.8132	0.1358	-2.8132	-2.52R
31	-80.8	0.0000	2.3023	0.1704	-2.3023	-2.07R
39	-77.8	0.0000	-0.9059	0.3359	0.9059	0.84 X
42	-77.7	0.0000	2.5102	0.1990	-2.5102	-2.26R
47	-79.0	0.0000	0.6204	0.3784	-0.6204	-0.58 X
48	-78.0	0.0000	2.6351	0.2522	-2.6351	-2.40R
49	-85.8	0.0000	2.4135	0.1701	-2.4135	-2.17R
54	-85.2	0.0000	2.6315	0.1142	-2.6315	-2.35R
69	-81.2	3.3673	1.4968	0.3435	1.8704	1.74 X
76	-86.5	2.2116	3.1743	0.3894	-0.9628	-0.91 X
98	-79.3	2.6355	0.9989	0.4196	1.6366	1.57 X
143	-81.3	3.6636	1.5087	0.3317	2.1548	2.00RX
187	-78.8	3.1781	0.9413	0.1667	2.2367	2.01R
195	-87.5	3.9474	2.4624	0.3327	1.4850	1.38 X
196	-87.5	3.9474	2.4615	0.3328	1.4859	1.38 X
210	-81.9	4.6052	2.1655	0.1324	2.4397	2.18R
224	-69.3	3.2734	0.6013	0.2821	2.6720	2.45R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for ln(swe+1)



02-02-2009

Regression Analysis: swe versus sdc, easter, elev

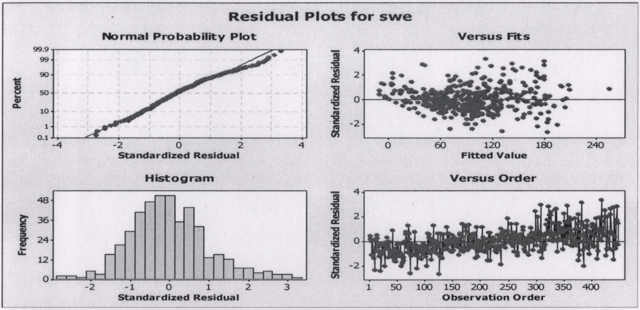
```
The regression equation is
swe = - 11.5 + 2.15 sdc - 4.56 easter + 0.0270 elev


Predictor       Coef  SE Coef      T      P
Constant     -11.467    3.953  -2.90  0.004
sdc          2.14537  0.07167  29.93  0.000
easter        -4.559    1.957  -2.33  0.020
elev         0.02704  0.01175   2.30  0.022


S = 29.6726   R-Sq = 72.3%   R-Sq(adj) = 72.1%

PRESS = 399431   R-Sq(pred) = 71.75%


Analysis of Variance

Source           DF       SS      MS       F      P
Regression        3  1021907  340636  386.88  0.000
Residual Error  445   391805     880
Total           448  1413712


Source  DF   Seq SS
sdc      1  1011934
easter   1     5309
elev     1     4664
```

Unusual Observations

Obs	sdc	swe	Fit	SE Fit	Residual	St Resid
7	15	0.00	40.82	5.28	-40.82	-1.40 X
9	37	0.00	71.43	2.17	-71.43	-2.41R
20	48	25.40	96.90	2.53	-71.50	-2.42R
27	42	30.00	91.28	2.19	-61.28	-2.07R
58	46	27.94	88.89	2.44	-60.95	-2.06R
73	62	53.00	131.78	2.36	-78.78	-2.66R
86	79	106.00	172.47	3.31	-66.47	-2.25R
103	91	134.00	195.48	3.51	-61.48	-2.09R
126	84	104.14	184.12	4.94	-79.98	-2.73RX
236	84	127.00	183.49	4.94	-56.49	-1.93 X
299	78	236.00	164.68	3.02	71.32	2.42R
303	60	192.02	132.73	2.44	59.29	2.00R
308	66	228.60	142.13	2.05	86.47	2.92R
323	81	262.89	179.83	3.83	83.06	2.82R
325	48	157.48	96.16	2.53	61.32	2.07R
326	48	157.48	96.17	2.53	61.31	2.07R
327	69	215.90	146.43	2.10	69.47	2.35R
328	23	58.42	58.34	6.24	0.08	0.00 X
335	60	208.79	133.05	2.44	75.74	2.56R
337	69	231.65	154.86	3.17	76.79	2.60R
351	80	236.22	170.65	2.90	65.57	2.22R
356	24	124.00	38.94	3.24	85.06	2.88R
376	3	68.58	4.87	3.50	63.71	2.16R
382	81	271.78	180.01	3.83	91.77	3.12R
386	48	40.00	100.98	2.15	-60.98	-2.06R
397	119	280.00	255.95	5.14	24.05	0.82 X
418	51	211.00	112.52	2.50	98.48	3.33R
424	68	218.44	144.37	2.16	74.07	2.50R
434	46	180.00	103.21	2.68	76.79	2.60R
435	47	165.00	100.51	2.87	64.49	2.18R
441	55	200.66	118.93	2.56	81.73	2.76R
446	23	124.21	41.46	2.23	82.75	2.80R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for swe



04-01-2010

General Regression Analysis: swe+1 versus sdC, y, vm

Box-Cox transformation of the response with rounded lambda = 0.382887
The 95% CI for lambda is (0.305, 0.455)

Regression Equation

swe+1^0.382887 = -9.0797 + 0.0415262 sdC + 0.264081 y + 6.4099e-005 vm

Coefficients

Term	Coef	SE Coef	T	P
Constant	-9.07970	1.75889	-5.16219	0.000
sdC	0.04153	0.00695	5.97714	0.000
y	0.26408	0.04186	6.30918	0.000
vm	0.00006	0.00002	2.83107	0.005

Summary of Model

S = 0.980734 R-Sq = 51.36% R-Sq(adj) = 50.89%
PRESS = 310.829 R-Sq(pred) = 49.78%

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	3	317.835	317.835	105.945	110.148	0.000000
sdC	1	275.880	34.363	34.363	35.726	0.000000
y	1	34.245	38.287	38.287	39.806	0.000000
vm	1	7.709	7.709	7.709	8.015	0.004940
Error	313	301.056	301.056	0.962		
Lack-of-Fit	310	298.861	298.861	0.964	1.318	0.482020
Pure Error	3	2.194	2.194	0.731		
Total	316	618.891				

Fits and Diagnostics for Unusual Observations for Transformed Response

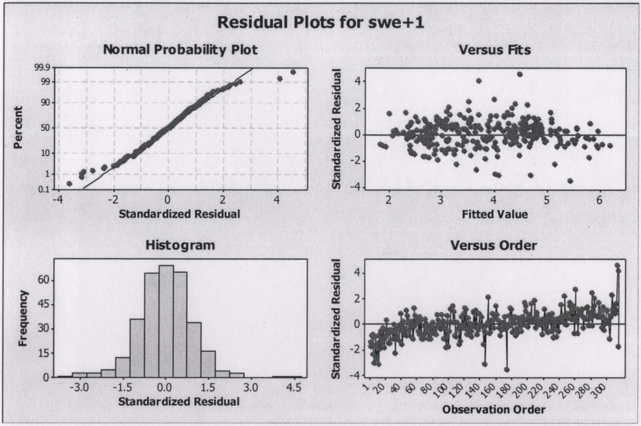
Obs	swe+1^0.382887	Fit	SE Fit	Residual	St Resid		
5	1.00000	3.27404	0.158279	-2.27404	-2.34951	R	
6	1.00000	4.00476	0.095109	-3.00476	-3.07829	R	
9	1.00000	3.66170	0.120544	-2.66170	-2.73472	R	
11	1.00000	4.07940	0.094303	-3.07940	-3.15451	R	
32	2.90748	5.15488	0.092648	-2.24740	-2.30184	R	
147	1.70028	4.69397	0.237452	-2.99369	-3.14611	R	X
150	5.14219	3.13421	0.090113	2.00797	2.05611	R	
160	5.21411	3.92471	0.282833	1.28940	1.37307		X
161	5.21411	3.92384	0.282735	1.29027	1.37395		X
174	1.98584	5.44724	0.197762	-3.46141	-3.60342	R	X
185	5.32091	6.20545	0.213327	-0.88454	-0.92404		X
248	6.75829	4.61631	0.087850	2.14198	2.19287	R	
261	6.55297	3.98894	0.114787	2.56403	2.63249	R	
282	5.88900	3.50753	0.126075	2.38147	2.44856	R	
289	4.54145	4.06009	0.199857	0.48136	0.50133		X
296	4.53309	4.06008	0.199878	0.47300	0.49263		X
308	6.61503	4.39878	0.074740	2.21625	2.26638	R	
315	8.94624	4.48530	0.076585	4.46095	4.56251	R	
316	7.66460	3.69092	0.115855	3.97368	4.08031	R	

Fits for Unusual Observations for Original Response

Obs	swe+1	Fit		
5	1.00	22.144	R	
6	1.00	37.478	R	
9	1.00	29.661	R	
11	1.00	39.329	R	
32	16.24	72.465	R	
147	4.00	56.740	R	X
150	72.00	19.759	R	
160	74.66	35.553		X
161	74.66	35.532		X
174	6.00	83.695	R	X
185	78.72	117.630		X
248	147.00	54.321	R	
261	135.62	37.092	R	
282	102.60	26.510	R	
289	52.05	38.845		X
296	51.80	38.845		X
308	139.00	47.887	R	
315	305.80	50.386	R	
316	204.20	30.284	R	

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Residual Plots for swe+1



Appendix-C SWE Prediction Tool Manual

1. Introduction

SWE_MAP_V1.0 is a spatial prediction tool which can be used to predict SWE in space and time. The tool can predict SWE within a range of 55^0 W to 97^0 W and 38^0 N to 60^0 N. The tool can be used for SWE prediction from January 2008. The tool is a combination of different Fortran codes and R scripts. SWE map and spatial and temporal gridded SWE product can be generated for a maximum period of one year at a time. This spatial modelling tool combines deterministic SWE prediction models with statistical Universal Kriging (UK) models. Comparison of different statistical and deterministic models can be done in this tool and the best models have been picked based on calculated Root Mean Square Error (RMSE) of SWE at validation and cross validation points. Later, the best deterministic and statistical models are combined by cressman data assimilation algorithm for final spatial prediction. CMC snow depth data product, Physiographical variables like vegetation mass, elevation, distance from ocean, and Canadian and USA snow course measurements have been used as input data. These data are processed by Fortran codes and R scripts, in order to be read in the main modelling part. So, the whole SWE_MAP_V1.0 tool can be divided into two parts, one is data processing part and another is SWE prediction part. SWE prediction part is the main modelling part which is written in R scripts. The data processing has mainly done by Fortran codes along with R scripts. The whole tool setup, file structure and detail run procedure will be discussed in this Appendix.

2. Concept and Logical Representation of Prediction Modelling Tool

In this part, only the logical representation behind the SWE prediction is described. In the current research, CMC snow depth analysis product have been used as a main predictor of SWE. But, the CMC product contains some error due to the representative measurements of snow depth from open area. Universal Kriging (UK) models is used to correct the associated bias by using the snow course point measurements which is a independent data set. These UK models is applied only within 150 km of radius of the measured point data. In the rest of the portion, CMC snow depth is used directly without adjustment. After the adjustment process, the SWE prediction can be done under two scenarios. Scenario I is applied to those days where more than 50 Canadian snow course measurements are available. On other days scenario-II is applied. The starting day for the SWE simulation period will always be in Scenario-I. One deterministic model named snow climate class model, developed by Sturm et al. 2009 is used in scenario-I. There are four UK models in this scenario. The best model among these four UK models is selected by calculating RMSE of SWE at validation and cross validation points and then combine with snow climate class model by Cressman algorithm. If snow climate class model predicts better than all the UK models, than the prediction from snow climate class model is used as the final SWE. In scenario-II, there are two deterministic models, one is snow climate class model and another is snow aging model used in Canadian Land Surface Scheme (CLASS) (Verseghy et al, 1993). In scenario II there are two UK models. The best deterministic model and the best UK model is selected based on calculated RMSE of SWE and then combined by Cressman algorithm for final SWE prediction. If any deterministic model provides good prediction than UK models, then the best deterministic model is used for the final prediction of SWE. If there is more than 20% of zero values present in the snow course data, then log transformation of snow depth and SWE is necessary. If there is more than 60% of zero snow course values, then for final prediction deterministic model is used. All this logical representation has been done by the following flow chat.

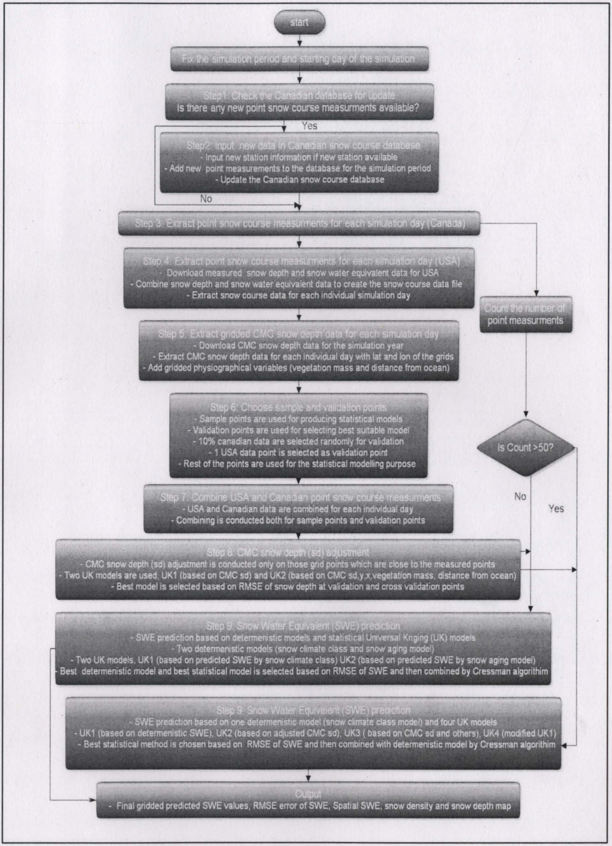


Figure: Flow Chart of the total Process for SWE prediction

3. Installation

This prediction tool is a combination of different Fortran codes and R scripts. So, to run this SWE_MAP_V1.0 tool first we need to install CRAN-R and Fortran compiler. The tool has been developed and setup in an UNIX environment. So, UNIX operating system will be helpful to run the tool and predict SWE in the simulation period. The SWE_MAP_V1.0 tool is in a main folder named SWE_MAP_V1.0, which contains three sub folders; data, model and output. The tool can be placed at any location in the home directory. There is option to input the directory of SWE_MAP_V1.0 in the main R script. The data folder contains preloaded three Fortran codes named mgrid.f90, canada.f90 and usa.f90. There is also an R script in this folder named sample.R. There are some other files which are necessary like Canadian database file final.txt, CMC snow depth analysis lat lon file, basic gridded file with physiographic variables and timeseries.txt. CMC snow depth analysis product file for the simulation period need to be downloaded from the web site (ftp: //sidads.colorado.edu/pub/ DATASETS/nsidc0447_CMC_snow_depth_v01/) and placed in the data folder. The model folder mainly contains R scripts for SWE prediction. There are four R scripts named main_swe.R which is the main script that combines other R scripts and make the run for SWE prediction, UK_sdepth_tm.R is a r script which is used for the snow depth adjustment, swe_inter_tm.R is a r script which is used to predict SWE in days when significant Canadian snow course data are available and swe_gap_tm.R is a r script for constructing SWE at time series gap. In spite of these r scripts, there are some other files necessary like day of year file and snow climate class model's parameter file. Files and maps created from the modelling tool will be placed in this folder later.

Two software are necessary for running these codes. G-Fortran is a good free Fortran complier can be used for running Fortran codes. R is developed by CRAN is a language software with different packages. R version > 2.14.1 is required to run the included scripts in the prediction tool. To install these two software, following codes can be written in the terminal of UNIX

```
$ sudo apt-get update
$ sudo apt-get upgrade
$ sudo apt-get install gfortran
$ sudo apt-get install r-base
```

R studio is a graphical interface of R language and easy to use. R studio can be downloaded from http://rstudio.org/. It is necessary to install different R packages required for the tool. The necessary R packages are gstat, maps, mapproj, fields, mapdata, lattice, maptools, raster, spam, splus2R, RSAGA, date, automap. These packages can be installed by typing install.packages("package name") in R -studio console.

Here, it will be convenient to run the Fortran codes from the terminal by typing gfortran file name (like: gfortran usa.f90) and then (./a.out). Before run the Fortran code from the terminal it is necessary to select the directory , where the Fortran code has already placed. The r script can be run from r studio by the source file.

4. File Structure

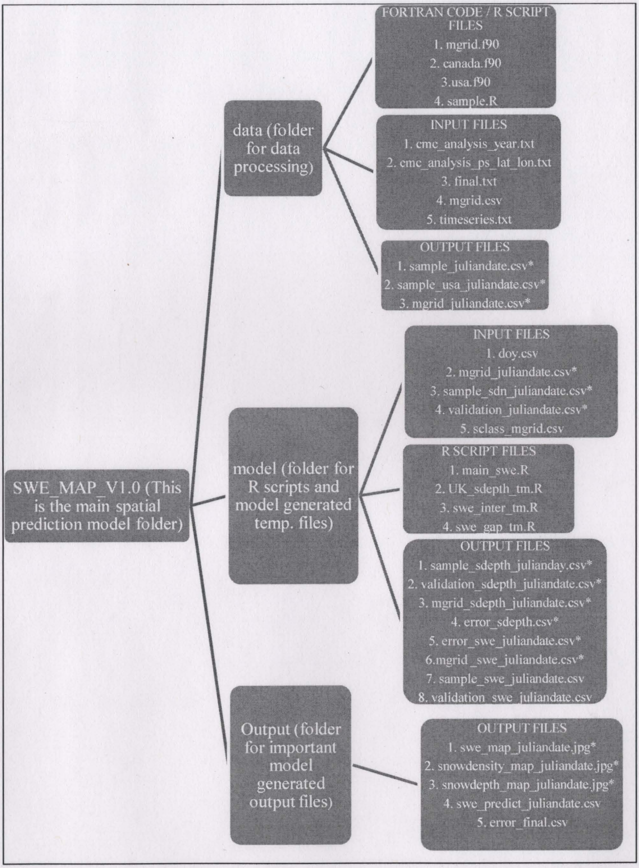

SWE_MAP_V1.0 (This is the main spatial prediction model folder)
data (folder for data processing)
FORTRAN CODE / R SCRIPT FILES
1. mgrid.f90
2. canada.f90
3.usa.f90
4. sample.R
INPUT FILES
1. cmc_analysis_year.txt
2. cmc_analysis_ps_lat_lon.txt
3. final.txt
4. mgrid.csv
5. timeseries.txt
OUTPUT FILES
1. sample_juliandate.csv*
2. sample_usa_juliandate.csv*
3. mgrid_juliandate.csv*
model (folder for R scripts and model generated temp. files)
INPUT FILES
1. doy.csv
2. mgrid_juliandate.csv*
3. sample_sdn_juliandate.csv*
4. validation_juliandate.csv*
5. sclass_mgrid.csv
R SCRIPT FILES
1. main_swe.R
2. UK_sdepth_tm.R
3. swe_inter_tm.R
4. swe_gap_tm.R
OUTPUT FILES
1. sample_sdepth_julianday.csv*
2. validation_sdepth_juliandate.csv*
3. mgrid_sdepth_juliandate.csv*
4. error_sdepth.csv*
5. error_swe_juliandate.csv*
6.mgrid_swe_juliandate.csv*
7. sample_swe_juliandate.csv
8. validation_swe_juliandate.csv
Output (folder for important model generated output files)
OUTPUT FILES
1. swe_map_juliandate.jpg*
2. snowdensity_map_juliandate.jpg*
3. snowdepth_map_juliandate.jpg*
4. swe_predict_juliandate.csv
5. error_final.csv

Figure: File Structure of SWE_MAP_V1.0

In the above figure, total file structure of SWE_MAP_V1.0 is provided. The * sign means this file will be created by running the codes and will not be in the folder at initial set up stage. Year and juliandate are used as variables in the filename . Year will be the simulation year. Julian date is a variable and will be created for each day within the simulation period. So, the filename contain juliandate in the upper structure, is not a single file, but a number of files will be created. For each day, separate file with Julian date will be created. In data folder some intermediate temporary files will be also created but not mentioned in the upper diagram. The details about each files type and run procedure will be described in the next part.

5. Run of SWE_MAP_V1.0

1. Run the Fortran code mgrid.f90 to add time dependent CMC Snow depth data into physiographic gridded data set which is constant in time.

- To run this code three input files are necessary. These three input files are in the data folder.

  a. mgrid.csv- it is the basic grid file contains all the position of CMC snow depth grids with physiographical variables like elevation, vegetation mass, distance from ocean etc. The sample of this file has been provided below

```
y,x,elev,veg,slope,aspect,lslope,easter,north,bar_E,bar_N,bar_NE,bar_NW,bar_SE,vm,dist_water,dist_ocean
38.039631,-96.721809,441.271,7,88.7031238674,134.6023210618,0.6732013145,0.4672369585,-0.8841321307,603.58899,141.45001,624.1789
38.217369,-96.789833,430.50601,11,88.7724558877,128.8606694131,0.7228110872,-0.0553423267,-0.9984674391,634.94397,152.215,634.94
38.395321,-96.858414,423.62399,7,88.7899467452,130.914218233,0.7347868904,-0.8586721473,0.5125252613,641.82593,159.09702,641.825
38.57346,-96.927521,436.14001,10,88.7756761858,138.7980186599,0.7250327085,0.5378981923,0.843009807,629.30994,146.58099,629.3099
38.751808,-96.997177,388.01401,11,88.7610077686,143.049845332,0.7148527216,-0.9942403489,0.1071733581,677.43591,194.707,606.8049
38.092659,-96.495987,430.237,7,88.8034508525,128.1880727615,0.7438794243,0.5787752581,-0.8154870941,614.62299,152.48401,635.2129
38.270691,-96.56321,404.07199,7,88.9496942972,110.8404746919,0.8333273708,-0.7737451214,-0.6334970301,661.37793,178.64902,661.37
38.448929,-96.630959,404.07199,7,88.9913300824,101.9929194428,0.8556139439,0.9940825766,0.1086270264,661.37793,178.64902,661.377
38.627369,-96.699249,434.07101,7,88.8884605492,132.9413006386,0.7979379193,0.8384371208,0.5449983435,631.37891,148.64999,560.747
38.806011,-96.768089,388.01401,11,88.9691986709,167.9128800444,0.8439498083,-0.9868538899,-0.1616149748,677.43591,194.707,606.80
38.984859,-96.837479,388.01401,11,88.8746367863,143.759120159,0.7895298144,-0.6846634911,0.728859317,677.43591,194.707,606.80493
39.16391,-96.907417,370.61401,7,88.791719758,135.0981996007,0.735988366,-0.0097153435,-0.9999528049,624.20496,212.10699,565.4719
39.343151,-96.977943,395.91901,11,88.7272843503,139.2868439406,0.690868795,0.8707704924,0.4916896882,598.89996,186.802,540.16699
38.145012,-96.269653,430.237,7,88.8886980128,124.185110778,0.7980810255,-0.9957525017,0.0920703831,635.21295,133.75403,635.21295
38.32333,-96.336037,381.10901,7,89.2790126417,99.8439500696,0.9673147865,-0.6342320608,0.7731427378,684.34094,182.88202,684.3409
38.50185,-96.402969,401.49701,7,89.1574107401,117.8871462745,0.9294120414,-0.9970044319,0.077344442,663.95294,181.224,663.95294,
38.68058,-96.470444,401.49701,7,89.1273319214,129.139259942,0.9178929121,-0.3277878871,-0.9447513435,663.95294,181.224,593.32196
38.85952,-96.538437,432.01401,7,89.2031409996,193.7643842355,0.9453109741,-0.8491428527,0.5281632472,633.43591,150.707,562.80493
39.03867,-96.606987,363.54999,11,89.0023687621,103.4211679891,0.8612754798,0.2487501007,-0.9685676989,701.89996,219.17102,631.26
39.21801,-96.676086,363.54999,11,88.9182335138,102.5042798367,0.8155269826,0.9200772866,-0.3917368844,631.26898,219.17102,572.53
39.39756,-96.745758,391.064,7,88.8274080458,140.2985306867,0.7596753533,0.8787031587,-0.4773685775,603.755,191.65701,545.02197,4
39.577301,-96.815987,406.13599,11,88.6217219299,143.565030511,0.6108451754,-0.8123855583,0.5831206604,588.68298,176.58502,529.95
39.757252,-96.886803,401.33801,11,88.6584358539,144.4147836829,0.6394952608,-0.0983192856,0.9951549216,391.39197,181.383,534.747
39.937401,-96.958183,401.33801,11,88.6828097841,143.0547897338,0.6580420131,-0.9936982898,0.1120879518,330.42297,181.383,534.747
```

Figure: Preview of mgrid.csv file

b. CMC snow depth analysis data product is a gridded data and can be downloaded from ftp://sidads.colorado.edu/pub/DATASETS/nsidc0447_CMC_snow_depth_v01/ . CMC snow depth data files are yearly and need to be downloaded for the simulation year and to be placed in the data folder. The lat lon file is a constant file and is pre placed in the data folder with the SWE_MAP_V1.0 tool. The file name for lat lon is cmc_analysis_ps_lat_lon.txt and the data file is cmc_analysis_year.txt (like cmc_analysis_2009.txt). Details about this file format can be found in the documentation at the National Snow and Ice Data Center site (NSIDC).

- Open the terminal and change the directory where the mgrid.f90 code is available. Type gfortran mgrid.f90 in the terminal and then ./a.out. On the screen input the starting simulation year ,month and day also input the number of days for which you want to simulate or process the data.
- On screen, the Julian date for the starting simulation date will be shown. Please keep this in record and it will be necessary in future simulation.
- After compiling the program, mgrid_juliandate.csv file will be created for each individual day with Julian date as a variable and it will also represent each day separately. These created output files store gridded CMC snow depth analysis data and physiographical variables. Below a preview of mgrid_2454892.csv is provided.

```
JD,y,x,elev,vm,dist_ocean,CMC_sd
    2454892 ,   38.039631    , -96.721809    ,  441.27100    ,  3649.2866    ,  934548.56   , 2.1
    2454892 ,   38.217369    , -96.789833    ,  430.50601    ,  3551.3218    ,  934996.75   , 2.2
    2454892 ,   38.395321    , -96.858414    ,  423.62399    ,  4107.7725    ,  934622.31   , 2.2
    2454892 ,   38.573460    , -96.927521    ,  436.14001    ,  2681.1626    ,  934533.56   , 2.2
    2454892 ,   38.751808    , -96.997177    ,  388.01401    ,  2787.5237    ,  934962.69   , 2.2
    2454892 ,   38.092659    , -96.495987    ,  430.23700    ,  3840.2571    ,  926766.94   , 2.2
    2454892 ,   38.270691    , -96.563210    ,  404.07199    ,  3896.6360    ,  929680.69   , 2.4
    2454892 ,   38.448929    , -96.630959    ,  404.07199    ,  3979.6985    ,  931596.56   , 2.6
    2454892 ,   38.627369    , -96.699249    ,  434.07101    ,  3866.9514    ,  933106.75   , 2.7
    2454892 ,   38.806011    , -96.768089    ,  388.01401    ,  3181.5874    ,  934595.38   , 2.7
    2454892 ,   38.984859    , -96.837479    ,  388.01401    ,  3215.4233    ,  936087.19   , 2.7
    2454892 ,   39.163910    , -96.907417    ,  370.61401    ,  3329.6587    ,  966864.44   , 2.7
    2454892 ,   39.343151    , -96.977943    ,  395.91901    ,  2984.3333    ,  973633.00   , 2.8
    2454892 ,   38.145012    , -96.269653    ,  430.23700    ,  3376.8025    ,  918630.56   , 2.3
    2454892 ,   38.323330    , -96.336037    ,  381.10901    ,  3577.8345    ,  923711.44   , 2.6
    2454892 ,   38.501850    , -96.402969    ,  401.49701    ,  3946.2913    ,  928374.69   , 2.9
    2454892 ,   38.680580    , -96.470444    ,  401.49701    ,  4089.0737    ,  932033.94   , 3.1
    2454892 ,   38.859520    , -96.538437    ,  432.01401    ,  3788.2136    ,  934932.88   , 3.3
    2454892 ,   39.038670    , -96.606987    ,  363.54999    ,  3825.6238    ,  966077.94   , 3.4
    2454892 ,   39.218010    , -96.676086    ,  363.54999    ,  3876.4263    ,  973946.69   , 3.5
    2454892 ,   39.397560    , -96.745758    ,  391.06400    ,  3535.1167    ,  981174.56   , 3.7
    2454892 ,   39.577301    , -96.815987    ,  406.13599    ,  2363.7312    ,  987545.63   , 3.9
    2454892 ,   39.757252    , -96.886803    ,  401.33801    ,  2120.8557    ,  992941.81   , 4.1
    2454892 ,   39.937401    , -96.958183    ,  401.33801    ,  2028.7667    ,  997345.63   , 4.2
    2454892 ,   38.018291    , -95.977737    ,  346.76999    ,  2785.0540    ,  913136.06   , 2.0
```

Figure: mgrid.csv file

2. Run canada.f90 Fortran code and it will create the Canadian snow course data file for each day within the simulation period

- To run this code, Canadian snow course database file (final.txt) is the only file which is necessary. In below the preview of this database file has been provided. In first column Julian day, then station id, lat, lon, snow depth in cm and at last SWE in mm.
- Open the terminal and change the directory to the directory where the canada.f90 code is available. Type gfortran canada.f90 in the terminal and then type ./a.out. On the screen input the starting simulation year ,month and day and also input the number of days for which you want to simulate or process the data.
- Individual snow course data file with Julian date, lat, lon, snow depth, Snow Water Equivalent (SWE) and snow density will be created as the output file. The quality of the data from the main database has been checked based on density, before stored in each individual Canadian snow course file. Only data of the spatial domain of current research is extracted from the main database final.txt. Below a preview of sample_2454892.csv has been provided. In the created output file snow depth is in cm and swe is in mm.

2427884	356	50.53	-117.28	152.0	389.0
2427885	232	49.37	-122.80	3.0	3.0
2427886	0	0.00	0.00	-999.0	-999.0
2427887	0	0.00	0.00	-999.0	-999.0
2427888	0	0.00	0.00	-999.0	-999.0
2427889	0	0.00	0.00	-999.0	-999.0
2427890	380	49.73	-120.18	91.0	246.0
2427891	0	0.00	0.00	-999.0	-999.0
2427892	381	49.82	-120.02	94.0	259.0
2427893	0	0.00	0.00	-999.0	-999.0
2427894	0	0.00	0.00	-999.0	-999.0
2427895	0	0.00	0.00	-999.0	-999.0
2427896	383	49.99	-118.87	122.0	320.0
2427897	0	0.00	0.00	-999.0	-999.0
2427898	382	49.78	-119.20	79.0	188.0
2427899	0	0.00	0.00	-999.0	-999.0
2427900	0	0.00	0.00	-999.0	-999.0
2427901	0	0.00	0.00	-999.0	-999.0
2427902	0	0.00	0.00	-999.0	-999.0
2427903	0	0.00	0.00	-999.0	-999.0
2427904	0	0.00	0.00	-999.0	-999.0
2427905	0	0.00	0.00	-999.0	-999.0
2427906	0	0.00	0.00	-999.0	-999.0
2427907	0	0.00	0.00	-999.0	-999.0
2427908	0	0.00	0.00	-999.0	-999.0
2427909	0	0.00	0.00	-999.0	-999.0

Figure: Preview of final.txt file

JD,y,x,sd,swe,d

2454906 ,	48.430000	, -81.169998	, 92.000000	, 180.00000	, 0.19565217
2454906 ,	49.779999	, -94.370003	, 45.000000	, 119.00000	, 0.26444444
2454906 ,	50.630001	, -93.180000	, 54.000000	, 84.000000	, 0.15555556
2454906 ,	51.169998	, -90.220001	, 63.000000	, 140.00000	, 0.22222222
2454906 ,	50.119999	, -91.919998	, 59.000000	, 155.00000	, 0.26271185
2454906 ,	48.700001	, -89.620003	, 55.000000	, 107.00000	, 0.19454545
2454906 ,	50.279999	, -89.029999	, 88.000000	, 170.00000	, 0.19318181
2454906 ,	49.080002	, -87.070000	, 56.000000	, 122.00000	, 0.21785714
2454906 ,	49.770000	, -86.919998	, 60.000000	, 96.000000	, 0.16000000
2454906 ,	47.970001	, -81.599998	, 68.000000	, 117.00000	, 0.17205882
2454906 ,	47.480000	, -81.419998	, 63.000000	, 135.00000	, 0.21428572
2454906 ,	46.820000	, -80.949997	, 55.000000	, 173.00000	, 0.31454545
2454906 ,	48.549999	, -80.699997	, 88.000000	, 183.00000	, 0.20795454
2454906 ,	46.630001	, -80.769997	, 73.000000	, 236.00000	, 0.32328767
2454906 ,	47.990002	, -80.699997	, 79.000000	, 163.00000	, 0.20632911
2454906 ,	47.830002	, -80.430000	, 83.000000	, 193.00000	, 0.23253012
2454906 ,	48.520000	, -79.919998	, 78.000000	, 158.00000	, 0.20256411
2454906 ,	46.680000	, -79.989998	, 75.000000	, 155.00000	, 0.20666666
2454906 ,	45.349998	, -80.029999	, 46.000000	, 137.00000	, 0.29782608
2454906 ,	45.919998	, -79.220001	, 55.000000	, 127.00000	, 0.23090909
2454906 ,	46.770000	, -79.029999	, 75.000000	, 196.00000	, 0.26133335
2454906 ,	46.349998	, -78.750000	, 59.000000	, 117.00000	, 0.19830509
2454906 ,	44.950001	, -78.699997	, 48.000000	, 147.00000	, 0.30625001
2454906 ,	45.500000	, -78.220001	, 58.000000	, 180.00000	, 0.31034482
2454906 ,	44.990002	, -77.970001	, 45.000000	, 132.00000	, 0.29333332

Figure: Preview of sample_2454892.csv file

3. Run usa.f90 Fortran code and this code will first download the snow depth and SWE data for the simulation period and then combine these two data set for each individual days and create the USA snow course measurements file.

- To run this code no input data file is necessary as the code downloads the necessary data directly from the web. The code downloads USA data from this web site (http://www.nohrsc.noaa.gov/nsa/). One input file named timeseries.txt is necessary to run the code. This file contains the time series date in chronological order which will need to choose the download address based on date.
- Open the terminal and change the directory to the directory where the usa.f90 code is available. Type gfortran usa.f90 in the terminal and then type ./a.out. On the screen input the starting simulation year ,month and day also input the number of days for which you want to simulate or process the data.
- This code downloads the USA snow depth and SWE data separately, combine them and then check the data quality based on density and store the data separately in file for each individual day. The output file name is sample_usa_juliandate.csv . An example of a file sample_usa_2454892.csv has been provided below.
- USA data from Northeast, Northern Great Lakes , Southern Great Lakes and Alleghany Front has been only downloaded and processed.

- It needs to be keep in mind that the usa.f90 code should not be run for more than 1 month simulation period at a time, otherwise it will take very long time. If the simulation period will be 1 year then 12 run of usa.f90 is necessary to cover the total simulation period. So, it is always necessary to divide the total simulation period in monthly basis.

```
JD,y,x,sd,swe,d
2454892 ,  44.533298   , -72.833298   ,  89.153999      ,  312.42001   , 0.35042736
2454892 ,  43.049999   , -73.033302   ,  59.436001      ,  241.29999   , 0.40598288
2454892 ,  43.933300   , -71.716698   ,  61.467999      ,  175.25999   , 0.28512394
2454892 ,  43.783298   , -72.033302   ,  58.419998      ,  160.02000   , 0.27391306
2454892 ,  43.057251   , -78.861702   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  41.432320   , -81.371597   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  46.493301   , -86.305801   ,  116.84000      ,  289.56000   , 0.24782608
2454892 ,  46.531101   , -87.548302   ,  78.739998      ,  187.95999   , 0.23870967
2454892 ,  46.533329   , -87.550003   ,  78.739998      ,  187.95999   , 0.23870967
2454892 ,  45.071659   , -83.564438   ,  45.720001      ,  106.68000   , 0.23333333
2454892 ,  48.058102   , -92.752197   ,  38.099998      ,  99.059998   , 0.26000002
2454892 ,  44.907501   , -84.718880   ,  30.480000      ,  76.199997   , 0.25000000
2454892 ,  44.479439   , -88.136658   ,  17.780001      ,  27.939999   , 0.15714285
2454892 ,  41.633301   , -87.083298   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  43.242199   , -88.288902   , 3.00000003E-03 ,  0.0000000   ,  0.0000000
2454892 ,  42.833302   , -84.766701   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  41.721390   , -83.732399   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  42.464001   , -83.116203   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  41.408920   , -87.389900   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  41.781429   , -88.067398   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  41.557690   , -87.659401   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  41.331902   , -86.307404   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  43.320068   , -88.168198   ,  0.0000000      ,  0.0000000   ,  0.0000000
2454892 ,  39.066700   , -78.966400   ,  0.0000000      ,  10.160000   ,  0.0000000
```

Figure: sample_usa_2454892.csv file

4. Run r script main_swe.R, this is the main prediction modelling part and it combines four other R scripts within it to produce the final SWE prediction

- main_swe.R is the only r script which needs to be run as the prediction model. Four different r scripts for different purposes are integrated in this main r script. We will discuss this four integrated r scripts separately.

- To run main_swe.R, first open the r studio and then open the main_swe.R file and run the file from the source. It requires to input the starting Julian date which can be obtained during running the mgrid.f90 Fortran code or the starting date can be converted to Julian date by Julian date converter web site (http://www.iasfbo.inaf.it/~mauro/JD/). The prediction tool also requires to input the number of simulation days. We need to also enter the directory where we place SWE_MAP_V1.0 prediction tool. Example- If SWE_MAP_V1.0 has been placed at /home/shinjan/SWE_MAP_V1.0 then we just to enter /home/shinjan/ as the directory.
- The first r script to be run within the main_swe.R is sample.R. The main purpose of running this r script is to chose modelling and validation points randomly for each single day, combine Canadian and USA snow course data for both modelling points and validation points and at last place these snow course data files and daily gridded CMC snow depth file to the model folder. The input file necessary for this script to run is sample_juliandate.csv, sample_usa_juliandate.csv and mgrid_juliandate.csv, and the output files are sample_sdn_juliandate.csv and validation_sdn_juliandate.csv. In below a preview of sample_sdn_2454892.csv and validation_sdn_2454892.csv has been provided.

```
JD,y,x,sd,swe,d,CMC_sd,elev,vm,dist_ocean
2454892,48.630001,-81.400002,96.5,147.3,0.15264249,104.3,305.099,11675.797,294204.91
2454892,50.279999,-89.029999,66,124,0.18787879,46.3,353.30301,10766.453,556770.19
2454892,45.52,-77.900002,53,150,0.28301886,27.3,442.30701,10904.539,554575.88
2454892,46.700001,-73.889999,61.700001,158,0.2560778,32.3,468.91101,10681.572,215498.03
2454892,47.669998,-81.720001,97.300003,185.39999,0.19054469,90.6,393.056,11574.215,398834.31
2454892,47.970001,-81.599998,72,127,0.17638889,97.8,359.548,11100.913,389125.97
2454892,48.549999,-80.699997,93,145,0.15591398,98.8,273.44601,10024.651,297806.75
2454892,47.830002,-80.43,96,178,0.18541667,70.2,316.14899,11828.217,352534.72
2454892,46.380001,-71.650002,76.599998,191,0.24934727,44.8,155.228,10789.63,85515.742
2454892,49.779999,-94.370003,44,109,0.24772727,58.1,355.21301,9489.9736,793413.38
2454892,44.950001,-78.699997,63.5,172.7,0.27196851,37.3,331.29599,11560.695,598880.06
2454892,47.279999,-79.5,93.5,147.3,0.15754011,46.6,236.62601,8727.2197,425563.13
2454892,45.25,-76.75,35.299999,106.7,0.3022663,7.3,167.767,10980.852,459457.69
2454892,51.169998,-90.220001,54.599998,109.2,0.2,54.4,371.367,9569.4199,570899.94
2454892,45.919998,-79.220001,35,79,0.22571428,38.4,429.23599,10869.604,572964.69
2454892,48.330002,-81.32,87,147,0.16896552,106.3,319.30399,10805.814,335528.69
2454892,49.77,-86.919998,46,71,0.15434782,45.2,320.30899,10012.904,465468.94
2454892,45.919998,-79.220001,34.799999,78.699997,0.22614942,38.4,429.23599,10869.604,572964.69
2454892,47.080002,-79.279999,64.5,144.8,0.22449613,51.4,263.68399,10002.105,440399.5
2454892,47.990002,-80.699997,88,152,0.17272727,77.3,342.552,11436.429,375143.28
2454892,47.630001,-69.57,78.199997,212,0.27109975,57.7,173.25999,6729.1372,27503.17
2454892,49.080002,-87.07,53.299999,132.10001,0.24784242,39.1,417.74701,11434.67,520555.47
2454892,46.630001,-81.769997,97.300003,228.60001,0.23494348,68.5,406.461,12212.555,508263.59
2454892,47.470001,-79.830002,94,147,0.15638298,34.4,202.569,9026.1182,401844.94
2454892,46.580002,-74.169998,82.800003,190,0.22946858,38.6,471.57199,10924.043,242701.16
```

Figure: sample_sdn_2454892.csv file

```
JD,y,x,sd,swe,d,CMC_sd,elev,vm,dist_ocean
2454892,48.700001,-89.620003,55,127,0.23090909,52.1,459.96899,10261.796,684286.88
2454892,47.48,-81.419998,73,132,0.18082191,86.3,376.487,11759.498,420700.66
2454892,48.049999,-67.099998,77.5,186,0.23999999,65.2,247.84599,10246.392,15979.454
2454892,45.02,-73.5,27.1,76,0.2804428,0.3,65.727798,9749.293,266377.31
2454892,48.18,-78.330002,95,211,0.22210526,68.3,319.93201,10706.105,356213.31
2454892,50.279999,-89.029999,66,124.5,0.18863636,46.3,353.30301,10766.453,556770.19
2454892,44.990002,-77.970001,51.799999,130.8,0.25250965,28.2,372.93201,11276.472,563923.31
2454892,47.830002,-80.43,96.400002,177.8,0.18443984,70.2,316.14899,11828.217,352534.72
2454892,47.279999,-81.25,72.900002,132.10001,0.18120714,78.5,398.71701,12340.142,444685.59
2454892,41.377201,-75.5364,0,0,0,0,485.15601,12118.175,134204.81
```

Figure: validation_sdn_2454892.csv file

- The second R script integrated in the main program is UK_sdepth_tm.R. The purpose of running this r script is to adjust the CMC snow depth analysis product in the prediction grids. The outputs from this script is used as the final snow depth on grid. The input file necessary is sample_sdn_juliandate.csv, validation_sdn_juliandate.csv and mgrid_juliandate.csv. The output files are sample_sdepth_juliandate.csv, validation_sdepth_juliandate.csv and mgrid_sdepth_julianddate.csv. Below a preview of sample_sdepth_2454892.csv and mgrid_sdepth_2454892.csv is provided. An error file of snow depth is also generated from this program and a preview is provided below.

```
JD,y,x,elev,vm,dist_ocean,CMC_sd,sda,sdepth
2454892,38.039631,-96.721809,441.271,3649.2866,934548.56,2.1,2.1,2.1
2454892,38.217369,-96.789833,430.50601,3551.3218,934996.75,2.2,2.2,2.2
2454892,38.395321,-96.858414,423.62399,4107.7725,934622.31,2.2,2.2,2.2
2454892,38.57346,-96.927521,436.14001,2681.1626,934533.56,2.2,2.2,2.2
2454892,38.751808,-96.997177,388.01401,2787.5237,934962.69,2.2,2.2,2.2
2454892,38.092659,-96.495987,430.237,3840.2571,926766.94,2.2,2.2,2.2
2454892,38.270691,-96.56321,404.07199,3896.636,929680.69,2.4,2.4,2.4
2454892,38.448929,-96.630959,404.07199,3979.6985,931596.56,2.6,2.6,2.6
2454892,38.627369,-96.699249,434.07101,3866.9514,933106.75,2.7,2.7,2.7
2454892,38.806011,-96.768089,388.01401,3181.5874,934595.38,2.7,2.7,2.7
2454892,38.984859,-96.837479,388.01401,3215.4233,936087.19,2.7,2.7,2.7
2454892,39.16391,-96.907417,370.61401,3329.6587,966864.44,2.7,2.7,2.7
2454892,39.343151,-96.977943,395.91901,2984.3333,973633,2.8,2.8,2.8
2454892,38.145012,-96.269653,430.237,3376.8025,918630.56,2.3,2.3,2.3
2454892,38.32333,-96.336037,381.10901,3577.8345,923711.44,2.6,2.6,2.6
2454892,38.50185,-96.402969,401.49701,3946.2913,928374.69,2.9,2.9,2.9
2454892,38.68058,-96.470444,401.49701,4089.0737,932033.94,3.1,3.1,3.1
2454892,38.85952,-96.538437,432.01401,3788.2136,934932.88,3.3,3.3,3.3
2454892,39.03867,-96.606987,363.54999,3825.6238,966077.94,3.4,3.4,3.4
2454892,39.21801,-96.676086,363.54999,3876.4263,973946.69,3.5,3.5,3.5
2454892,39.39756,-96.745758,391.064,3535.1167,981174.56,3.7,3.7,3.7
2454892,39.577301,-96.815987,406.13599,2363.7312,987545.63,3.9,3.9,3.9
2454892,39.757252,-96.886803,401.33801,2120.8557,992941.81,4.1,4.1,4.1
2454892,39.937401,-96.958183,401.33801,2028.7667,997345.63,4.2,4.2,4.2
```

Figure: mgrid_sdepth_2454892.csv file

```
JD,y,x,sd,swe,d,CMC_sd,elev,vm,dist_ocean,sdepth
2454892,48.630001,-81.400002,96.5,147.3,0.15264249,104.3,305.099,11675.797,294204.91,91.3351284878911
2454892,50.279999,-89.029999,66,124,0.18787879,46.3,353.30301,10766.453,556770.19,65.3985582646265
2454892,45.52,-77.900002,53,150,0.28301886,27.3,442.30701,10904.539,554575.88,57.7790949966221
2454892,46.700001,-73.889999,61.700001,158,0.2560778,32.3,468.91101,10681.572,215498.03,68.3054276265502
2454892,47.669998,-81.720001,97.300003,185.39999,0.19054469,90.6,393.056,11574.215,398834.31,87.2344392436528
2454892,47.970001,-81.599998,72,127,0.17638889,97.8,359.548,11100.913,389125.97,82.9241315144314
2454892,48.549999,-80.699997,93,145,0.15591398,98.8,273.44601,10024.651,297806.75,93.5409049020335
2454892,47.830002,-80.43,96,178,0.18541667,70.2,316.14899,11828.217,352534.72,94.7786653271365
2454892,46.380001,-71.650002,76.599998,191,0.24934727,44.8,155.228,10789.63,85515.742,79.4843250242423
2454892,49.779999,-94.370003,44,109,0.24772727,58.1,355.21301,9489.9736,793413.38,48.4744767448289
2454892,44.950001,-78.699997,63.5,172.7,0.27196851,37.3,331.29599,11560.695,598880.06,58.7746869375164
2454892,47.279999,-79.5,93.5,147.3,0.15754011,46.6,236.62601,8727.2197,425563.13,80.4215765474488
2454892,45.25,-76.75,35.299999,106.7,0.3022663,7.3,167.767,10980.852,459457.69,39.0824646955776
2454892,51.169998,-90.220001,54.599998,109.2,0.2,54.4,371.367,9569.4199,570899.94,59.6378965427638
2454892,45.919998,-79.220001,35,79,0.22571428,38.4,429.23599,10869.604,572964.69,50.396939478495
2454892,48.330002,-81.32,87,147,0.16896552,106.3,319.30399,10805.814,335528.69,91.0385315174427
2454892,49.77,-86.919998,46,71,0.15434782,45.2,320.30899,10012.904,465468.94,50.3711018994067
2454892,47.080002,-79.279999,64.5,144.8,0.22449613,51.4,263.68399,10002.105,440399.5,74.8209403607216
2454892,47.990002,-80.699997,88,152,0.17272727,77.3,342.552,11436.429,375143.28,92.2933392499659
2454892,47.630001,-69.57,78.199997,212,0.27109975,57.7,173.25999,6729.1372,27503.17,80.3360783253623
2454892,49.080002,-87.07,53.299999,132.10001,0.24784242,39.1,417.74701,11434.67,520555.47,56.0341081856488
2454892,46.630001,-81.769997,97.300003,228.60001,0.23494348,68.5,406.461,12212.555,508263.59,90.0616017969942
2454892,47.470001,-79.830002,94,147,0.15638298,34.4,202.569,9026.1182,401844.94,34.4
2454892,46.580002,-74.169998,82.800003,190,0.22946858,38.6,471.57199,10924.043,242701.16,75.5426056515412
```

Figure: sample_sdepth_2454892.csv file

```
Julian_date,RMSE_cmc_cv,RMSE_cmc_v,RMSE_cv,RMSE_v,RMSE_nn_v,RMSE_nn_cv
2454892,26.6055646289601,18.521798063903,13.6757848340199,13.4113595569947,12.9198945361409,19.7888886116753
2454893,21.7410417003892,25.5346011463909,12.2193178752385,14.6642351993568,19.7115152903934,14.5760933703211
2454894,26.8214060648016,32.963018616322,13.1042272285564,21.6335609548279,27.4817618147423,20.8860005703559
2454895,30.387687965406,17.4999297241447,17.1206672663811,13.2679309484625,13.1258488268242,22.043934000255
2454896,19.5037944805208,10.3247275993122,17.1389270886034,24.4234022912238,11.9163529668807,8.60748463363036
2454897,17.247274326,NA,15.8106696197971,NA,NA,14.800342118753
2454898,17.4979556230144,NA,17.090380385589,NA,NA,9.95793163724892
2454899,27.5803974509132,NA,25.1000660182663,NA,NA,26.0729278444847
2454900,20.1919447423659,5.65685424949238,13.7132810007504,2.10562716757648,2.02407638978417,12.261785595669
2454901,20.5808388604316,NA,14.8985348534276,NA,NA,12.6542067049453
2454902,14.3936138303358,NA,13.1733825317354,NA,NA,9.99113461855385
2454903,25.6970559147069,NA,25.0233078833134,NA,NA,20.5846221513647
2454904,20.4481241943193,32.3147799002252,14.9632503851226,1.14781799200581,32.3310170950715,18.6764379289958
2454905,19.5368044811294,16.1220346110533,16.4357246208922,3.68977537811967,5.03016328031496,14.2610614459131
2454906,23.3974337063427,16.7266588229024,16.281795165709,0.888468763137992,2.07756404226501,15.2510480284774
2454907,24.9343648360221,18.5486482291679,15.015087752457,11.2430718509124,11.0841831922433,16.1076081764517
2454908,25.2974174094916,22.0446948590488,11.9856428698568,24.2720762305812,23.4676689573885,18.6289751273841
2454909,19.8629326177758,35.3282216223801,14.7001078889183,2.89719312412434,33.5319653550156,13.730002189889
2454910,13.9443832539496,15.6977705423414,18.9458456223411,14.154409852325,15.6977705423414,13.9443832539496
2454911,12.4771512889547,0.494974746830583,11.9960192095683,2.02628093002526,1.80919192664173,7.98511651618035
2454912,13.516813507692,NA,12.506692224201,NA,NA,11.0013727910291
2454913,14.7994950875169,NA,10.1277487352071,NA,NA,9.9957641468687
2454914,28.3017297887356,NA,15.649123315493,NA,NA,21.5556313808528
2454915,20.6751364138617,NA,14.9879566907114,NA,NA,10.8664259542154
```

Figure: error_sdepth.csv file

- There are two other r scripts included in the main_swe.R. One is swe_inter_tm.R and another is swe_gap_tm.R. Both these two r scripts are used to predict the final swe from adjusted CMC snow depth product. The first R script is used for scenario-I and the second R script is used for scenario-II. Input files in these R scripts are the output from UK_sdepth_tm.R. There are lots of output files produced from this R script. sample_swe_juliandate.csv, mgrid_swe_juliandate.csv, validation_swe_juliandate.csv and also error file for each individual day. These files are not the final swe product. Final swe product file is swe_predict_juliandate.csv. A preview of swe_predict_2454892.csv is provided below. An error file for the full simulation period is also generated. This file contains associated RMSE error of SWE for each simulation day. A preview of error_final.csv has been also provided.

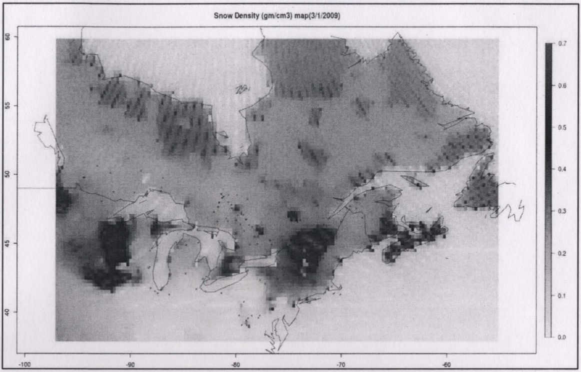
```
x,y,swe.pred,snowdepth,snowdensity
-96.721809,38.039631,5.18823483710177,2.1,0.247058801766751
-96.789833,38.217369,5.39687505699736,2.2,0.245312502590789
-96.858414,38.395321,5.36100573636532,2.2,0.243682078925696
-96.927521,38.57346,5.32867455768509,2.2,0.242212479894777
-96.997177,38.751808,5.29999351550882,2.2,0.240908796159492
-96.495987,38.092659,5.28658842248524,2.2,0.240299473749329
-96.56321,38.270691,5.72418828056021,2.4,0.238507845023342
-96.630959,38.448929,6.15878086045518,2.6,0.236876186940584
-96.699249,38.627369,6.3550651977996,2.7,0.235372785103689
-96.768089,38.806011,6.31810708426152,2.7,0.23400396608376
-96.837479,38.984859,6.2859428603918,2.7,0.232812698533029
-96.907417,39.16391,6.25868914778096,2.7,0.231803301769665
-96.977943,39.343151,6.46850850920638,2.8,0.231018161043085
-96.269653,38.145012,5.36931588398239,2.3,0.233448516694887
-96.336037,38.32333,6.0217435932028,2.6,0.231605522815492
-96.402969,38.50185,6.66792945553498,2.9,0.229928601914999
-96.470444,38.68058,7.0799948185211,3.1,0.228386929629713
-96.538437,38.85952,7.49175841601243,3.3,0.227022982303407
-96.606987,39.03867,7.67743038032258,3.4,0.225806775891841
-96.676086,39.21801,7.86729832255845,3.5,0.224779952073099
-96.745758,39.39756,8.2873726587895,3.7,0.223983044832149
-96.815987,39.577301,8.71194572895635,3.9,0.223383223819394
-96.886803,39.757252,9.14234958601116,4.1,0.222984136244175
-96.958183,39.937401,9.35558251592937,4.2,0.222751964664985
```

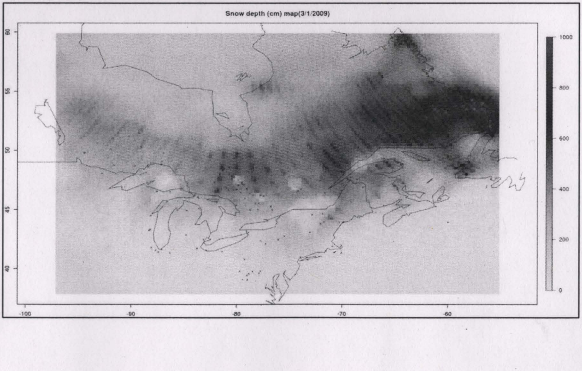
Figure: swe_predict_2454892.csv file

```
Julian_date,RMSE_nn_mm
2454833,7.89446621728533
2454834,20.8629476262892
2454835,12.9372600429426
2454836,22.6362377725243
2454837,20.3673241933546
2454838,16.8571523914477
2454839,13.6742369125769
2454840,13.0413775523625
2454841,7.63625891735768
2454842,9.37771393927058
2454843,12.5412022274669
2454844,38.8810852087993
2454845,32.5051396463845
2454846,25.3294117867267
2454847,27.3594347324435
2454848,12.4081540224632
2454849,30.5811759610902
2454850,11.4208752793285
2454851,21.6667715398841
2454852,29.4191692935845
2454853,21.2798270833644
2454854,22.2801485749371
2454855,15.3522536023927
2454856,32.8327324210718
```

Figure: error_final.csv file

- Different spatial maps are produced by these two r scripts. On each day SWE, snow density and snow depth maps are produced by the tool. In below a preview of SWE, snow depth and snow density map are provided.


Snow Density (gm/cm3) map(3/1/2009)
60
55
50
45
40
-100
-90
-80
-70
-60
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0


Snow depth (cm) map(3/1/2009)
60
55
50
45
40
-100
-90
-80
-70
-60
1000
800
600
400
200
0

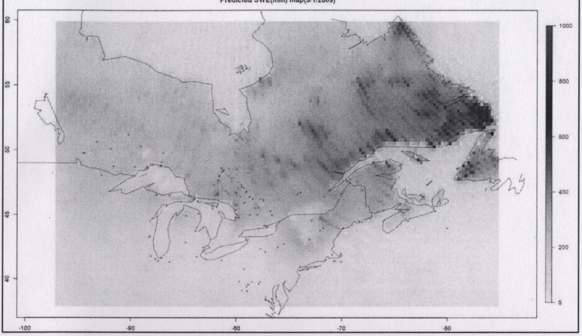


Figure: Preview of snow density_map_2454892.jpg, snow depth_map_2454892.jpg, swe_map_2454892.jpg

6. Real Time Run / Incorporating more Data Points

The developed SWE prediction tool can be used for real time SWE prediction if real time snow course data will be available. The main thing necessary for real time SWE prediction is to incorporate real time data. The CMC snow depth data product and USA data are collected from the web, so if these two datasets are updated in regular basis, then these data can be downloaded and used in the prediction tool. There is no web base data for Canadian snow course measurements, We have developed a Canadian snow course database in extension of Ross Brown's CD (MSC,2009). We have tried to get all the available snow course measurements from great lakes to Labrador. Unfortunately, data from Quebec Hydro was not collected. So, when we get data from Quebec Hydro, there must be an option to input this new data set. In case of real time analysis, new snow course measurements data need to input in daily basis.

So, in real time data analysis the only thing need to keep in mind, is to update the Canadian data base daily. So, the main objective of real time analysis, is that there must be an option to incorporate new data points in the existing database. A new folder named datainput within the SWE_MAP_V1.0 has been created for updating the Canadian database with new data sets. In this folder there are number of Fortran codes , the most recent database file and station database file. The final.txt file is updated after adding new data and can be used as input file in the data folder and to run canada.f90 Fortran code.

First thing in the datainput folder is the station information file. The station information file is named as Station_List.txt. In below, a preview of the file is provided. If we need to include new stations, then new station information (station id, station name, latitude, longitude at least) need to be included in the Station_List.txt file. input_station.f90 Fortran code can be run by gfortran to add information in the station info database file. To input station information the new file must be in a specific format. st1.prn is such a new station info file which has been inputted to Station_List.txt file by running input_station.f90 code. The Fortran code read format for the new file will be (5X,A11,2X,A31,I2,2X,I2,4X,I2,2X,I2) where at first is station id, then in serial station name, latitude in degree, latitude in minute, longitude in degree, longitude in minute. In Fortran code the name of the new station info file need to be inputted on screen.

Station Id	Station Name	Lat		Long		Elev	Start	End	#Recor
ALE-05AD803	AKAMINA	49	2	114	3	1800	19800131	20030529	140
ALE-05AA803	ALLISON PASS	49	44	114	36	1980	19630328	20030505	107
ALE-07OA801	ASSUMPTION	58	36	118	28	370	19860304	20030401	37
ALE-07BC802	BARRHEAD NORTH	54	16	114	21	670	19730315	20030327	72
ALE-07BB808	BARRHEAD WEST	54	11	114	48	670	19730314	20030327	73
ALE-05EC801	BELLIS	54	7	112	5	670	19730319	20030403	74
ALE-05CC801	BENTLEY	52	29	114	4	910	19730308	20030328	72
ALE-07GE802	BEZANSON	55	14	118	31	690	19730315	20030328	73
ALE-05FA803	BIGSTONE	53	2	113	51	850	19870304	20030328	35
ALE-06AC801	BONNYVILLE	54	30	110	40	540	19730320	20030327	71
ALE-05BA801	BOW RIVER	51	25	116	11	1580	19370330	20030401	141
ALE-05BA813	BOW SUMMIT (NEW)	51	42	116	28	2080	19790227	20030528	119
ALE-05BA811	BOW SUMMIT (OLD)	51	42	116	28	2080	19680229	19800430	35
ALE-05DD801	BRAZEAU RES.	52	57	115	41	970	19770228	20030402	54
ALE-05DD802	BROWN CREEK	52	45	116	33	1340	19770301	20030401	54
ALE-05EE802	BRUCE SNOW PL	53	17	112	4	670	19750313	20030328	94

Figure: Preview of Station List.txt file

STATION_NUMBER	NAME	lat_	lat_	mlon_	dlon_m
SNOW-MNR-2401	CHRISTIES CORNERS	43	17	-80	2
SNOW-MNR-2402	MOUNT ALBION	43	12	-79	50
SNOW-MNR-2403	VALENS	43	23	-80	8
SNOW-MNR-2404	DUNDAS VALLEY	43	15	-80	1
SNOW-MNR-2501	ETHEL	43	43	-81	7
SNOW-MNR-2502	GALBRAITH	43	38	-80	55
SNOW-MNR-2503	MORLEY (GORRIE)	43	55	-81	9
SNOW-MNR-2504	HARRISTON	43	55	-80	52
SNOW-MNR-2505	WESTFIELD (WAWONOSH)	43	50	-81	27
SNOW-MNR-2506	KINLOSS TWP	43	59	-81	26
SNOW-MNR-2507	WALLACE TWP	43	44	-80	53
SNOW-MNR-2508	MCKILLOP TWP	43	39	-81	18
SNOW-MNR-2601	BLAKENEY	45	15	-76	15
SNOW-MNR-2602	BRIGHTSIDE	45	7	-76	30
SNOW-MNR-2605	MABERLEY	44	50	-76	34
SNOW-MNR-2606	BON ECHO PARK	44	54	-77	12

Figure: Preview of new station info (st1.prn) file

To input new snow data in the database, we use input_snow.f90 code. It will create new database file by combining old database file and new data file. There is option in this input_snow.f90 program to input the old database file name, new snow course data file name. In Canadian snow course data base file the first column is station id, then date in yyyymmdd format after that snow depth in cm and than swe in mm. The new data file must be in a specific format to input in the database. It also needs to be noted that the data in new data file must be arranged in ascending order. The Fortran code read format for the new file will be (5X,A11,4X,I4,I2,I2,2X,F6.1,2X,F6.1) which in serial station id, year, month, day, snow depth in cm and swe in mm. A preview of both database file and new data file has been provided below.

BCE-2D01	19350322	152.0	389.0
BCE-1D01	19350323	3.0	3.0
BCE-2F01	19350328	91.0	246.0
BCE-2F02	19350330	94.0	259.0
BCE-2F04	19350403	122.0	320.0
BCE-2F03	19350405	79.0	188.0
BCE-2C01	19360304	48.0	89.0
BCE-2D01	19360322	163.0	401.0
BCE-2F02	19360322	74.0	185.0
BCE-3A01	19360323	320.0	1316.0

Figure: Preview of Canadian snow course database file

SNOW-MNR-0301	20041115	0.0	0.0
SNOW-MNR-0302	20041115	0.0	0.0
SNOW-MNR-0303	20041115	0.0	0.0
SNOW-MNR-0304	20041115	0.0	0.0
SNOW-MNR-0305	20041115	0.0	0.0
SNOW-MNR-0306	20041115	0.0	0.0
SNOW-MNR-0307	20041115	0.0	0.0
SNOW-MNR-0308	20041115	0.0	0.0
SNOW-MNR-0401	20041115	0.0	0.0
SNOW-MNR-0402	20041115	0.0	0.0
SNOW-MNR-0450	20041115	0.0	0.0
SNOW-MNR-0607	20041115	0.0	0.0
SNOW-MNR-1401	20041115	0.0	0.0
SNOW-MNR-1501	20041115	0.0	0.0
SNOW-MNR-1601	20041115	0.0	0.0

Figure: Preview of new snow course data file

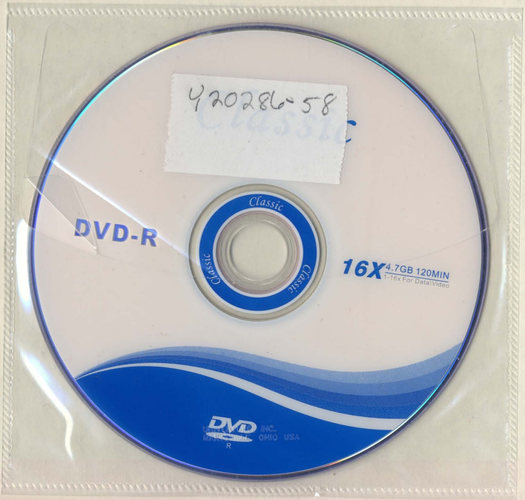
In this SWE_MAP_V1.0 Julian date has mainly been used inspite of normal date. To convert this database file to Julian day database, to combine station database file and the snow course database file and to assign numerical station id, we use another Fortran code named compile.f90. This Fortran code has been used to convert database file in final.txt. These final.txt is used as input file in the data processing part of the tool. A new station id file named new_id.txt has also been created. In compile.f90 code there is an option to input the database file name which will convert to final.txt later. A preview of final.txt has been provided below. In the first column there is Julian day, then new numerical station id, after that Lat and Lon in degree and at last snow depth in cm and swe in mm.

2427884	356	50.53	-117.28	152.0	389.0
2427885	232	49.37	-122.80	3.0	3.0
2427886	0	0.00	0.00	-999.0	-999.0
2427887	0	0.00	0.00	-999.0	-999.0
2427888	0	0.00	0.00	-999.0	-999.0
2427889	0	0.00	0.00	-999.0	-999.0
2427890	380	49.73	-120.18	91.0	246.0
2427891	0	0.00	0.00	-999.0	-999.0
2427892	381	49.82	-120.02	94.0	259.0
2427893	0	0.00	0.00	-999.0	-999.0
2427894	0	0.00	0.00	-999.0	-999.0
2427895	0	0.00	0.00	-999.0	-999.0
2427896	383	49.99	-118.87	122.0	320.0
2427897	0	0.00	0.00	-999.0	-999.0
2427898	382	49.78	-119.20	79.0	188.0
2427899	0	0.00	0.00	-999.0	-999.0
2427900	0	0.00	0.00	-999.0	-999.0
2427901	0	0.00	0.00	-999.0	-999.0
2427902	0	0.00	0.00	-999.0	-999.0
2427903	0	0.00	0.00	-999.0	-999.0
2427904	0	0.00	0.00	-999.0	-999.0
2427905	0	0.00	0.00	-999.0	-999.0
2427906	0	0.00	0.00	-999.0	-999.0
2427907	0	0.00	0.00	-999.0	-999.0
2427908	0	0.00	0.00	-999.0	-999.0
2427909	0	0.00	0.00	-999.0	-999.0

Figure: Preview of final.txt file

7.0 Conclusion

In this appendix, more basic things related with SWE_MAP_V1.0 is discussed. The concept and logics behind developing this tool has been described first. The procedure of using this tool to predict SWE, is also described here.


420286-58
Classic
DVD-R
16X 4.7GB 120MIN
1-16x For Data\Video