ADAPTING REMOTELY SENSED SNOW DATA FOR DAILY FLOW MODELING ON THE UPPER HUMBER RIVER, NEWFOUNDLAND AND LABRADOR

MELISSA TOM

ADAPTING REMOTELY SENSED SNOW DATA FOR DAILY FLOW MODELING ON THE UPPER HUMBER RIVER, NEWFOUNDLAND AND LABRADOR

By

© Melissa Tom

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

June 2010

St. John's, Newfoundland

Abstract

This thesis investigated the use of remotely sensed snow information to help improve flood forecasting in western Newfoundland's Humber River Basin. Flood forecasting on the Humber River is important because of the large population settlements within the Humber Valley. In this research, two types of remotely sensed snow data were considered for analysis: (1) snow cover (or snow extent) and (2) snow water equivalent (SWE). The majority of this thesis focuses on the remotely sensed snow cover data. Moderate Resolution Imaging Spectroradiometer (MODIS) Terra snow cover images were acquired over the Humber Valley watershed throughout the snowmelt period, from March to June, for the years 2000 to 2009. MODIS is an optical sensor on NASA's (National Aeronautics and Space Administration) Earth Observing System (EOS) Terra and Aqua satellites. Its daily temporal data are advantageous and the data are free and easily accessible. Daily snow cover data were extracted from the National Snow and Ice Data Center (NSIDC) daily snow product, specifically MOD10A1: a product derived from MODIS data, using a custom EASI script run in PCI Geomatica. PCI Geomatica is a robust remote sensing and image processing software. One major obstacle, regarding the acquisition of MODIS imagery over the Humber Valley watershed, is the presence of over 50% cloud cover for 80% of the days on average from March to June every year. This was a concern for data collection: affecting the sample size of acquired data and the accuracy of the snow cover data. When cloud cover is high there is a greater chance that it may be misclassified as snow and/or snow is misclassified as cloud cover. For this

reason, a cloud-cover threshold was determined. The Rango-Martinec snowmelt runoff model, a widely used degree-day model which incorporates snow cover data as a direct input, was evaluated. It was found that the next day's flow is highly dependent on the previous day's flow and less dependent on the meteorological data: rainfall, snow cover, and temperature. The results from the snowmelt runoff model using the snow cover data provided very good final Nash-Sutcliffe coefficients of 0.85 for the calibration stage and 0.81 for the validation stage, but a consistent one-day lag of the modeled flow values was also observed. Although these results were not superior to currently employed flood forecasting models for the Upper Humber (because of a one-day lag in the modeled flows), the methodology developed herein may be useful for other river basins in NL where the flows are dominated by snowmelt during the spring such as the Exploits River Basin located in central NL. Remotely sensed snow water equivalent (SWE) data obtained from an advanced microwave scanning radiometer (AMSR-E), aboard the Aqua satellite, was also investigated for daily flow modeling applications. SWE often provide a better estimate of snowmelt than snow cover but this data had several disadvantages in the Humber River Basin. The major obstacles included large spatial resolution (25 km), data inaccuracy for wet snow, boreal forest, mountainous regions, and time step irregularities. Extremely large variances in the SWE data rendered the information inaccurate and ineffective for streamflow forecasting on Newfoundland and Labrador's Humber River. This research makes significant contributions to the field of hydrology providing a valuable methodology in adapting remotely sensed snow data to daily flow simulation and will be helpful to local authorities.

Acknowledgements

I would like to give my sincere thanks to Dr. Leonard Lye for his unwavering support, knowledge, and encouragement. I could not have wished for a better supervisor – one who is dedicated to his students and a natural giver. I am grateful for the enormous amount of support, guidance, and resources provided to me by the Department of Environment and Conservation, Government of Newfoundland and Labrador and C-CORE. I would especially like to thank Dr. Amir Ali Khan, Mr. Thomas Puestow, and Mr. Keith Abbott. Of course, I would like to thank the Institute for Biodiversity, Ecosystem Science, and Sustainability (IBES) as well as Memorial University, specifically the Faculty of Engineering and Applied Science, for their support throughout my research. Lastly, I would like to express my appreciation to my fiancé, David McComiskey, for his love, support, patience, and laughter all throughout my research and thesis preparation.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

a – Degree-Day Factor

AMSR-E – Advanced Microwave Scanning Radiometer – Earth Observing System

ANN – Artificial Neural Network

ANOVA – Analysis of Variance

BPNN – Back Propagation Neural Network

c_r – Rain Runoff Coefficient

c_s – Snow Runoff Coefficient

CCD – Central Composite Design

CDC – Conventional Depletion Curve

CEQUEAU – Model (Canada): Developed at the University of Quebec

D_v – Volume Difference

DC – Depletion Curve

DEM – Digital Elevation Model

DLPC – Deer Lake Power Company

DOE – Design of Experiments

E – Nash-Sutcliffe Model Efficiency Coefficient

EASE-Grid – Equal-Area Scalable Earth Grid

ECMWF – European Centre for Medium-Range Weather Forecasts

EO – Earth Observation

EOS – Earth Observing System

FTP – File Transfer Protocol

GIS – Geographic Information Sciences

GRNN – General Regression Neural Network

HDF – Hierarchal Data Format

IBES – Institute for Biodiversity, Ecosystem Science, and Sustainability

IR – Infrared

k – Recession Coefficient

MDC – Modified Depletion Curve

MDCI – Modified Depletion Curve I

MDCII – Modified Depletion Curve II

MODIS – Moderate Resolution Imaging Spectroradiometer

MUN – Memorial University of Newfoundland

NASA – National Aeronautics and Space Administration

NDSI – Normalized Difference Snow Index

NDVI – Normalized Difference Vegetation Index

NN – Neural Network

NSIDC – National Snow and Ice Data Center

OLM – Object-Level Metadata

QA – Quality Assessment

R^2 – Coefficient of Determination

SCU – Snow Cover Units

SRM – Snowmelt Runoff Model

SSARR – Streamflow Synthesis and Reservoir Regulation

SWE – Snow Water Equivalent

TCP/IP – Transmission Control Protocol/ Internet Protocol

UTM – Universal Transverse Mercator

VB – Visual Basic

VIR – Visual Image Retrieval

WIST – Warehouse Inventory Search Tool

WMO – World Meteorological Organization

WRMD – Water Resources Management Division

List of Appendices

- Chapter 1 -

Introduction

1.1 Purpose

The aim of this research project is to determine the role of remotely sensed snow data in daily flow modeling on the Humber River, Newfoundland and Labrador, using the Rango-Martinec snowmelt runoff model.

1.2 Overview

Daily flow predictions are required for forecasting floods. Rain-runoff models are used to forecast flow rates and water levels using real-time or periodic rainfall and discharge data. These predictions can range from hours to days ahead. There are several rationales on flood forecasting. The main reason is to implement flood control and mitigation; this includes protection of settlements through proper and timely management and warning protocols. Other reasons for flood forecasting are to control reservoir levels and handle water volumes for appropriate hydroelectric power production year-round. To be specific, operators of large reservoirs would be able to plan for expected inflows and therefore maximize the hydropower generation from the reservoir (Bettwy 2004).

The level of importance of snowmelt in flood generation depends on the region within Canada. Generally, the larger the basin, the more the snowmelt runoff will dominate over rainfall runoff contributions (Watt 1989). According to the Canadian Flood Guide of 1993, the four main causes of flooding in Newfoundland are (1) rainfall alone, (2) rainfall plus snowmelt, (3) tidal effects (in some coastal areas), and (4) ice jamming. For the Humber River Basin, flood forecasting is of great significance because of the large settlement of people in and around the area, a growing population of over 30,000 (Statistics Canada 2006). The Deer Lake hydropower generating system is also affected by the predicted flow rates. Currently, the provincial government of Newfoundland and Labrador is not including any sort of snow cover or snow water equivalent data into its flood forecasting model. In the past, however, snowmelt has been assessed by the Water Resources Management Division (WRMD) of the Department of Environmental and Conservation of Newfoundland and Labrador Provincial Government, using the deterministic model: Streamflow Synthesis and Reservoir Regulation (SSARR) model. The WRMD halted the operation of this model because, over time, the SSARR model became inaccurate in its flow forecasts: overestimating the amount of snowmelt in the spring (Cai 2009).

1.3 The Study Area

The Humber River Basin is located in western Newfoundland, Canada, shown in Figure 1.1. It is approximately at latitude and longitude coordinates 49° N, -58°E. It is the second largest river system on the island with a drainage area of over 8,000 km^2. Its outlet

is located in the Bay of Islands, close to the Humber Village Bridge hydrometric station, Figure 1.2, flowing into the Atlantic Ocean. Over half of its drainage area is regulated by the Deer Lake Power Company (DLPC) for hydroelectric power generation.

Figure 1.1: Map of Canada and the Island of Newfoundland

The basin's climate, during the winter and early spring, is snowy and rainy with average temperatures ranging from -20 °C to 0 °C. The Humber River Basin can experience freezing rain when temperatures hover around 0 °C and it frequently endures average wind speeds of 20 km/h. The region is categorized as dense forest with canopy cover

greater than 75%. It contains Black Spruce (*Picea mariana*) and is located in Canada's boreal forest (Water Resources Management Division 2009). Compared to adjacent prairie and tundra areas, the coniferous boreal forest experiences a great deal more snow accumulation and delayed snowmelt, attributable to the forest canopy coverage (Seidel and Martinec 2004).

This watershed is divided into two parts, based on elevation and location. The Upper Humber is smaller in size but higher in elevation, located in the northern, mountainous area of the basin. The Lower Humber makes up the remainder, southern part of the basin which includes Grand Lake, Deer Lake, and the Deer Lake power generating station. The average elevation in the Lower Humber is approximately 100 m, whereas the elevation in the Upper Humber ranges from 600 m to 800 m. The high elevation is one of the major reasons for almost 100% snow cover over the Upper Humber from October to April (Cai 2009).

The focus of this research assesses the use of remote sensing of snow distribution to improve flood forecasting for the Upper Humber River basin above Black Brook, as shown in Figure 1.2 (highlighted and shaded in red). The yellow outline, in Figure 1.2, delineates the entire Greater Humber watershed with its outlet into the Atlantic Ocean located close to the Humber Village Bridge station. The hydrometric stations are shown with black dots and labelled by their unique station names (i.e. 02YL008). The

hydrometric stations are the locations where hourly water levels are recorded and flow rates are derived from stage-discharge curves.

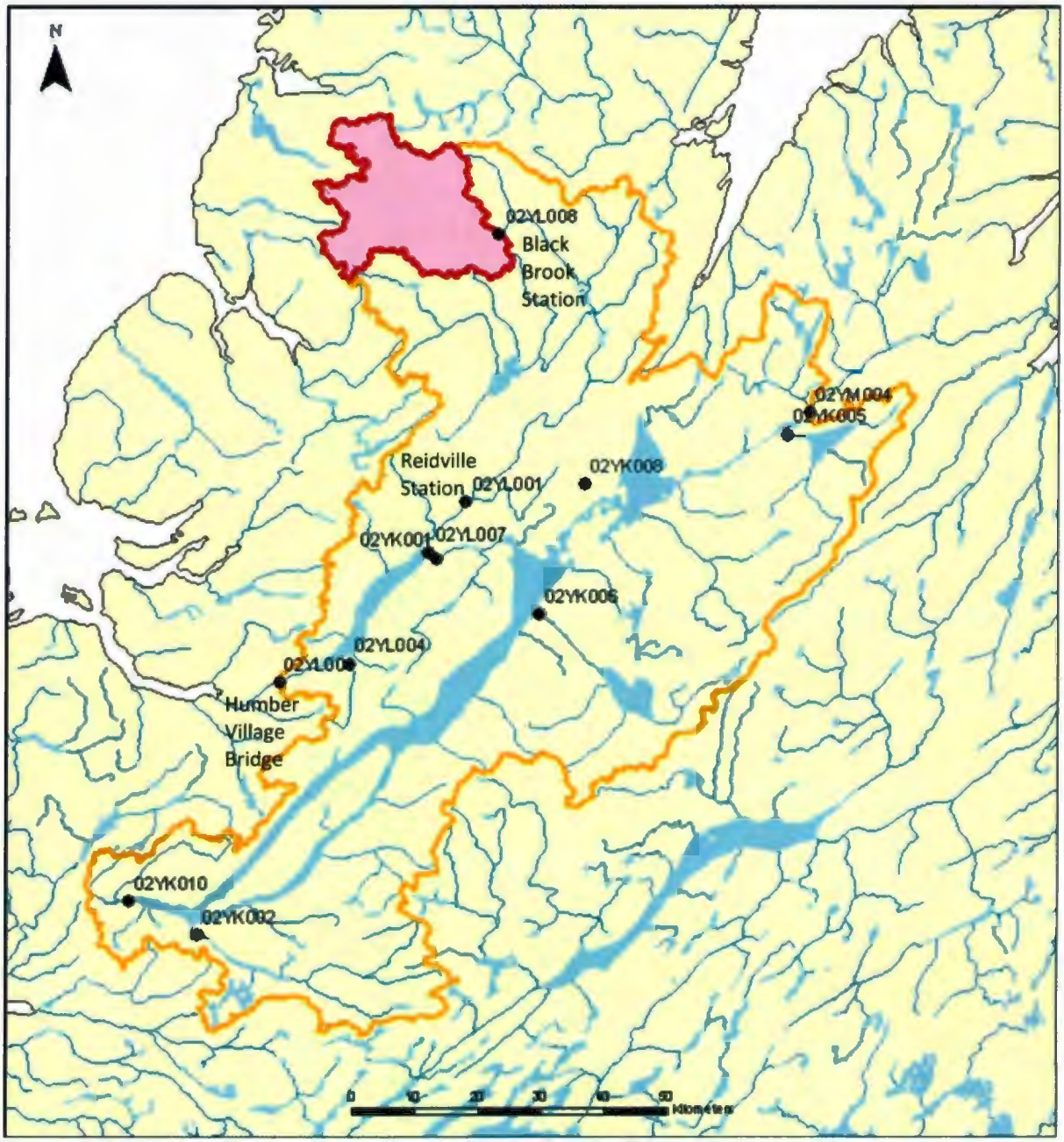


Figure 1.2: Upper Humber River above Black Brook (Water Resources Management Division 2009)

The Upper Humber above Black Brook region was chosen as a study area for several reasons. The main reason was to monitor the flows in this northern part of the basin because it retains its snow cover longer than all other areas of the watershed. The completion of the seasonal snowmelt is often followed by increased flow rates. This increase in water volume sometimes leads to flooding in the Humber River. The occurrence of this potential flooding is often unexpected in the Lower Humber, as the snow in the Lower Humber melts earlier than the snow in the Upper Humber. Additionally, this region, as opposed to the western portion of the Humber River, does not have any associated snow cover monitoring program (Water Resources Management Division 2009). Lastly, the Upper Humber has natural, unregulated flows which provide direct and comparable flows for modeling input.

1.4 Current Flood Forecasting Methods on the Humber River, NL

The Government of Newfoundland and Labrador has mandated the WRMD, Department of Environment and Conservation, to provide flood forecasting services for the Humber River, NL. Over the past 20 years the department has used various forecasting models. The most recent model, still being used by the WRMD but only as an interim model, is the dynamic regression model. This is a statistically based model. The flow is predicted based on a linear time series of lagged flows and precipitation data. This model predicts better than its predecessor, but it does not incorporate any snowmelt from the Upper Humber region for flood predictions. Also, the dynamic regression model, being a simple linear regression model, does not capture any nonlinear hydrological effects (Cai 2009).

Over the past year, another model was developed and is being tested by the WRMD alongside the dynamic regression model. This is an artificial neural network (ANN) model developed by Haijie Cai, a Civil Engineering master's student at Memorial University of Newfoundland (MUN) (Cai 2009). An ANN, usually called a "neural network" (NN), is a mathematical model which attempts to simulate the structure and/or functional aspects of reality. In most cases, an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. It is a non-linear statistical data modeling tool, applicable in many diverse fields of study, but in this case is used for hydrological modeling. Neural networks (NN) are used to find patterns in data and simulate complex relationships between inputs and outputs (Cai 2009).

Two types of ANN models were tested: general regression neural network (GRNN) and back propagation neural network (BPNN). The models were tested for the snowmelt period in 2009 at three locations on the Humber River: Black Brook Station (Upper Humber), Reidville Station, and Humber Village Bridge (refer to Figure 1.2). Both models were good predictors for the non-snow areas (Reidville and Humber Village Bridge). Both models were still good predictors for the snowy area (Upper Humber Black Brook), but the GRNN provided slightly better results than BPNN, with model efficiency Nash-Sutcliffe coefficients of 0.82 and 0.80 respectively (Cai 2009).

It is no surprise that the WRMD is interested in improving and advancing this flood forecasting service even further, given that the region of interest experiences heavy and frequent snow falls, leading to expected large snowmelt volumes. Incorporating snow data into the prediction model was the next logical step. When a river basin area remains fully-covered throughout the forecasting period, the forecasts are accurate when based solely on an index of the energy available for melting the snow (i.e., degree-day factor). For the Upper Humber, however, during the spring period forecasts, the catchment becomes partially to completely bare and assumably the snowmelt plays a significant role in flow predictions. Incorporating the snowmelt information obtained from the remotely sensed snow cover images can help predict more accurate flows (Maidment 1993).

This thesis integrates remotely sensed snow cover data into the Rango-Martinec snowmelt runoff model (SRM) to investigate possible improvement of daily flow modeling on the Upper Humber River. The Rango-Martinec SRM is one of the first and still most widely used hydrologic models with satellite snow cover as a direct input variable (Seidel and Martinec 2004). The use of remotely sensed snow water equivalent (SWE) data is also explored in hopes of further improving the Humber River's daily flow modeling.

1.5 Research Objectives

The research objectives for this thesis are fourfold. The first three encompass the primary objectives. The remaining one included as research progressed is considered a secondary objective. They are:

1. To determine the most advantageous satellite snow data available with regard to timeliness, quality, and cost. Also to determine how the snow data is obtained and what type of snow data is available through various sensors onboard satellites.
2. To acquire satellite images of snow cover data and to manipulate, validate, and manage this snow cover data through methods/processes such as geo-referencing and metadata analysis. Also to implement a snow algorithm and automate data extraction by running programming scripts.
3. To incorporate the remotely sensed snow cover data into the Rango-Martinec snowmelt runoff model used to forecast daily flow rates in the Humber River Basin. This includes the calibration of parameters using design of experiments (DOE) and a validation phase to evaluate the model's prediction accuracy.
4. To investigate and obtain remotely sensed snow water equivalent (SWE) data for the possibility of further snowmelt analysis. If the data are deemed reliable, it is likely to improve snowmelt estimates in terms of predicted total volume of water and timing of snowmelt.

1.6 Thesis Outline

This first chapter provided a brief overview of the research to be presented in this thesis. It also described the study area and one-day ahead flood forecasting models currently being used by the WRMD on the Humber River Basin. Chapter two introduces remote sensing of snow cover and provides in depth information on how the data were obtained, validated, and manipulated. This manipulation transforms the data into a practicable format for SRM model input. Snow cover depletion curves are also discussed and derived from the satellite images. Chapter three continues to address the primary objectives, with information on the snowmelt runoff model used in the daily flow modeling. In addition, this chapter provides the methodology and results for the calibration and validation stages in modeling on the Upper Humber River, NL. Chapter four addresses the secondary objective, investigating and obtaining SWE data from remote sensing methods. This chapter also explores the possibility of updating and improving the snowmelt runoff model using new snowmelt data. Chapter five entails a summary of the results from this research, a discussion on the methodologies developed in this thesis, and the potential applications for this research. Chapter six concludes this thesis with a conclusion and recommendations for further research; this final chapter is followed by references and appendices.

- Chapter 2 -

Remotely Sensed Snow Data

Remote sensing uses a real-time sensing or a recording device which is connected wirelessly to a platform. The platform ideally would be a satellite but could also be an aircraft or any object which does not physically touch the object being observed, such as ground-based supports. Remote sensing allows real-time observation of the Earth's surface and/or the events at particular locations. This is useful in observing vast, dangerous, and/or inaccessible areas. In Geographic Information Systems (GIS), remote sensing is considered a primary data source. A primary data source is described as obtaining data directly from the source without any type of mediator (Longley et al. 2005).

There are several types of sensors used in remote sensing: all of which provide unique information about the Earth's surface properties. For example, thermal sensors measure changes in surface temperatures, multispectral scanners measure reflective solar radiation and albedo to differentiate between snow and no snow, and microwave sensors measure dielectric properties to determine moisture content for snow and soil. Remote sensing is based on measuring components of the electromagnetic spectrum. Reflected or emitted energy is measured from the Earth's surface and a unique spectrum signal returns for a

specific Earth property that is being investigated. Certainly, the key feature in remote sensing is that the sections that can be used within the electromagnetic spectrum are limited by the properties of the Earth's surface and/or landscape characteristics required for analysis (Maidment 1993).

Remote sensing can provide significant data used to complement the conventional data. This new direction allows for exciting expansions in hydrology; it can help hydrologists undertake previously unsolvable problems such as exploring vast remote areas in a timely manner (Maidment 1993). Its practical applications to aid in flood forecasting are fairly new: practical because of the daily temporal data available via satellite. For this analysis, remote sensing is specifically used to collect snow cover data for the Upper Humber Basin. The flow diagram in Figure 2.1 illustrates how this data acquisition interconnects with predicting flows in rivers.

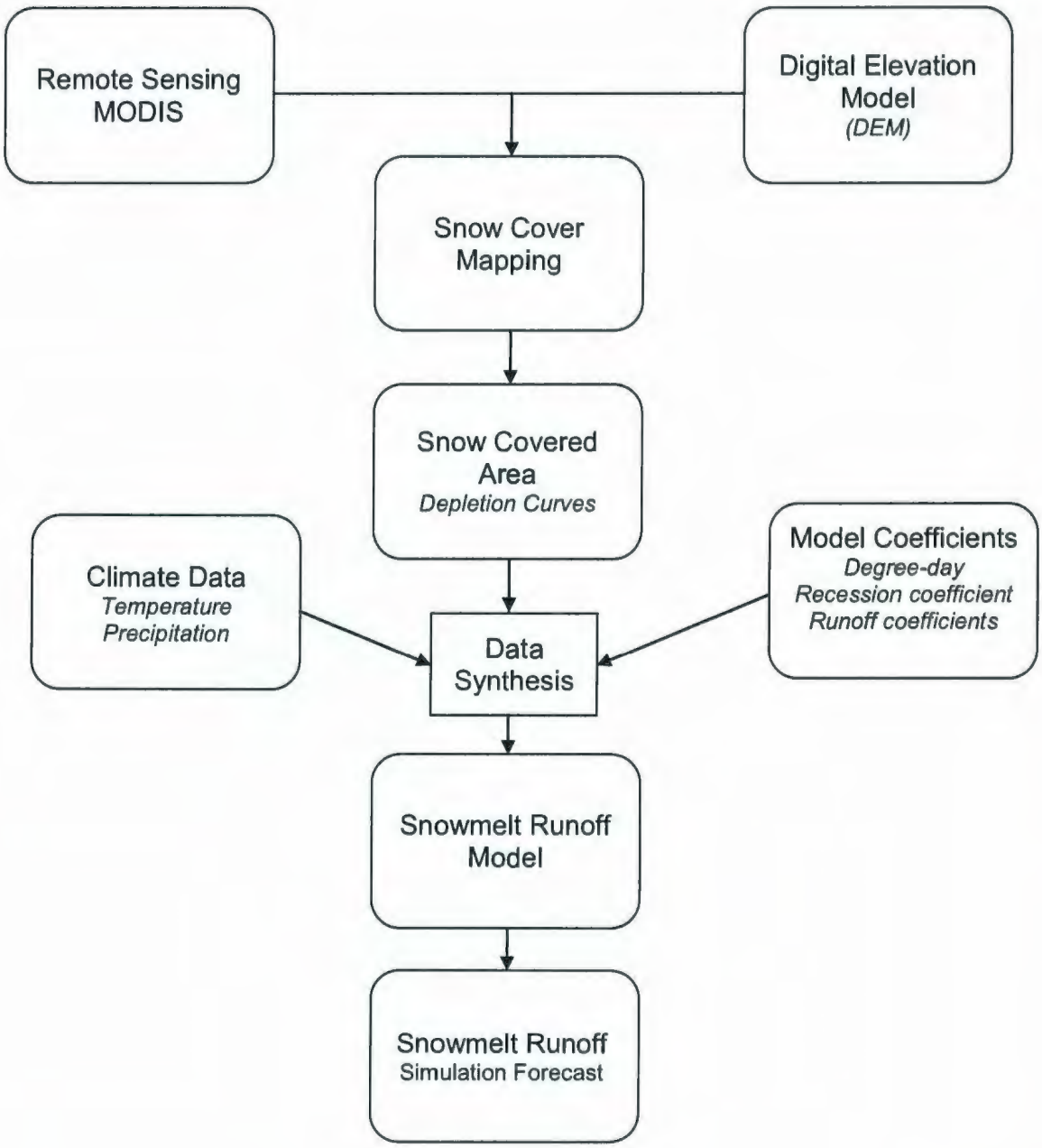


Figure 2.1: Flow Diagram Tying in Remote Sensing with Flood Forecasting

2.1 Data Collection

Remotely sensed data are very useful in monitoring the progress of snowmelt and quantifying the amount of snowmelt being added to spring runoff. Interest in remotely sensed data collection was focused on snow cover data over the Humber River Basin. The snow extent information was used as in indirect measure of snowmelt, given that snow depth cannot normally be obtained directly from visual image retrieval (VIR) imagery (Rees 2006). This section will describe how these data were obtained: mainly choosing an appropriate sensor and sensor details, the format of the raw data, the download process, and how the snow cover data are derived.

2.1.1 MODIS Sensor

MODIS (Moderate Resolution Imaging Spectroradiometer) is regarded as the optimal source for snow cover data and was the remote sensor selected to capture the images over the Humber River Basin. MODIS is an optical sensor aboard NASA's Earth Observing System (EOS) Terra and Aqua satellites. MODIS Terra images were used for this application because their snow cover data was declared by NASA to be favoured over the MODIS Aqua images (Riggs et al. 2006). The reliability for snow cover data extraction was compromised on the MODIS Aqua sensor when band six, sensor detection required for snow cover data acquisition, failed shortly after launch (Riggs et al. 2006).

The Terra satellite was launched in December 1999. The first views of Earth from MODIS were in February 2000 and data acquisition began in March 2000. The Terra

satellite has a near-polar, sun-synchronous orbital period of 98.1 minutes. Its nominal swath coverage is 2,330 km (across track), providing tile sizes of 1,200 km by 1,200 km, and a spatial resolution of 500 m for bands three to seven (i.e. pixel size = 0.25 km^2) (Riggs et al. 2006). Terra orbits the Earth in which the location that it passes over and collects only daytime data (sun-synchronous). Two main reasons MODIS Terra images were chosen over other satellite images are (1) the daily temporal data are advantageous and (2) the data are free and easily accessible. Some pertinent technical specifications of MODIS are summarized in Table 2.1.

Table 2.1: Technical Specifications of MODIS (National Aeronautics and Space Administration 1999)

Orbit:	704 km, 10:30 a.m. descending node (Terra) or 1:30 p.m. ascending node (Aqua), sun-synchronous, near-polar, circular
Scan Rate:	20.3 rpm, cross track
Swath Dimensions:	2330 km (cross track) by 10 km (along track at nadir)
Size:	1.0 x 1.6 x 1.0 m
Weight:	228.7 kg
Data Rate:	10.6 Mbps (peak daytime); 6.1 Mbps (orbital average)
Spatial Resolution:	250 m (bands 1-2) 500 m (bands 3-7) 1000 m (bands 8-36)
Design Life:	6 years

Primary Use	**Band**	**Bandwidth**[a]	**Spectral Radiance**[b]	**Required SNR**[c]
Land/Cloud/Aerosols Properties	3	459-479	35.3	243
	4	545-565	29.0	228
	5	1230-1250	5.4	74
	6	1628-1652	7.3	275
	7	2105-2155	1.0	110

[a] Bandwidth is in nm

[b] Units for Spectral Radiance = $W/m^2/\mu m/sr$.

[c] SNR = signal-to-noise ratio

Although the design life for MODIS was six years, it has been in orbit for about 10 years now and continues to operate without any irreparable problems. The bands on MODIS range from band 1 to band 36. For the purpose of this research on snow cover data, only the sections with the bands of interest were provided in Table 2.1. Snow covered land, snow covered ice on inland water, and fractional snow cover are all components that are identified or computed from the MODIS snow cover algorithm (Riggs et al. 2006). Other

primary uses for the remaining MODIS bands range from phytoplankton biogeochemistry to atmospheric water vapour (Seidel and Martinec 2004).

2.1.2 Raster Data

The remotely sensed MODIS data are stored in hierarchal data format (hdf); used to store raster data. Raster is a grid-like format, as opposed to vector, which stores its data as lines and polygons. In raster representation the area is divided into an array of rectangular (usually square) cells. Each cell is assigned properties or attributes which describes all geographic variations. The cells are sometimes called pixels (short for picture elements) (Longley et al. 2005).

The default for remotely sensed data storage is raster format and the resolution is often described by pixels. A pixel, by definition, is the smallest element in an image that can be individually processed. The size of a pixel helps describe the resolution of an image. It is also important to know that the information inside each cell is assumed to be homogeneous and there is only one classification for each pixel. The pixel size of the snow extent MODIS/Terra data is 500 m by 500 m or 0.25 km^2, which translates into approximately 1,880 pixels within the Upper Humber Basin above Black Brook region.

2.1.3 Data Download

The initial images rendered by MODIS can be considered unrefined or raw data, but the National Snow and Ice Data Center (NSIDC) snow extent MOD10A1 product is

classified as pre-processed data. Although the MOD10A1 product is processed, these images still need to be downloaded and clipped to a reasonable size in order to focus on and assess the area of interest. The MODIS sensor is used to capture a variety of products. The snow and ice product was downloaded from the NSIDC based in Boulder, Colorado. The NSIDC provides various data source links where specifics can be chosen for proper data download. The source chosen for this data download was the Warehouse Inventory Search Tool (WIST), which was used to obtain the archived data of MODIS/Terra images from 2000 to 2009 (National Aeronautics and Space Administration WIST 2010).

WIST stores the entire archived data for MODIS/Terra, MODIS/Aqua, and many other EOS (Earth Observation System) data from other instruments. WIST allows the user to search by parameter, spatial sub-setting, and tile searching for select products. One drawback of WIST is that the maximum download per order maxes out at 1000 granules (The National Snow and Ice Data Center 2008). A data granule is the smallest aggregation of data which is independently managed. The Humber River Basin is located under one granule; a data granule consisted of a per day image of snow cover (National Aeronautics and Space Administration WIST 2000).

The snow and ice product of specific interest was listed under cryosphere: MODIS/Terra Snow Cover Daily L3 Global 500 m SIN Grid V005, primary data search. This infers a download of the latest version of daily data with 500 m resolution in the gridded

sinusoidal equal area map projection. A file-transfer-protocol (FTP) was used to obtain the data. FTP is a means to exchange and manipulate files over a TCP/IP (Transmission Control Protocol/ Internet Protocol) network, for example the internet. It is often accessed by user-based passwords or anonymous user access. This method of downloading data is used when large amounts of information are being transferred. In this case, the MODIS/Terra snow and ice product was downloaded by FTP over the internet through email authentication from WIST. See Appendix A for WIST screenshots providing the sequence of steps to correct MODIS/Terra snow cover data download.

2.1.4 Re-Projection

It is important to know the projection of the downloaded data. Different projections preserve various aspects of an image. There are conformal, equidistant, and equal area projections. The choice of projection depends on the type of information required. For the purpose of snow covered area, an equal area projection was desired. The sinusoidal projection is an equal area projection and displays the proper areas equal to their corresponding areas on a globe. They do, however, distort the image of the land masses, but this is merely a visual drawback and does not affect this analysis (Longley et al. 2005).

Locations are specified using the UTM (Universal Transverse Mercator) coordinate system. It divides the Earth into a grid with 60 longitudinal zones, each with a different map projection: a specific secant transverse Mercator projection. The transverse Mercator

projection is known for its ability to map sections of large north-south extents with minor distortion. Each zone is divided into 20 latitude bands labelled by letters of the alphabet from C to X (omitting the letters I and O, because of similarities to the numbers zero and one) (Longley et al. 2005). Newfoundland's Humber River Basin is located in zone 21U. The data is collected from the cryosphere, which is the part of the Earth's surface that is covered in frozen water. The details of the data download are as follows: MODIS/Terra snow cover daily L3 global 500 m SIN Grid V005. V005 is the latest version of available data and is the most advanced in pre-processing with a much improved method for properly identifying and classifying snow and cloud (Riggs et al. 2006). The MODIS snow and ice product was searched for on the NSIDC website and downloaded using FTP Pull, which is an easy way for the user to copy files over the internet.

The MODIS data was downloaded for the 10 available years, from 2000 to 2009. The product was downloaded for every year to monitor the snow cover over the Humber River Basin, from October 1st to June 30th, but data range for analysis was only used from March 1st to June 30th, to cover the period of snowmelt. Although Newfoundland and Labrador can experience some snowmelt in October, November, and December, quantifying the spring snowmelt was of primary concern for the WRMD because of higher observed flow rates during that period. The images for MODIS Terra snow cover only began in March 2000.

2.2 Data Validation

Data validation is a step that is very important but often mistakenly regarded as peripheral. This section will first discuss how the MODIS data were reviewed for imperfections and errors in sensor readings through metadata and flagged data. Second, the method of pixel classification for this data is clarified. Lastly, this section will explain how the snow cover data accuracy was improved by setting a cloud-cover threshold.

2.2.1 Metadata

Metadata are essential when using data for an analysis and is commonly referred to as 'data about data'. It is structured information provided along with the data itself to inform users about its quality and applicability. Metadata includes information about the currency (date), processing, projections, scale, resolution, source, and contact information for further questions. Specifically, object-level metadata (OLM) provides crucial documentation which describes the contents of a single dataset. OLM allows the user to decide whether the data satisfy their requirements for analysis. It also provides information about the data which allows the user to handle it efficiently and effectively (Longley et al. 2005).

2.2.2 Flagged Data

Within the metadata file for each data set is a section for flagged data. The snow and ice product was flagged in the metadata for a few of the days, over the 10 year data collection period. The yellow flags described as "other quality", indicate a failure in the

sensors "no surface reflectance input" from the quality assessment (QA) checks. These data were removed and not used in the analysis. The QA provides an indication on the quality of data. Unless the data are unusable or missing it is often determined to be of good quality. When the majority of pixels covering the region of interest are classified as either zero (missing data) or one (no decision), the data is also removed from analysis. Missing data classification is self explanatory; this describes data that has been lost along the way and termed missing. No decision data classification is determined when the data are deemed unusable or when the sensor is unable to detect any reflectance's relevant for proper classification. The usable good quality data are input for the snow algorithm (Riggs et al. 2006).

2.2.3 Pixel Differentiation

The process used to classify pixels can never be 100% accurate because land cover is never homogeneous, at any level of detail. Regardless of the image resolution, there will always be some variation within a pixel. There is a basic assumption that the information within one pixel is the same throughout that given area. A mixed pixel or "mixel" is the term used to express a pixel whose area is divided into more than one class, which can be described as a transition zone. It is actually quite uncommon for a pixel to be completely classified as mixel-free at any resolution (Longley et al. 2005).

There are two main techniques used to classify mixels. The more common technique is to identify and assign the land class with the highest percent coverage within that pixel

area. The other technique is to discover the land cover class identified at the center of the pixel and assign that land cover class to the entire pixel (Longley et al. 2005). MODIS/Terra pixels are classified by the first technique described: the land class with the highest percent coverage. The MODIS ground resolution is not precise enough to be able to pinpoint a center pixel classification. The general reflectance combination within the ground resolution cell or pixel is determined and classified based on its most likely category of classification (i.e. snow, lake ice, inland waters, no snow, cloud, and ocean).

2.2.4 Cloud Cover Threshold

The main snow mapping obstacle for MODIS, being an optical sensor, is cloud cover. When cloud cover is high, there is a greater chance that it may be misclassified as snow and/or snow is misclassified as cloud cover. Cloud cover can be misclassified as snow and snow can be misclassified as cloud. Given that Newfoundland and Labrador is an exceptionally cloudy province, the snow cover derived MODIS images of the Humber River Basin can often be influenced when percent cloud cover over the basin is high. It has been discerned, from the MODIS cloud cover data, that cloud cover over the Humber River Basin is over 50% cloud cover for 80% of the days on average from March to June every year for the past 10 years (2000 to 2009).

This snow and cloud misclassification problem persists in the snow algorithm. The technical reasoning behind this misclassification is associated with parts of ice clouds appearing yellow in MODIS bands one, four, and six color display (bands four and six

being pertinent to snow classification). This error occurs when parts of the clouds are in the shadows from other clouds. These arrangements lead to parts of the cloud not able to be picked up as cloud when the cloud mask is generated because of different reflectance levels. These missed clouds are then processed in the snow algorithm and often have spectral features closer to "snow" than "not snow". According to the most recent MODIS Snow Products user guide, this problem is typically very small due to a great deal of improvements having been recently implemented (Riggs et al. 2006).

As stated, the Humber River Basin experiences well above average cloud cover compared to other areas of the world. This large portion of cloud cover was still a concern for data collection, despite the MODIS Snow Products user guide reassurance. This cloud cover was a concern because of possible affects on either the sample size of acquired data or the accuracy of the snow cover data. It was therefore important to assess and implement a cloud cover threshold to reduce the possibility of misclassified snow and/or cloud. Although MODIS provided daily images, very few remained after the cloud cover threshold was realized.

Snow depletion curves, plotting percent snow cover over time, were created to set the cloud cover threshold. These were created for each cloud cover threshold being tested and only included snow cover points with cloud cover less than or equal to the particular specified cloud cover threshold limit. It was important to observe little change in the rapidity and date of decline from curve to curve as the tested percent of acceptable cloud

cover increased. These criteria are essential to avoid significantly altering the correctly classified data. Table 2.2 summarizes the various cloud cover thresholds tested along with the number of data points it provided, on average per year, over 10 years for the snowmelt time periods. See Appendix B for plot comparisons of the data summarized in Table 2.2.

Table 2.2: Percent Cloud Cover and Coinciding Number of Snow Cover Data Points

Cloud Cover (%)	**Approx. Number of Snow Cover Data Points**	**Change in Rapidity and Date of Decline**
5	5	-
10	7	Minimal change
20	10	Minimal change
30	13	Noticeable difference

It is crucial to find a good balance between cloud cover and number of snow cover data points per year. Finding this balance is similar to choosing a filter size. On one side, the cloud cover becomes too high and the number of snow cover data points increase, but the accuracy of these points decreases because it may not be classifying the cloud and snow properly. On the other side, as the cloud cover threshold decreases, the number of snow cover data points diminishes rapidly. This affects the statistical integrity of the assessment because as the number of data points decreases, more of the daily snow cover must be interpolated (in other words more of the daily snow cover points must be estimated). It is important to ensure that the data are independent and random; otherwise

the data will not accurately portray reality. A cloud cover threshold of 20% was found to provide a proper balance between number of data points and data validity. The 20% cloud cover threshold was determined based on the plotted snow cover data sets over the past 10 years (2000 to 2009). For flood forecasting in the future, this cloud cover threshold should be re-assessed with appropriate snow cover and cloud cover data to ensure accurate modeling.

2.3 Data Manipulation

The MODIS images downloaded from the NSIDC constitute pre-processed data. First, this data is manipulated to derive the snow cover maps through radiometric and geometric corrections, re-projections, and multispectral classification. Second, the robust image processing program, PCI Geomatica version 10, is used to compile and calculate percent snow cover, while the daily output files are handled in Microsoft Excel using a Visual Basic (VB) program. Finally, snow cover depletion curves are derived based on a specified cloud cover threshold which is set to eliminate possibly skewed and misclassified snow cover data.

2.3.1 Deriving Snow Cover Maps

There are three steps which lead-up to the derivation of snow cover maps from remotely sensed data:

1. pre-processing,
2. multispectral image classification, and

3. integration of interpreted results.

2.3.1.1 Pre-Processing

Metadata is an important part of data processing as it explains where this data is from and how they have been manipulated. There are two steps in pre-processing: radiometric and geometric corrections. Radiometric correction is essential to compensate atmospheric distortions. This provides a clearer visual of the Earth's features and leads to a more reliable and robust interpretation of the data. Geometric correction or geocoding is a process in which all raw data is transformed in various ways to ensure that they all belong to the same georeference system. The standard georeference system varies from country to country (Seidel and Martinec 2004). This NSIDC pre-processing is efficient, making the data available within days of capture.

2.3.1.2 Multispectral Image Classification

MODIS is an optical sensor. This means that it uses the visible and infrared spectrums to generate images of the Earth's surface. It does this by detecting solar radiation from targets on the ground. These targets are differentiated by their spectral reflectance (Seidel and Martinec 2004). See Figure 2.2 for a visual on the visible and infrared (IR) portions of the electromagnetic spectrum with respect to the other parts of the spectrum.

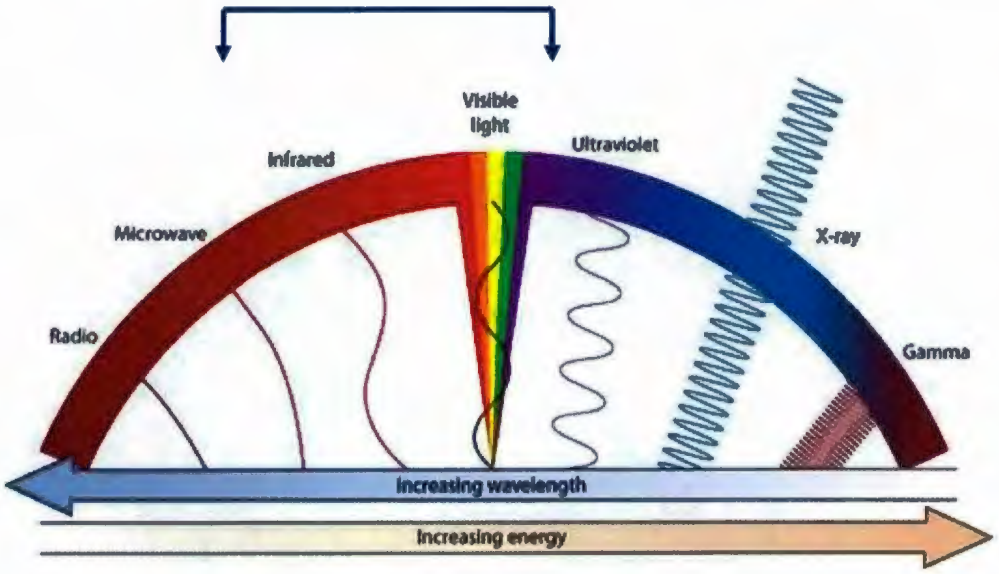


Figure 2.2: Electromagnetic Spectrum with Emphasis on the Visible and Infrared Sections (Science Learn 2007)

Deriving snow cover maps from MODIS data is based on a method developed by NASA (National Aeronautics and Space Administration) using the normalized difference snow index (NDSI). This is the difference between IR reflectance of snow in visible and shortwave wavelengths. Terra uses bands four and six for snow mapping. MODIS Aqua band six (1.6 μm) detectors failed after launch, leaving it only about 30% functional; 70% of the band six detectors became non-functional. This is why the snow cover data are compromised. Aqua now uses MODIS band seven (2.1 μm) for the NDSI calculations (Riggs et al. 2006). The NDSI is not affected by the wide range of illumination settings and it does not rely on the reflectance of a single band. (Seidel and Martinec 2004).The NDSI calculations are as follows, see Equation [2.1].

[2.1] $NDSI = \frac{band\ 6 - band4}{band\ 6 + band\ 4}$

Where:

band 4 = green band reflectance; and
band 6 = shortwave IR reflectance.

The NDSI allows the differentiation between snow and many other land cover types by observing the strong reflectance of snow in the visible bands (e.g. band four) and the strong absorption of snow in shortwave IR (e.g. band six) (Abbott 2009).

To create a snow cover map from remotely sensed data, the NDSI technique is implemented to identify and classify snow on a pixel-by-pixel basis. Other spectral threshold tests are used in conjunction with the NDSI test to identify other types of land cover. The NDSI method is useful for numerous reasons. The two principal reasons are: (1) it is easier to detect snow and ice in the visible region because it is considerably more reflective in the visible region than in the shortwave IR region; and (2) it can be considered a great snow/cloud discriminator because the reflectance in the shortwave IR region of *most* clouds remains high, while the reflectance of snow is low (Seidel and Martinec 2004). There is, however, one type of cloud, which remains difficult for optical sensors to differentiate from snow and that is the thin cirrus cloud (Rees 2006).

The Humber River Basin is classified as having a dense forest canopy with over 75% coverage with coniferous trees, specifically Black Spruce (Water Resources Management

Division 2009). Snow cover mapping is frequently hindered when a pixel is partially or fully covered by dense forest cover (i.e. snow cover remains unnoticed). Snow that falls on a coniferous tree canopy does not often remain there for the entire winter as it can often disappear due to sublimation. The snow on the ground below, however, will most likely remain unaccounted. Measuring reflectance specifically the NDSI and NDVI (Normalized Difference Vegetation Index) together can often provide a strong signal used to exploit and classify snow covered forests (Seidel and Martinec 2004). Still, for the Upper Humber, the NDVI was not used in conjunction with the NDSI. Given that the plotted conventional depletion curves showed no significant changes or abnormalities throughout any of the 10 plotted snowmelt seasons, it is assumed that the snow mapping is not greatly hindered by the dense forest canopy.

The snow cover algorithm screens each pixel for temperature, before a conclusive snow decision is made from calculating the difference in bands ratio. This ensures that the classification makes logical sense. Any pixel classified as snow with an estimated temperature greater than 283 K (or 10°C) is changed to land. This extra step has proven useful in reducing the occurrence of erroneous snow identification in some situations, but often only along warm coastal regions with wide, sandy beaches. The proper location and alignment of snowy coastlines in Canada have been problematic in the past, but improvements have been realized when the land/water mask was implemented. This has reduced erroneous snow mapping along coastlines and coastal differences remain a minor problem (Riggs et al. 2006).

The snow algorithm is approximately 93% to 100% accurate at mapping snow under ideal illumination conditions (i.e. clear skies and several centimetres of snow on a smooth surface). Ideal conditions are rare in any part of the world, and never the case in the Humber Valley, NL regarding completely clear skies and a smooth surface, but the snowy region often will accumulate several centimetres of snow over the winter season. The NDSI has proven to be a robust indicator of snow when snow is present, although patchy snow or thin snow cover on vegetated surfaces may be missed by the NDSI (Riggs et al. 2006).

2.3.1.3 Integration of Interpreted Results

This third step in deriving the snow cover maps is used to manage and display results, normally through use of a geographic information system (GIS). A GIS is a broad term which encompasses a large range of applications. These applications commonly fulfill the five M's of GIS: mapping, measurement, monitoring, modeling, and management (Longley et al. 2005). With technology advances, using technical tools for multispectral image analysis and GIS's for managing and storing large databases, processing remotely sensed information is becoming less expensive. Furthermore, Earth Observation (EO) data with practical time steps and ground resolution are becoming available in steadily increasing numbers (Seidel and Martinec 2004).

Before this step can be completed, the data should be assessed based on quality and any uncertain data should be removed. The section on data validation addressed this process which includes reviewing metadata files and assessing the flagged data.

2.3.2 PCI Geomatica – Post-Processing

Many steps are carried out to extract the desired snow cover information from the raw satellite data. Snow cover analysis requires a number of steps to process the data from the sensors. These sensors are unable to capture adequate information to classify the objects in a single step. While there are many steps, the main idea is to develop an algorithm used to extract the snow covered area over long periods of time. This is usually accomplished by counting the number of snow covered units (SCUs) over the given area (Seidel and Martinec 2004).

PCI Geomatica is a powerful integrated software system with many applications used for remote sensing data and image processing. Geomatica FOCUS is an application used for viewing, enhancing, and examining remotely sensed imagery. This application (version 10) was used for the analysis along with EASI modeling, which was applied within FOCUS for processing the remotely sensed data using a written script.

For this thesis, the downloaded MODIS images were clipped to the area of interest and assessed for percent snow cover over the Humber River Basin for all of its 12 sub-watersheds. The EASI snow algorithm was implemented and used for the extraction of

daily snow cover data. Each processed day was exported into an individual text file containing information on the area/pixel count of particular classes: no snow, lake, cloud, lake ice, and snow. Problem pixels were classified under missing data and no decision. Each land cover classification is linked to a pixel number. Table 2.3 provides a legend of significant pixels and land classification for proper snow cover mapping of western Newfoundland's Humber River catchment area.

Table 2.3: Legend for Land Classifications and Corresponding Pixel Identification Numbers (Riggs et al. 2006)

Pixel	Land Classification
0	Missing Data
1	No Decision
25	No Snow
37	Lake
39	Ocean
50	Cloud
100	Lake Ice
200	Snow
254	Detector Saturated
255	Fill (data used to fill gaps in the swatch)

Ocean identified pixels are not analyzed for snow. Inland waters, lakes, and rivers, however, are assessed for possible snow covered ice conditions. A snow/no-snow decision is made on the MODIS swath data if all of the following three criteria are met: (1) data are classified as either land or inland water, (2) data are captured in daylight, and

(3) cloud mask is applied (Riggs et al. 2006). Figure 2.3 provides MODIS/Terra processed data displaying various land classifications, specifically snow cover. In Figure 2.3, the Upper Humber watershed is outlined in red, snow and lake ice indicated by white, clouds by grey, and ocean and inland lakes by blue.

Figure 2.3: Processed MODIS/Terra Snow Covered Area for the Upper Humber Basin, NL

The AREAREPORT program in PCI Geomatica version 10 is used to generate a snow area report from the MODIS/Terra data set. The two inputs required to run this algorithm are an input raster (MODIS/Terra images) and a bitmap mask (watershed boundaries). The reporting units were set to km^2. Figure 2.4 is an example of the AREAREPORT text file output for the Upper Humber above Black Brook sub-watershed on 2002-102 (this

implies the date the information was obtained: year 2002 and julian date 102, being April 12, the 102nd day in the year starting January 1).

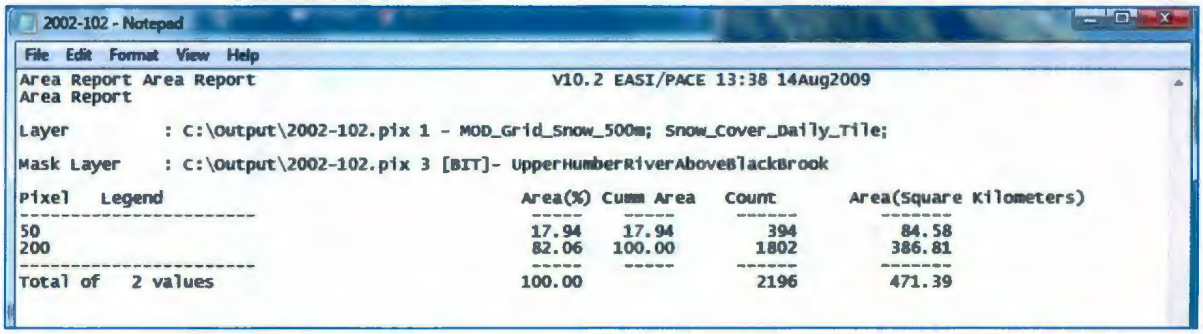


Figure 2.4: Example of an AREAREPORT Daily Text File Output

The number of different pixel values reported in an individual text file varied throughout the 10 years of data. This is because only the land classes that were classified within the specified region (approx. 470 km^2 with 1,880 pixels), on that given day, are reported. For example, on April 12, 2002 only two types of land classes were reported: cloud cover (pixel 50) and snow cover (pixel 200). The following land classes were considered for this research: snow cover, cloud cover, snow and lake ice, and inland lakes. For the Upper Humber region, however, there were no inland lakes and therefore no lake ice to consider for flow modeling.

A program was written in EASI script to automate the extraction of snow cover information, see Appendix C for details. The process of extracting snow cover information from the MODIS imagery followed these steps:

1. import all hdf files, rename files, save in pix format (i.e., 2008-306.pix),
2. clip hdf files to cover only the island of Newfoundland, and
3. append all watershed bitmap files to the pix file and run AREAREPORT program (this creates daily individual text files with pixel classification information).

2.3.3 Visual Basic

Visual Basic (VB), a programming language used in Microsoft Windows, was used to write a program that could pick out the percentages of snow cover and cloud cover for each day and amalgamate the individual text files to import them into MS Excel. This made the data easier to view, manipulate, and combine with meteorological data for analysis. See Appendix D for the VB script written to import, combine, and manage all daily individual output text files from the snow cover data extraction into one Excel spreadsheet. These imported data were further manipulated in Excel for data analysis.

2.3.4 Snow Cover Depletion Curves

Snow cover depletion curves are useful plots which can easily and accurately depict how a snowpack melts over its seasonal snowmelt period. These snow depletion curves are the final product developed from the satellite images. It is typical to observe a gradual decrease in snow covered area over the seasonal melt period. Snow covered area, however, is not a clear measure for snow reserves in terms of water equivalent (Seidel and Martinec 2004) and it is complicated by spatial distribution of slope, aspect, and type of forest cover (Watt 1989).

Snow depletion curves are typically reverse s-shaped: defined as being steep in the middle and flat on both ends. The reason behind the reverse s-shape is that the frequency distribution of the snow depths follows this form and snowmelt starts at lower elevations of the basin, progresses across the medium elevations, and finishes at the upper parts due to temperature lapse rate.

Snow cover depletion curves are never completely smooth between measured points. These blips in the curve are caused by climate irregularities during the snowmelt season. During periods of extremely cold temperatures the snowmelt decline is temporarily halted, whereas during periods of exceptionally warm temperatures the decline is steeper (Seidel and Martinec 2004). Of course, the more frequently this snow cover data is obtained, the more accurate the decline is plotted with smaller variations.

There are three basic types of depletion curves originally defined by Hall and Martinec (1985). First, the conventional depletion curve (CDC), type I, plots 'Snow Covered Area' vs. 'Time'. While this is the simplest curve to plot, since the data is the most easily attainable, it also provides the least amount of information as the rapidity of decline is solely based on initial snow water equivalent (SWE). SWE is the amount of water in a snowpack based on density. Second, the modified depletion curve I (MDCI), type II, plots 'Snow Covered Area' vs. 'Cumulative Degree-Days'. This claims to improve the observation of seasonal snowmelt because the rapidity of decline is based on both the initial SWE and temperature conditions. This adaptation eliminates the effect of

temperature differences from year to year. Third, the modified depletion curve II (MDCII), type III, plots 'Snow Covered Area' vs. 'Cumulative Snowmelt Depth'. This depletion curve (DC) provides the best and most accurate information on snowmelt out of the three plots. The rapidity of decline is based on the actual volume of water, provided that all of the snowpack were to melt. It offers information on the likelihood that the degree-day factor alters throughout the season (Rango and van Katwijk 1990). There are secondary Type II and Type III depletion curves which take into account the melting of new snow fallen during the snowmelt period.

Type I and Type II depletion curves were derived for the 10 years (2000 to 2009) of snow cover data over the Upper Humber River basin (see Appendix E). At this time, no snow water equivalent data was available. Type III curves were not plotted. Figure 2.5 shows typical Type I snow depletion curves at the 20% cloud cover threshold for 2002 and 2005.

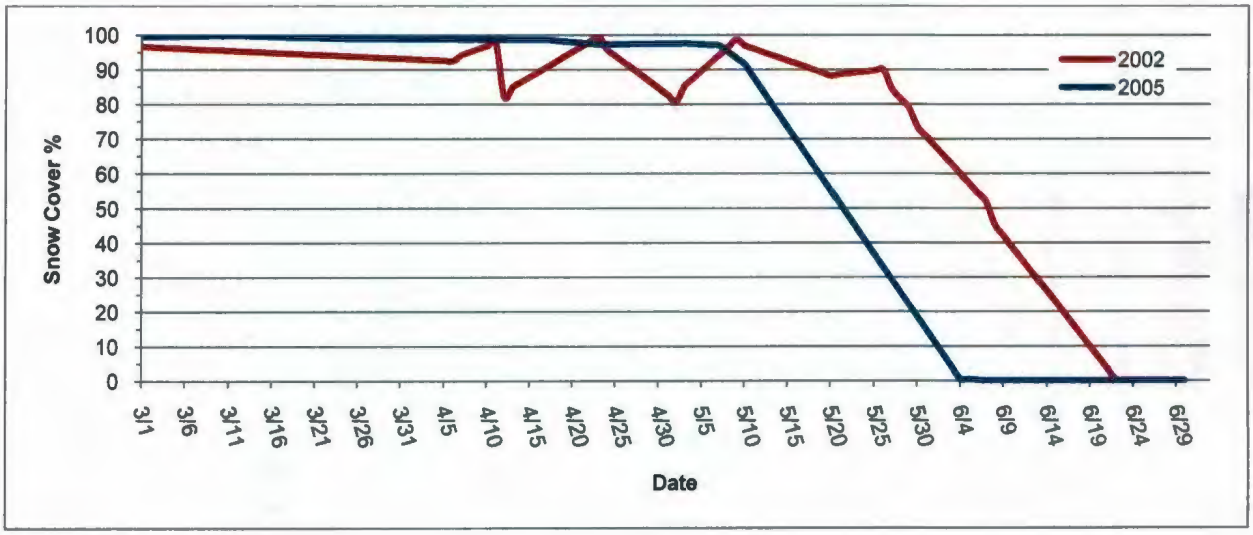


Figure 2.5: Typical Type I Conventional Snow Depletion Curves at 20% Cloud Cover

The average snowmelt period over the 10 years of data began mid-May and ended mid-June, lasting an entire month. The MODIS/Terra images in Figure 2.6 provide a better visual of the snow ablation period over the Upper Humber region lasting on average approximately 30 days.

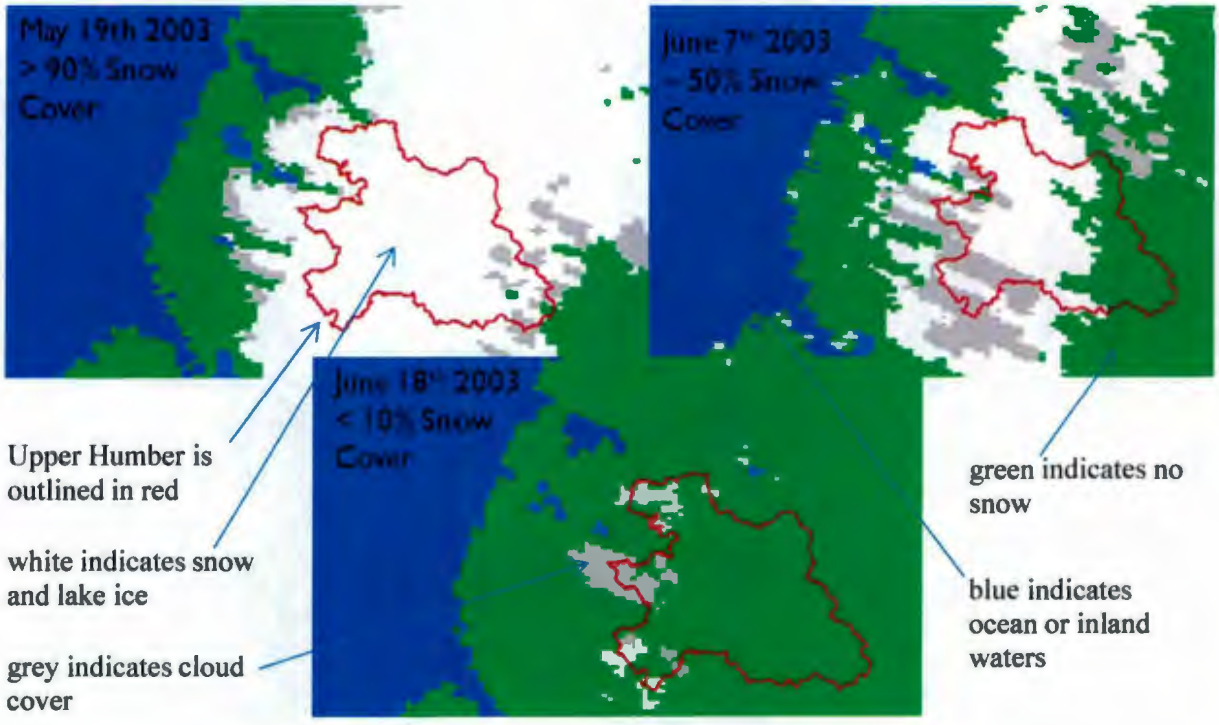


Figure 2.6: Upper Humber Basin above Black Brook Snowmelt Period for 2003 Depicted by Processed MODIS/Terra Images using PCI Geomatica

The snow depletion curves are not naturally smooth. The sharp peaks and edges are created because of the low number of data points (approx. 10/year), from implementing the 20% cloud cover threshold. There were only approximately 10 data points over a 90 days period for each year. To obtain the percent of daily area snow coverage, linear interpolation was used between known values to fill in the gaps. Certainly, the instability and uncertainty lies in between the data points. These time intervals between the points can create an inaccuracy for shape of the depletion curves and a single point can skew the understanding of the satellite images (Seidel and Martinec 2004). It is vital, for real-time runoff forecasts, to obtain the satellite snow cover data within days after a satellite

overflight and also to extrapolate the depletion curves of the snow coverage to the future weeks (DeWalle and Rango 2008).

This chapter has explained every detail of remotely sensed snow cover data required for its implementation into a snowmelt runoff model for daily flow modeling. This entailed: selecting the proper sensor; advantages and disadvantages of snow cover data for the Humber River, NL; data collection, validation, and management; and snow cover depletion curves. Although the snow cover DCs are susceptible to many accuracy pitfalls, they provide a sufficient estimate of snow cover in an area where no significant amount of snow data has been archived. These curves also provide an estimate on the percent snowmelt over time.

- Chapter 3 -

Daily Flow Modeling

The purpose of daily flow modeling is to be able to forecast the next day's flow rates in the water body being analyzed. These daily forecasted flow rates enable one to prepare for and manage possible flooding in populated areas. The predicted daily flows can also help water management for hydropower companies. This includes playing a role in hydropower generation and sales.

The use of snow to help predict daily flow rates is considered a more complex addition to rainfall runoff models. The lag between when it falls and when it produces runoff and groundwater recharge is the differentiating factor between how snow and rain are treated in hydrology (Maidment 1993).

3.1 Choosing a Snowmelt Runoff Model (SRM)

Choosing a rainfall runoff model that incorporates snowmelt was an important task. Many factors were considered. The general criteria for choosing a runoff model are: (1) reliability, (2) ease of use (including input data requirements and data availability), (3) performance and accuracy of results, (4) characteristics of study watershed (most

important being basin relief), and (5) cost of setting-up and running model (Watt 1989). It was important to identify the purpose of the model in terms of desired output and available input data. An SRM is essentially made up of two parts (1) calculating the amount of snowmelt and rainfall and (2) converting these numbers into runoff.

Streamflow predictions are based on two groups of key terms used to describe rainfall inputs. The first group contains the water storage terms. This group encompasses interception, soil moisture, and surface storage. The second group are the flux terms. This group includes infiltration, evapotranspiration, snowmelt, interflow, groundwater baseflow, and surface runoff from rainfall and snowmelt. These terms can all affect streamflows and can each be used at varying levels of complexity (Maidment 1993).

Of course snowmelt is the most important additional direct input required for this model. Background information on snow is essential to understand before moving forward in choosing a model. Snow is a form of precipitation made of falling or deposited ice particles and is often formed from the freezing of the water vapour in the air. For modeling, the focus is on snow cover rather than falling snow. The model must also look at temporary and seasonal snow cover (lasting several months). This snow should not last throughout the summer, hence the terms: temporary and seasonal. Snow cover represents an important geophysical variable for climate, especially in affecting the ground's albedo effect caused by its strong reflection properties (Rees 2006). Albedo is a measure of how

strong light is reflected from light sources, like the sun. It is a specific form of reflectivity.

The type of forecasting and how often the model is updated are other factors considered in choosing an appropriate model. Forecasting can be used for estimating conditions at a specific future date or during a particular time period. It is also used to predict the occurrence of extreme events (floods and droughts), to operate water resource systems, or to negotiate contracts in hydropower sales. The frequency of updating the model was also important to consider because often as the forecasting "lead time" increases, the forecasting accuracy decreases. Of course, a model with the ability for real-time uplink would be ideal.

SRM's can be divided into two main methods: (1) the energy-budget method and (2) the degree-day method. The energy-budget method is considered a complex water balance with many parameters for physical model representation. This method uses conservation of energy to a fixed volume. The sum of the energy fluxes by radiation, convection, conduction, and advection in addition to the change in internal energy in the volume yields zero change and all energy is in check, see Equation [3.1]. Simply, the energy-budget method estimates the amount of energy available for snowmelt (Maidment 1993).

[3.1] $Q_m = Q_n + Q_h + Q_e + Q_g + Q_a - \frac{\Delta U}{\Delta t}$

Where:

Q_m = energy available for melt;
Q_n = net radiation (flux of energy at the surface due to exchange of radiation);
Q_h = sensible energy (flux of energy at the surface due to the difference in temperature between the surface and overlying air);
Q_e = latent energy (flux of energy exchanged from vapour movement at the surface from the difference in vapour pressure between the surface and overlying air);
Q_g = ground heat (flux of energy exchanged by conduction);
Q_a = advective energy (energy derived from external sources, i.e. rain); and
$\Delta U/\Delta t$ = rate of change of internal energy over time (Maidment 1993).

A method such as the energy-budget requires many detailed inputs. Problems may arise with detailed input series such as: (1) dew point, (2) wind speed, and (3) solar radiation; given that numerous continuous simulation models need testing over an extensive time period to 'warm-up' (i.e., validate) the model. This 'warm-up' ensures that initial parameter settings are correct (Watt 1989). Often there are just not enough resources available to obtain all inputs with enough accuracy and/or the time period of available data is simply not long enough for proper model start-up. Additionally, although the energy-budget methods can provide a solid understanding of all variables involved in flood forecasting, they may not all be significant for the specific study region.

The second method for runoff models is the degree-day method. It is a more basic model fundamentally based on temperature index methods. Temperature index methods do not incorporate a complex or even adequate physical description of the melt process as the

energy balance methods do, but the temperature based models can still yield practical results. Instead, the temperature index method relates snowmelt to air temperatures. As temperature increases, the volume of snowmelt increases. Equation [3.2] provides the most common calculation used to relate snowmelt to temperature, referred to as the degree-day method (Maidment 1993).

[3.2] $M = M_f(T_i - T_b)$

Where:

M = depth of meltwater produced over the given interval of time [mm/day];
M_f = melt factor [mm/day°C)];
T_i = index air temperature (often an average of the interval of time) [°C]; and
T_b = base temperature (often set to 0 °C) [°C].

Air temperature is a realistic index for heavily forested areas, such as the Humber Valley. Forest canopy lessens major fluctuations in parameters such as wind velocity and longwave radiation exchange (low energy radiation entering and leaving the Earth). It also reduces the significance of shortwave radiation (which is radiant energy in certain wavelengths: energy given off by the sun) (Watt 1989).

After careful consideration, Martinec and Rango's SRM (Martinec et al. 1983 and 2008) was chosen. The two key reasons for choosing this simple degree-day SMR model were: (1) it is a degree-day method specifically useful when incorporating snowmelt into the balance and (2) it uses short-term forecasting and can be updated daily. Many other

factors were significant in choosing the best model for this research: input parameters, ease of use, robustness, output quality, reliability, simplicity, effectiveness, cost, and modeling with a daily time-step.

3.2 Martinec and Rango's Snowmelt Runoff Model

The snowmelt runoff model (SRM) was originally developed by Martinec (1975) (Martinec 1975). It is a well-known model used to predict next day flow rates. This short-duration forecast model was specifically developed to predict snowmelt runoff. Over the years, the SRM has been modified and improved in collaboration with Al Rango (NASA). The most recent update was in 1998, version 4.0. This hydrological model has been applied to many mountainous terrains, where the basin is subdivided into elevation zones (DeWalle and Rango 2008).

Martinec and Rango's SRM is one of the first and still most widely used hydrological models which incorporates satellite snow cover mapping as a direct input variable. Various other models consider satellite snow cover data but as non-binding auxiliary information. It also has modest input variable requirements and the degree-day model is preferred for dense forest canopy coverage. SRM performance can worsen when air temperature and precipitation data are forecasted too far (i.e., weeks or months) in advance and deviate from the observed values (DeWalle and Rango 2008). Of course, periodic updating will reduce these inaccuracies. For this analysis and for the best possible results the model is updated daily.

Martinec and Rango's SRM has also been applied to numerous basins of varying characteristics with acceptable results. In 1979 the World Meteorological Organization (WMO) performed a comparison of 11 snowmelt runoff models on six different basins for a 10 year period. This model was tested across various geographical regions and various basin sizes ranging from an area of 10 km^2 to 2200 km^2. Differences in elevation also ranged from 350 m to 3500 m. Computed daily runoff values were compared with the measured values. Model performance was based on two especially informative criteria: coefficient of determination, R^2 and volume deviation, D_v. Based on WMO's test, Martinec and Rango's SRM best represents remote sensing in snow hydrology. The model only requires six parameters and was at least as accurate as the CEQUEAU model (developed at the University of Quebec), which requires 31 parameters (Seidel and Martinec 2004).

The Rango-Martinec model is an ideal model as there are no set limits with regard to basin size and elevation range. Basin elevation is a significant characteristic for flood forecasting, especially when predicting snowmelt rates. A basin with high relief is often divided into elevation zones for separate analysis. These zones are assigned using a digital elevation model (DEM). Basins of relatively low relief are considered as a single unit with only one elevation zone (Watt 1989). For Martinec and Rango's SRM it is recommended that the basin be divided into elevation zones if the elevation range of the basin is $\geq$ 500 m (DeWalle and Rango 2008). The Upper Humber Basin only ranges from

600 m to 800 m, an elevation difference of 200 m, so this basin can be treated as a single unit.

3.3 Background Information: Region of Interest

The Upper Humber River basin, as measured above Black Brook, has an area of approximately 470 km^2. This watershed has 200 m change in elevation (much less than 500 m) so this basin was not subdivided into elevation zones. As a result, the temperature lapse rate was not required. The lapse rate accounts for the decrease in temperature with an increase in elevation. Again, the main objective from Chapter 2 was to use the percent snow cover data extracted from the remotely sensed images and display them as conventional snow depletion curves (CDCs) for every snowmelt season from 2000 to 2009. Snow depletion plots can be viewed in Appendix E.

For each snowmelt season, lasting approximately 122 days, there were on average 10 snow data points considered valid from quality assessment and cloud cover threshold. For simplicity, the CDC's were derived using linear interpolation. As discussed in Chapter 2, there are three types of snow depletion curves introduced by Hall and Martinec (1985). First, Type I is the conventional depletion curve (CDC). It is plotted as the percent snow covered area vs. time elapsed. The second snow depletion curve, Type II-A is the modified depletion curve (MDC). The percent snow covered area is plotted against cumulative degree days. There is a Type II-B curve that takes into account the melting of newly fallen snow by subtracting the degree days required to melt the new snow from the

cumulative degree days. The third depletion curve, Type III-A is named the second modified depletion curve (MDCII). For this curve, percent snow covered area is plotted against cumulative snow depth. This is the most refined and accurate depletion curve as it can assess the likelihood that the degree-day factor alters throughout the season. This variation of the degree-day factor over the snowmelt season is due to change in the density and albedo of the snowpack. The Type III-B depletion curve accounts for new snow fall during the melt season, very similar to the Type II-B curve, but the Type III-B curve plots against cumulative snow depth as opposed to cumulative degree-days (Rango and van Katwijk 1990).

3.4 Input Parameters

This section explains how the SRM calculates the daily flows and the required input parameters. This model consists of two main terms. Simply, the first term is a portion of the flow from yesterday (Q_n) and the second term consists of additional water from precipitation and predicted snowmelt. Today's flow (Q_{n+1}) is calculated by Equation [3.3].

$$[3.3]\ Q_{n+1} = kQ_n + (1-k)[c_s a_n (T_n + \Delta T_n) S_n + c_r P_n] \frac{A \times 1000}{86400}$$

Where:

Q = average daily flow [m^3/s];
k = recession coefficient [$(m^3/s)/(m^3/s)$];
c = runoff coefficient [dimensionless];
c_s refers to snow runoff coefficient [dimensionless];
c_r refers to rain runoff coefficient [dimensionless];
a = degree-day factor [$mm°C^{-1}d^{-1}$];
T = degree-days [°C*d];
ΔT = correction by lapse rate [°C*d];
S = fraction of snow covered area [fraction];
P = precipitation [mm];
A = area of basin [km^2]; and
n = number of days [d].

The first term in the model is the term that takes into account autocorrelation between flows in adjacent time periods. The recession coefficient, k, can be considered an autoregressive coefficient. This means that for larger watersheds the flows in adjacent time periods would be highly autocorrelated, whereas for smaller watersheds the flows would be more sensitive to the daily climate conditions. Flows that are influenced more by the daily meteorological conditions would make the flows in adjacent time periods less significant (McCuen 1998).

The second term in the model represents the additional water from precipitation and predicted snowmelt based on temperature and area of the basin. Logically, the factor (1-k) works in balance with the k factor in the first term: it increases the second term when

the autoregressive coefficient, k, is small and vice-versa. Breaking down the second term: the precipitation, P, is multiplied by the rain runoff coefficient (the fraction of rainfall contributing to runoff), c_r, and the watershed area, A, to determine the amount of rainwater being added to the flow prediction during a specific forecast period. The predicted melt from the snowpack is generated from the product $c_s aTS$, again multiplied by the watershed area, A. The two runoff coefficients for rain and snow, respectively (c_r and c_s), are used to assess the amount of precipitation that contributes directly to runoff, whereas the rest of the water may evaporate or be absorbed into the ground (McCuen 1998). The last bit of the equation is a unit conversion from millimetre-square kilometers per day to cubic meters per second.

A critical temperature threshold must be set. T_{crit} is used for precipitation differentiation between rain and snow (i.e. rain when $T \geq T_{crit}$ and new snow when $T < T_{crit}$). When precipitation is classified as rainfall the contribution is immediate, whereas snow has a delayed effect on runoff since its conversion to melt water takes time. Often T_{crit} is set to 0 °C (any temperature equal to or above 0 °C is classified as rainfall and any temperature below 0 °C is classified as snowfall). Provided that new snow falls over a previously snow covered area, it is assumed to become part of the seasonal snowpack. This means that the effect of the new snow is included in the derived snow depletion curve (DeWalle and Rango 2008). New snow over the Upper Humber Basin was assumed to fall over previously snow covered area as it is a highly snow covered area and the analysis always took place well beyond snow accumulation periods (starting March 1st).

Degree-days, DD, in Martinec's SRM model were simply calculated as the expression in Equation [3.4].

[3.4] $DD_n = \sum T_n, when\ T > 0\ ℃$

Where:

DD_n = total degree-days up to n days from appropriate starting point; and
T_n = average daily temperatures (°C) on n^{th} day.

In this case, the cumulative degree-days measure the heating of the snowpack. It is assumed that as the number of degree-days increases, so does the amount of snowmelted from the snowpack (leading to increased runoff). Starting March 1^{st} of every snowmelt season, the average daily temperature was observed and if it exceeded 0 °C (set base temperature, T_{crit}) then it was included in the degree-days for melting the snowpack, the degree-days were cumulative every day forward until June 30^{th} of the given season. Although it is possible that for any given year the temperature may rise above 0 °C before March 1^{st}, it is unlikely in Newfoundland and Labrador. Even if the temperature did rise above 0 °C before March 1^{st}, it is likely an irregularity and would not contribute to any quantifiable increase in runoff volume due to snowmelt. Starting March 1^{st} was also convenient for a controlled method of calculating degree-days over the past 10 years since MODIS snow cover data was provided starting March 1^{st} 2000.

The SRM parameters described from Equation [3.3] can be predetermined in four ways: (1) actual measurements, (2) hydrological judgements on basin characteristics, (3) theoretical relations, or (4) empirical regression relationships (Martinec and Rango 1986). Of course, these parameter ranges must make physical sense and remain within their acceptable ranges. For example, the runoff coefficients (c_r and c_s) should not exceed 1.0. The critical temperature for determining whether precipitation is classified as rain or snow should not be less than 0°C. The degree-day factor (a) should fall within the range of values recommended for similar basin conditions. For example, the degree-day factor includes a radiation component and therefore higher values are expected in the Himalayas and lower values in Scandinavia (northern Europe) (Seidel and Martinec 2004). The density of melting snow usually ranges from 0.3 to 0.55 g/cm^3 and so the degree-day factors often end up ranging between 3.5 to 6 mm/°C/day. There are always exceptions to the rule, the Upper Humber experiences snow cover under a forest canopy, so lower 'a' values are expected. Low snow densities correspond to low degree-day factors (Martinec and Rango 1986). Furthermore, 'a' has seasonality effects and is expected to increase over the melt season concurrent with increasing snow density and decreasing albedo (as the snow becomes 'older' and 'dirtier'). The recession coefficient (k) expresses the losses and requires hydrological judgement and analysis from past discharge data (Seidel and Martinec 2004). This autoregressive input parameter has a large impact on snowmelt runoff computations and can range from about 0.4 to 0.95 (Martinec and Rango 1986), indicating a next day flow rate prediction influenced up to as much as 95% from the previous day's flow.

3.4.1 Daily Flow, Temperature, and Precipitation Data

The daily flow, temperature, and precipitation data are all retrieved from the Government of Newfoundland and Labrador via the WRMD Hydrologic Modelling sector. Considering the satellite snow cover data were available only from 2000 to 2009, only the corresponding flow, temperature, and precipitation data were acquired.

The WRMD has numerous hydrometric stations, used to measure water level (stage) in and around the Humber River Basin, NL. These water level measurements are converted to flow rates (m^3/s) by the proper stage-discharge curve. The stage-discharge curves are developed and adjusted from measurements taken throughout the year across Newfoundland and Labrador by the WRMD. The stage readings are taken hourly and averaged every day (from midnight to midnight) (Wills, H., personal communication, April 19, 2010). There are two types of flow data: (1) the real-time data, from which the measurements are taken as is and (2) the archived hydrometric data, which goes through editing and correction before they are stored. The archived hydrometric data are adjusted for ice and other possible obstructions in the river. For this analysis the real-time flows, directly from the Humber River above Black Brook station, were used for a direct measurement of runoff.

Rain weight gauges are used to obtain daily precipitation measurements in millimetres. The weight is recorded every hour and then is converted into millimetres of rain. This hourly data is summed up over the day from the 24 readings from midnight to midnight.

The daily rainfall is a measure of accumulated rainfall. Rainfall intensity is not measured. The rain weight gauges are calibrated yearly by adding weights and obtaining the readings (Wills, H., personal communication, April 19, 2010).

The climate stations are used to record air temperature at various locations throughout the Humber River Basin, NL. These sensors measure and record the data hourly. Daily measurements in degrees Celsius are averaged over 24 readings from midnight to midnight (Wills, H., personal communication, April 19, 2010).

3.4.2 Snow Cover Data

Type I curves were used in this analysis. Although Type IIA curves were also assessed, the exchange of time for cumulative degree days made no significant difference in modeling flow accuracy, much less than a 5% difference.

Small steps have been made towards obtaining more accurate and influential data for flow predictions. The WRMD performed their first snow cover survey of the Upper Humber River watershed during the 2008-2009 snow season, in March 2009. Twenty stations were established during the field visit, snow depth and snow weight were measured manually with a snow tube and spring/digital scales. Snow depth was used to calculate both snow water equivalents (SWE) and percent density of the Upper Humber River watershed snowpack (Water Resources and Management Division 2009). Although some SWE data for the Upper Humber has been measured and recorded, the data has not

yet reached a satisfactory level of quantity and quality. In effect, Type III curves were not an option since the knowledge of the snowpack and corresponding snowmelt for this area is limited. There was simply not enough snow depth data to create the more desired Type III curves. Typically, at least a few years of snow data would be required to establish preliminary assessments of the snowmelt behaviour for a particular region. Although Deer Lake Power has been conducting annual snow surveys in western NL since 1928, their snow survey sites are located in the Corner Brook watershed and the Grand Lake watershed. Some of the Grand Lake watersheds are close to the Upper Humber watershed but neither the Corner Brook nor Grand Lake watersheds actually flow into the Upper Humber River (Abbott, K., personal communication, June 16, 2010).

The Type I snow cover depletion curves were developed and daily values were read off the curves to be used as inputs for snow runoff computations in the Rango-Martinec SRM. Of course, error from snow cover depletion curve derivation propagates directly to runoff values (Seidel and Martinec 2004). See Equation [3.5] as an expression explaining this error propagation.

[3.5] $V_M = M \times S \times A$

Where:

V_M = meltwater volume [m^3];
M = melt depth [m];
S = snow coverage [fraction]; and
A = area of basin [m^2].

In attempt to minimize these errors, new snow throughout the season is not accounted for in the snowmelt depletion curves. It is still however recognized in Martinec's SRM as precipitation and therefore contributes to the runoff predictions (Seidel and Martinec 2004). In the specific case of the Upper Humber Basin, separation between new and old snow is not necessary since new snow on top of old snow can be added to the depletion curve with minimal error propagation.

3.4.3 Unknown Parameters

There are four unknown parameters in the SRM model: recession coefficient (k), snow runoff coefficient (c_s), rain runoff coefficient (c_r), and degree-day factor (a). The recession coefficient (k) can be determined through analysis of historical data. Often larger basins have a higher k factor than in smaller basins (DeWalle and Rango 2008). The runoff coefficients (c_r and c_s) indicate the percentage of precipitation (rain or snow) that appears as runoff. The degree-day factor (a) converts the number of degree-days into daily snowmelt depth.

These four parameters were optimized through Design of Experiments (DOE) methodology. DOE allows for the study of multiple factors in parallel using advanced matrix-based test plans (Anderson 2005). It can measure interaction effects, which are often significant in predicting responses. The use of a certain type of DOE allows one to fit curvature. DOE limits the number of runs required to perform the experiment because of its ability to study multiple factors simultaneously. DOE was used to learn which

combination of factors provided the best fit to the observed runoff. The most advantageous aspect of DOE is that it can measure interaction effects. Interaction effects are the effects that the parameters have in combination and they are often significant in predicting responses. One-factor-at-a-time method cannot detect these interactions since only one parameter is varied while the others remain fixed.

3.5 Model Efficiency Measures

Many measures were used to compare model efficiency at various parameter settings. Of course, a visual assessment of observed and modeled flows will immediately show whether the simulation is successful or not. The three numerical measures used were the Nash-Sutcliffe coefficient NSE, volume deviation D_v, and ratio of observed flow to modeled flow Q_o/Q_m.

These measures were used to help determine the highest model efficiency and ideal parameter settings. Although these measures were all used to determine model efficiency, some carried more weight than others. The Nash-Sutcliffe coefficient, NSE, provided the best information for this research, as it is able to quickly and easily quantify the accuracy of model outputs provided that there is observed data available for comparisons.

The Nash-Sutcliffe model efficiency coefficient is a measure of goodness-of-fit for hydrologic models. In other words, it is used to assess the predictive power of a hydrological model. The Nash-Sutcliffe coefficient, E, is calculated using Equation [3.6].

[3.6] $E = 1 - \frac{\sum_{t=1}^{T}(Q_o^t - Q_m^t)^2}{\sum_{t=1}^{T}(Q_o^t - \bar{Q}_o)^2}$

Where:

Q^t_o = observed flow at time t [m^3/s];
Q^t_m = modeled flow at time t [m^3/s]; and
$\bar{Q}_o$ = average daily observed discharge for the simulation season or for the multiple simulation seasons (depends on the time period) [m^3/s].

Essentially the goodness-of-fit is based on the complement of the residual variance between the modeled and observed flows divided by the observed flow data variance. The Nash-Sutcliffe model efficiency coefficient ranges from -∞ to +1 where E = 1, perfect match; E = 0, predictions are as accurate as $\bar{Q}_o$; and E < 0, $\bar{Q}_o$ is a better predictor than the model (DeWalle and Rango 2008). The Nash-Sutcliffe coefficient can be calculated with two different average daily discharge values: (1) the average daily discharge for each simulation season, or (2) the average daily discharge over multiple simulation seasons. Both methods provide different interpretations: (1) a season-by-season fit, or (2) an overall fit, respectively.

Volume deviation, D_v, is the second measure used to characterize the accuracy of the daily modeled flows. The volume is calculated as the accumulated flow multiplied by accumulated time period. The difference in water volume between observed and modeled values allows one to take note of the overall volume of water being carried in the water body and it provides an estimate of water quantity over a desired time period, whether the

approximation is underestimated, on target, or overestimated. D_v is defined in Equation [3.7] as:

$$[3.7]\ D_v = \frac{V_R - V_R'}{V_R} \times 100$$

Where:

D_v = the percentage difference between the total observed and modeled runoff [%];
V_R = observed runoff volume over the snowmelt season [m^3]; and
V_R' = modeled runoff volume over the snowmelt season [m^3].

The third measure is expressed as the ratio of observed flow over modeled flow, Q_o/Q_m. It is a simple measure implemented to quickly assess whether the seasonal daily modeled flows are able to closely follow the seasonal daily observed flows. The Q_o/Q_m ratio is calculated for each day in the snowmelt season and then the ratio is averaged out over a given time period, to a single estimate. This key explains the three possible conclusions from this measure:

- $Q_o/Q_m > 1$; Modeled flows are underestimated
- $Q_o/Q_m < 1$; Modeled flows are overestimated
- $Q_o/Q_m = 1$; Modeled flows are equal to observed flows

3.6 Design of Experiments Parameter Analysis

For this analysis, a 2^4 factor factorial was first implemented and used as a basis or starting point for these coefficients. This analysis was initially only performed on one

random year (2002) for testing. A 2^k factorial design is a design with k factors, each at two levels (low and high). A 2^4 factorial design implies four factors each tested at two levels. The low and high levels were set logically, but were essentially random pick-up starting points and refined based on the results of the initial 2^k design. A 2^k design contains k main effects and 2k-1 effects (which includes main effects and interaction effects). The 2^4 test has four main effects and seven effects in total. The analysis procedure for a 2^k design constitute six steps, they are: (1) estimation of factor effects, (2) form initial model, (3) perform statistical testing, (4) refine model, (5) analyze residuals, and (6) interpret results (Montgomery 2001). Statistical testing involves analysis of variance (ANOVA). The data is checked for normality, constant variance of residuals, and random run order. Design-Expert, a DOE software program, provides all of the information needed to interpret the results of the experiment properly.

As the main hydrological model efficiency measure, the Nash-Sutcliffe coefficient, E, was used to compare the best fits between the observed flows and the modeled flows using Martinec's model, referred to as the response variable. Table 3.1 provides the information used for the 2^4 factorial preliminary design: factors analyzed and their corresponding low and high levels.

Table 3.1: Preliminary 2^4 Factorial Design

Factors	Low	High
Recession Coefficient, k	0.2	0.6
Snow Runoff Coefficient, c_s	0.3	0.7
Rain Runoff Coefficient, c_r	0.3	0.7
Degree-Day Factor, a	1.0	8.0

The results from the 2^4 factorial experiment showed the highest E to be 0.60. The result is not ideal but was used as a starting point for further DOE analysis. This further analysis entailed checking for curvature and attempting to improve the model fit with a higher Nash-Sutcliffe value. From the preliminary analysis it was observed that the variation of the rain runoff coefficient, c_r did not significantly impact the results and so the c_r was left as a constant, set to 0.5 (average of low (0.3) and high (0.7) values tested) for the remainder of the tests. From the DOE analysis, the rain runoff coefficient was deemed non-significant in the model, at the 5% significance level. See Appendix F for the complete work and results from DOE testing, run sequence, tests, and results (labelled as 'Performing a DOE Analysis on Four Factors for Martinec's Snowmelt Runoff Model').

The Box-Behnken response surface design was then used to model the curvature and optimize the three remaining coefficients (k, a, and c_s). Modeling coefficients were tested

for all eight calibration years (2000 to 2007). A Box-Behnken design (1960) has three-levels and is described as a spherical design (also known as a rotatable design). It is called a spherical design because all testing points lie on a sphere (of radius$\sqrt{2}$). This design does not contain any points at the vertices of the cubic region, created by the upper and lower limits for each variable. It also does not contain an embedded factorial or fractional factorial design that can be used as a starter point, as opposed to the central composite design (CCD) – another DOE design used to test for curvature (Montgomery 2001).

The Box-Behnken design required 13 runs, which included only one center-point. Only one center-point was necessary because the combination is calculated from a formula and no variation exists, unlike in a physical experiment where center-point variability is inevitable. The three parameters (a, c_s, and k) were tested at three levels each. Table 3.2 shows the three factors tested each at three levels (low, center, and high).

Table 3.2: Refined Box-Behnken Design

Factors	Low	Center	High
Recession Coefficient, k	0.4	0.6	0.8
Snow Runoff Coefficient, c_s	0.2	0.3	0.4
Degree-Day Factor, a	1.0	1.5	2.0

The Box-Behnken design improved and refined the results a great deal with an optimized E = 0.81. Figure 3.1 presents the run order and optimal parameter combinations, proven by the Nash-Sutcliffe model efficiency coefficient. All assumptions for ANOVA were met.

Select	Std	Run	Block	Factor 1 A:Snow Runo	Factor 2 B:Degree-Day	Factor 3 C:Recession C	Response 1 Nash-Sutcliffe
	8	1	Block 1	0.40	1.50	0.80	0.58
	6	2	Block 1	0.40	1.50	0.40	-1.73
	10	3	Block 1	0.30	2.00	0.40	-1.73
	11	4	Block 1	0.30	1.00	0.80	0.81
	13	5	Block 1	0.30	1.50	0.60	0.33
	9	6	Block 1	0.30	1.00	0.40	0.38
	4	7	Block 1	0.40	2.00	0.60	-1.51
	1	8	Block 1	0.20	1.00	0.60	0.73
	7	9	Block 1	0.20	1.50	0.80	0.81
	2	10	Block 1	0.40	1.00	0.60	0.47
	12	11	Block 1	0.30	2.00	0.80	0.58
	5	12	Block 1	0.20	1.50	0.40	0.38
	3	13	Block 1	0.20	2.00	0.60	0.47

Figure 3.1: Run Order and Parameter Settings for Box-Behnken Design

3.7 Refining Recession Coefficient

The recession coefficient, k, can also be estimated through plotting Q_{t-1} vs. Q_t of the observed flows. Data was plotted for the first eight years (calibration stage from 2000 to 2007). An example of this plot is shown in Figure 3.2 for the snowmelt season in 2006. See Appendix G for Q_{t-1} vs. Q_t plots on all calibration years (2000 to 2007).

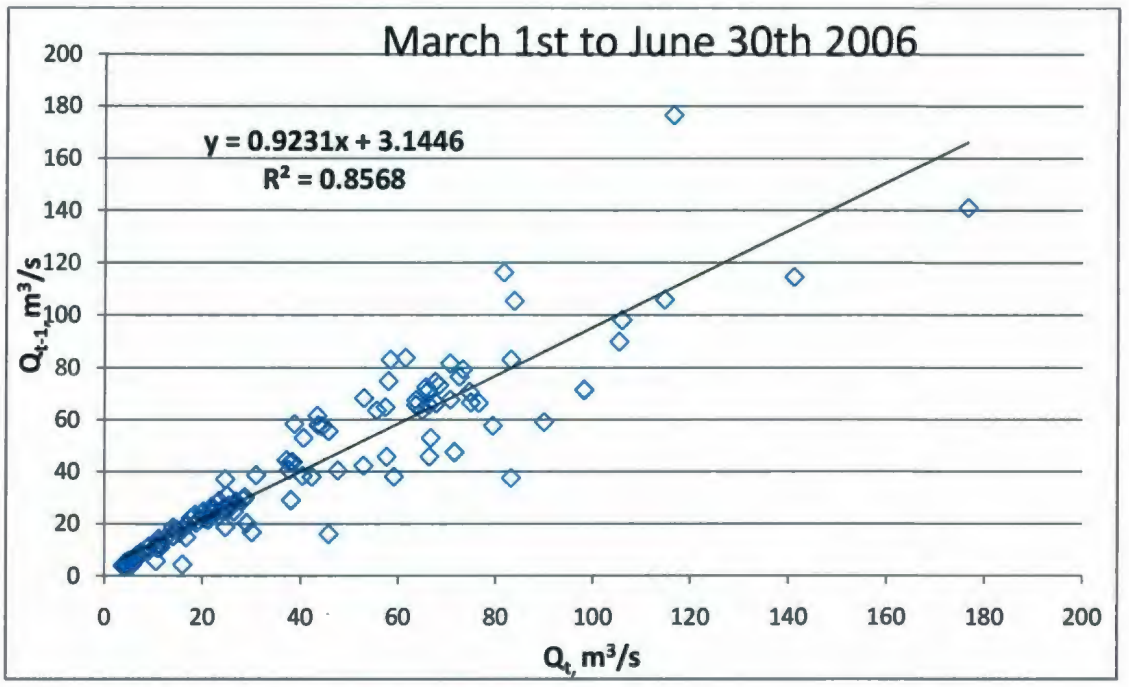


Figure 3.2: Plot of Q_{t-1} vs. Q_t with Linear Trend Line Fit to Estimate Recession Coefficient

A linear trend line was fit to each yearly snowmelt period plot. The slope provided the average approximate recession coefficient for that year. The slope being the ratio of yesterdays flow (Q_{t-1}) to today's flow (Q_t), see Equation [3.8]. In effect, the volume of today's flow can be estimated as being a part of yesterday's flow. Over the calibration stage years, the average k-value was discovered to be 0.90. See table 3.3 for the average k for each year and corresponding linear trend line fit, coefficient of determination, R^2.

[3.8] $k = \frac{Q_{t-1}}{Q_t}$

Where:

k = recession coefficient [$(m^3/s)/(m^3/s)$];
Q_{t-1} = yesterday's flow rate [m^3/s]; and
Q_t = today's flow rate [m^3/s].

Table 3.3: Average Recession Coefficient k-value and their Corresponding R^2 values for all Calibration Years 2000 to 2007

Year	Average k	R^2
2000	0.8925	0.8024
2001	0.9520	0.9124
2002	0.9366	0.8669
2003	0.9006	0.8162
2004	0.8484	0.7254
2005	0.9490	0.9107
2006	0.9231	0.8568
2007	0.8110	0.6579
	Overall Average k = 0.9017	Overall Average R^2 = 0.8186

The year 2007 has a significantly low correlation compared to the remaining calibration years. Being only one year out of the eight it was not considered to be of major concern. Year 2007s flows were likely to have been influenced more by rainfall, snowmelt, and

temperature throughout the snowmelt season compared to the remaining calibration years.

In some cases, the recession coefficient varies significantly. This variation can occur over a time period of many years, over a single year, or even over one snowmelt period – it depends on basin characteristics and meteorological occurrences. For these cases, coefficients x and y are calculated using Equation [3.9] (Martinec and Rango 1986).

[3.9] $k_{n+1} = xQ_n^{-y}$

Equation [3.9] explains how k varies, particularly in relation to the current flow rate. Coefficients x and y are solved by simply recording two coordinate pairs (one of low flow value and the other of high flow value) from the Q_{t-1} vs. Q_t plots. Equation [3.10] supplies two equations and two unknowns when both coordinate pairs are substituted and the coefficients are determined (Martinec and Rango 1986).

[3.10] $log(k) = \log(x) - ylog(Q_{t-1})$

Where $k = \frac{Q_{t-1}}{Q_t}$

For the Upper Humber Basin, however, the k-value over the eight calibration years did not vary significantly and was accepted as a single k-value. The largest variability of recession coefficient k is often observed in small basins with low flows less than 10 m^3/s.

As the basin size and flow rate both increase, the recession coefficient experiences much less fluctuation. For a basin such as the Upper Humber Basin, size 470 km^2 and flows on average of 140 m^3/s, the k-value can be said to range between 0.90 ± 0.05 (Martinec and Rango 1986). From this new estimation, a refined DOE Box-Behnken design was performed by varying k from a narrower range: 0.85 to 0.95. Table 3.4 shows the improved and more concise levels for the three factors at levels low, center, and high.

Table 3.4: Final Box-Behnken Design

Factors	Low	Center	High
Recession Coefficient, k	0.85	0.90	0.95
Snow Runoff Coefficient, c_s	0.2	0.3	0.4
Degree-Day Factor, a	1.0	1.5	2.0

This resulted in an optimal overall Nash-Sutcliffe coefficient of 0.85 averaged over all eight calibration years (2000 to 2007), an improvement from the 0.81 obtained earlier. Figure 3.3 depicts this final DOE analysis using Design Expert.

Select	Std	Run	Block	Factor 1 A:Snow Runo	Factor 2 B:Degree-Day	Factor 3 C:Recession C	Response 1 Nash-Sutcliffe
	8	1	Block 1	0.40	1.50	0.95	0.83
	6	2	Block 1	0.40	1.50	0.85	0.7
	10	3	Block 1	0.30	2.00	0.85	0.7
	11	4	Block 1	0.30	1.00	0.95	0.84
	13	5	Block 1	0.30	1.50	0.90	0.82
	9	6	Block 1	0.30	1.00	0.85	0.83
	4	7	Block 1	0.40	2.00	0.90	0.7
	1	8	Block 1	0.20	1.00	0.90	0.85
	7	9	Block 1	0.20	1.50	0.95	0.84
	2	10	Block 1	0.40	1.00	0.90	0.83
	12	11	Block 1	0.30	2.00	0.95	0.83
	5	12	Block 1	0.20	1.50	0.85	0.83
	3	13	Block 1	0.20	2.00	0.90	0.83

Figure 3.3: Final Design of Experiments Analysis using Design-Expert – Run Order, Level Settings, and Responses

Significant factors at the 5% significance level were the three parameters: A (snow runoff coefficient), B (degree-day factor), and C (recession coefficient). All three interactions between parameters: AB, AC, and BC, were found to be significant. All ANOVA diagnostics passed, with a two-factor interaction model (i.e. no significant curvature), and adjusted R^2 of 0.98.

3.8 Model Analysis

The model analysis was divided into two sections: (1) calibration and (2) validation. The first eight years of data (2000 to 2007) were used to calibrate the model and its coefficients using DOE. The remaining two years of data (2008 to 2009) were used to

validate the model. The coefficients calibrated from the 2000 to 2007 data were used in the validation period to assess whether the modeled flow values were good predictors of the observed flows. The model efficiency measures used were the Nash-Sutcliffe E, the volume difference D_v, and the Qo/Qm ratio.

3.8.1 Calibration Period

The optimal coefficients using the refined DOE Box-Behnken design for 2000 to 2007 were a = 1.0 $mm°C^{-1}d^{-1}$, k = 0.9, c_s = 0.2, and c_r = 0.5, with a goodness-of-fit Nash-Sutcliffe value, E = 0.85. A degree-day factor of 1 mm/°C day means that one degree-day of thaw can melt one millimeter of water from the snowpack. The snow runoff coefficient of 0.2, which means 20% of the snowmelt contributes to the flow and the rain runoff coefficient of 0.5, which means 50% of the rain contributes to the flow. The autoregressive coefficient is very strong at 0.90, 90% of today's flow is predicted from yesterday's flow. A high k means that the meteorological effects, like snow and rain runoff, have less of an impact on the prediction of today's predicted flow rate. A single value for each key parameter was determined through calibration; this was based on the important assumption that there is homogeneity of these parameters in the study watershed.

The modeled flow values were compared visually to the observed flows for every snowmelt season. Figure 3.4 illustrates one randomly chosen snowmelt season (2001)

plotting $Q_{modeled}$ and $Q_{observed}$ vs. Time. See Appendix H for all $Q_{modeled}$ and $Q_{observed}$ vs. Time plots from 2000 to 2009.

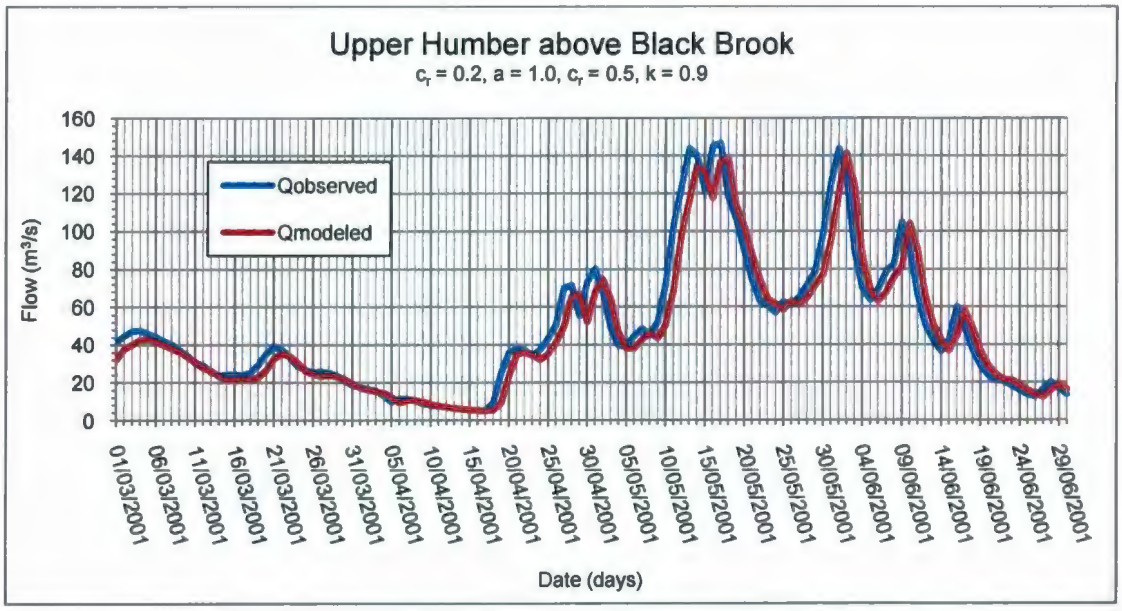


Figure 3.4: Comparison of $Q_{modeled}$ and $Q_{observed}$ over the 2001 Snowmelt Period

Although the modeled and observed values fit very closely, the modeled values are lagged by one day. This lag does not aid in forecasting, since the modeled flows are not actually predicting at all. Instead they are providing an estimate closer to the previous day's flow rather than an estimate for today's flow. This observation shows that the first term in the model is the dominant term and it means that there is a high autocorrelation between adjacent time-period flows. Another indication of this high autocorrelation was observed earlier from the high and dominant autoregressive coefficient, k = 0.9.

For the final Box-Behnken calibration experiment, the seasonal Nash-Sutcliffe coefficient was also calculated to assess the fit for each snowmelt season individually. Most seasons provided an excellent fit observing two seasons with an E above 0.90 (2001 and 2005) and only two seasons with an E below 0.80 (2004 and 2007). See Table 3.5 for the break down.

Table 3.5: Seasonal Fit Nash-Sutcliffe Coefficients for Final Box-Behnken Calibration Experiment 2000 to 2007

Snowmelt Year	**Nash Sutcliffe Coefficient, E**
2000	0.80
2001	0.91
2002	0.85
2003	0.84
2004	0.73
2005	0.91
2006	0.86
2007	0.64

A ratio of Q_o/Q_m was calculated to numerically assess whether the overall modeled flows are being overestimated or underestimated. Over the calibration period, the ratio of observed flows to modeled flows was 1.06. This draws the conclusion that on average the modeled flows are 6% lower than the observed flows.

The percent volume deviation, D_v, for each year was always less than 6.5%. Table 3.6 provides the yearly D_v from snowmelt seasons 2000 to 2007.

Table 3.6: Percent Volume Difference for Snowmelt Seasons 2000 to 2007

Snowmelt Year	**Dv, %**
2000	1.10
2001	4.34
2002	6.40
2003	1.57
2004	0.99
2005	4.33
2006	2.62
2007	2.41

All differences in volume report that the observed runoff volumes were greater than the modeled runoff volumes ($V_R > V_R'$). There are two possible reasons for this consistent underestimation (1) not all sources of runoff are accounted for (i.e. ground infiltration or evaporation), or, (2) not all accounted for sources are being modeled properly.

3.8.2 Validation Period

The two-year validation period from 2008 to 2009, was used to test the prediction power of the model and its estimated coefficients set from the calibration stage. After running

the SRM, the predictions for 2008 and 2009 showed good results with an overall average Nash-Sutcliffe coefficient of 0.81. Refer to Figures 3.5 and 3.6 as they illustrate the close fit between the modeled and observed flows for 2008 and 2009, respectively.

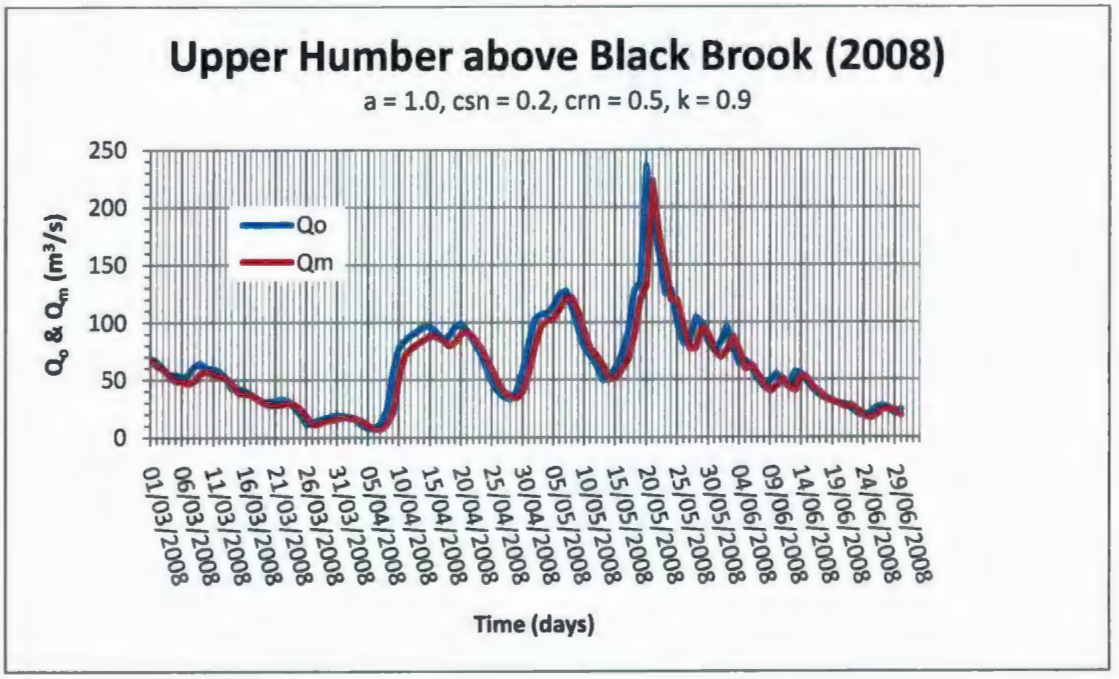


Figure 3.5: Comparison of $Q_{modeled}$ and $Q_{observed}$ over the 2008 Snowmelt Period

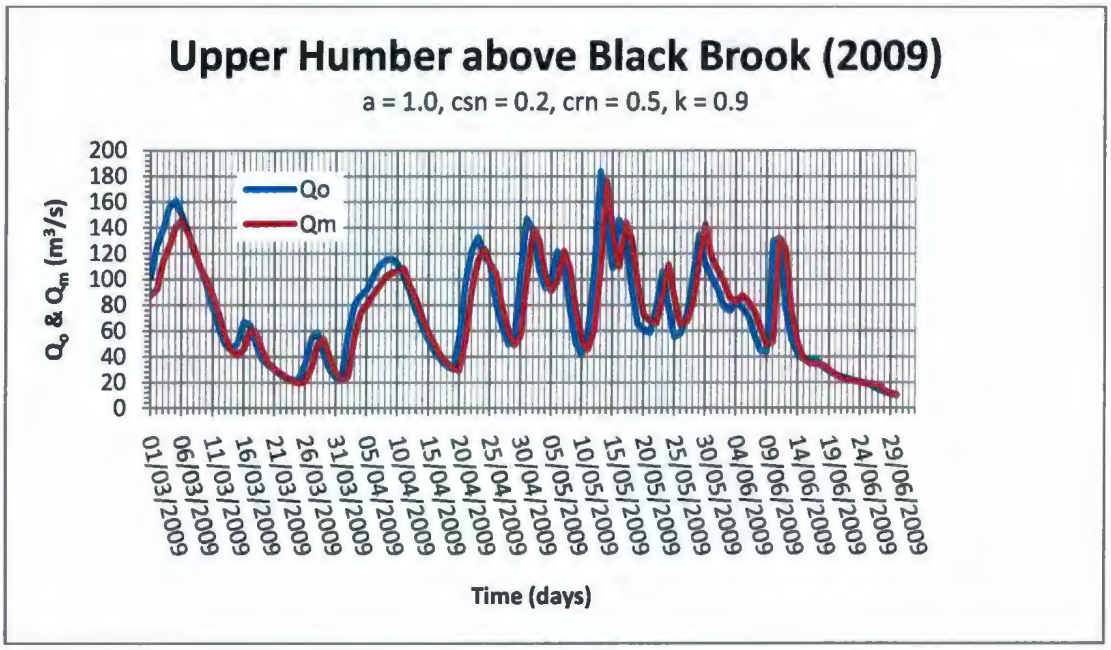


Figure 3.6: Comparison of $Q_{modeled}$ and $Q_{observed}$ over the 2009 Snowmelt Period

The seasonal Nash-Sutcliffe coefficients for the validation stage are reported in Table 3.7. Year 2009 was less predictable from using the calibrated parameters than year 2008. It may be that the 2009 snowmelt season was different from those of the calibration stages.

Table 3.7: Seasonal Fit Nash-Sutcliffe Coefficients for Final Box-Behnken Validation Experiment 2008 to 2009

Snowmelt Year	Nash-Sutcliffe Coefficient, E
2008	0.86
2009	0.75

The ratio of $Q_{observed}$ over $Q_{modeled}$ was 1.07 for the validation stage, slightly higher than the ratio from the calibration stage (1.06). This means that the two validation year flows

being modeled are on average seven percent lower than the actual observed flow rates. The one day lag was still present indicating a strong autocorrelation to adjacent flows for forecasting time periods. The strong influence of the first term in the SRM naturally made the second term much less significant (the first term being a percentage of yesterday's flow and the second term being the temperature effects as well as snow and rain runoff). It is difficult to assess whether the SRM is a model worth applying to the Humber River Basin with a validation period of only two years, but currently there is no improvement from adding the snow data into the prediction model. The difference in volume for the two validation years were both less than the 6.5% and within range from the volume differences calculated for the calibration years, see Table 3.8.

Table 3.8: Percent Volume Difference for Snowmelt Seasons 2008 and 2009

Year	D_v, %
2008	6.20
2009	2.50

Martinec (1972) analyzed information for a mountainous watershed of area 43.2 km^2. The data indicated that the percent snow cover area at the time of the peak runoff varied from about 25 to 70%. Peak flows for the Humber River Basin, watershed area of 470 km^2, ranged in snow coverage of 34 to 97% from years 2000 to 2009. This observation indicates that the peak runoff in the snowmelt season does not necessarily coincide with a certain percent snow coverage.

The two stages, calibration and validation, of this research both provided significant insight in flood forecasting on the Upper Humber River, NL, especially in adapting snow data to aid in flood forecasting. Martinec's snowmelt runoff model functioned well over the tested 10 year period. Design of experiments also proved successful in determining the significance and appropriate levels for each unknown parameter (concluding that the rainfall runoff coefficient was the only parameter of the four tested to be non-significant at the 5% significance level). Model efficiency measures such as the Nash-Sutcliffe coefficient, NSE, also supplied important information on the functionality of the model and its ability to predict daily future flows. The final chapters will provide more detailed information on research results and recommendations to improve flow predictions for the Upper Humber Basin.

- Chapter 4 -

Snow Water Equivalent (SWE) Data

The primary research objectives, involving remotely sensed snow cover data and daily flow modeling, have been completed. Although the analysis using snow cover data as an input variable to Martinec and Rango's SRM shows promise for daily flow predictions, other alternatives, if feasible, should be assessed. The secondary research objective was developed to discover whether or not recently available SWE data are: (1) accurate and useable, and if so, then (2) to determine whether or not this data will improve daily flow predictions for the Upper Humber Basin. SWE data via satellite became available for the Humber River Basin, NL in January, 2010. SWE is expressed as quantity of snow reserves: the amount of liquid water in the snowpack, if the snowpack were to melt completely. SWE is calculated by Equation [4.1] (Natural Resources Conservation Service 2010).

[4.1] $SWE = density \times depth$

Where:

SWE = snow water equivalent [mm];
density = relative density (ρ_{snow}/ρ_{water}) [$(kg/m^3)/(kg/m^3)$]; and
depth = depth of snow [mm].

To ensure proper units, the density must be represented as relative density (or specific gravity), with respect to liquid water. The snow depth is the vertical distance from the snow surface to the ground.

4.1 Remotely Sensed Snow Water Equivalent Estimates

The SWE data is available through the European Space Agency (ESA) Data User Element (DUE) Global Snow Monitoring for Climate Research or, more simply, the GlobSnow project. This SWE information was derived from AMSR-E (Advanced Microwave Scanning Radiometer – EOS) sensor data in combination with ECMWF (European Centre for Medium-Range Weather Forecasts) weather station observations. SWE estimates are available in the northern hemisphere for the years 2003 to 2008. By August 2010, the derived SWE dataset will provide daily SWE data for the last 30 years (Luojus et al. 2009).

The SWE data is saved in HDF4-format. Each day provides two files of information: (1) the SWE estimate and (2) its error estimate (i.e. data variance). The sensor used to obtain

the GlobSnow SWE data, AMSR-E, is one of the six sensors aboard NASA's Aqua satellite. The AMSR-E passive microwave observations, along with weather station observations collected by ECMWF, are integrated and used to produce maps pertaining to SWE estimates. The GlobSnow SWE product encompasses the entire northern hemisphere (except Greenland) in a single data field, projected in Equal-Area Scalable Earth Grid (EASE-Grid). This projection changes the shape of the land but the land mass areas are accurate and can be used for appropriate calculations and data processing. The SWE nominal resolution is 25 km x 25 km per pixel, providing a pixel area of 625 km^2. The geometry of the pixels can vary (Luojus et al. 2009).

4.2 Snow Water Equivalent Obstacles

Although SWE data availability for North America was a significant accomplishment, there are potential pitfalls in using the SWE data, particularly for the Upper Humber Basin. First, the area of one pixel is 625 km^2, compared to the area of interest: approximately only 470 km^2. Given that the resolution of the SWE product is larger than the area of interest, data accuracy may be a problem (i.e. information is averaged over such a large area and there may be large SWE derivation errors with only one pixel). Second, the Upper Humber Basin is a mountainous region. Mountainous regions provide less accurate SWE information because the data is obtained using radar. Difficulties with radar arise when differentiating elevations from topographic variability. Third, the Upper Humber Basin is located in the boreal forest with a dense forest canopy. According to GlobSnow SWE Product Guide from 2009, there are low correlations over the boreal

forest, especially in more remote areas with a sparse climate station observing network. Fourth, Newfoundland and Labrador is subject to significantly higher than average annual precipitation compared to the rest of Canada. During the spring months, the snow is frequently considered "wet snow" as opposed to "dry snow". The AMSR-E sensor has difficulties with the higher reflectance of wet snow (Luojus et al. 2009). These difficulties relate directly to SWE data accuracy. At times, the SWE estimates cannot be calculated at all.

4.3 Snow Water Equivalent Data Processing and Analysis

The SWE data was obtained in a similar manner as the snow cover area data from remote sensing: using PCI Geomatica, EASI script, and importing the individual text files into Microsoft Excel with a VB script.

From preliminary data analysis it was clear that the SWE data was not feasible for further analysis. The obstacles of the remotely sensed data described earlier, for the Upper Humber Basin in particular, were too much to overcome at this point in time. First, many days within the snowmelt period (March 1st to June 30th) provided void SWE estimates of zero or missing data. This does not mean there was no snow cover; this means that there was an error in determining the SWE for that region. Again only one pixel: size 625 km^2 was observed for the Upper Humber River. Second, for the few SWE estimates that were not zero (approximately six points per year), the SWE variance was large. For example, on April 14th 2008, the SWE estimate was 197.36 mm with a variance of 1,796.03 mm.

This is a large variance and means that this SWE information is not useable. All SWE estimates obtained from 2003 to 2008 for the Upper Humber River basin demonstrated these large variances and were therefore not useable.

4.4 Potential Improvement of Daily Flow Predictions

Technical advances may bring more accurate SWE data. It could then be used to test for improvements in the daily flow predictions. In slightly modifying the Rango-Martinec SRM, the change in SWE from one day to the next, in mm, is used as an input, see Equation [4.2].

$$[4.2]\ Q_{n+1} = kQ_n + (1-k)[c_s\boldsymbol{\Delta SWE} + c_r P_n]\frac{A\times 1000}{86400}$$

The change in SWE (mm) replaces the product of snow covered area in percent, the cumulative degree days, and the degree-day factor. This SWE difference between adjacent time periods will indicate the actual amount of melted snow rather than the snowpack's entire snow reserves.

By comparing the change in SWE to the product of: $S_n \times T_n \times a_n$, it would provide a good indication of whether or not the SWE data will impact the flow predictions. As the SWE data is currently not accurate or precise enough for proper analysis, the pursuit of flow prediction improvement will have to be investigated upon appropriate technical advancements.

- Chapter 5 -

Discussion

This research focused on a sub-watershed of the Humber River Basin: the Upper Humber River basin above Black Brook, NL. The main objective was to use satellite snow data to model more accurate flow rates, which would help forecast floods in the area. This chapter will discuss the methodology of this model, practical results, and potential applications for this research.

5.1 Methodology and Results Summary

The daily flow prediction analysis is broken down into two sections: 1) remote sensing technology and 2) SRM model analysis. Figure 5.1 shows how the daily flow prediction analysis is divided into these two sections.

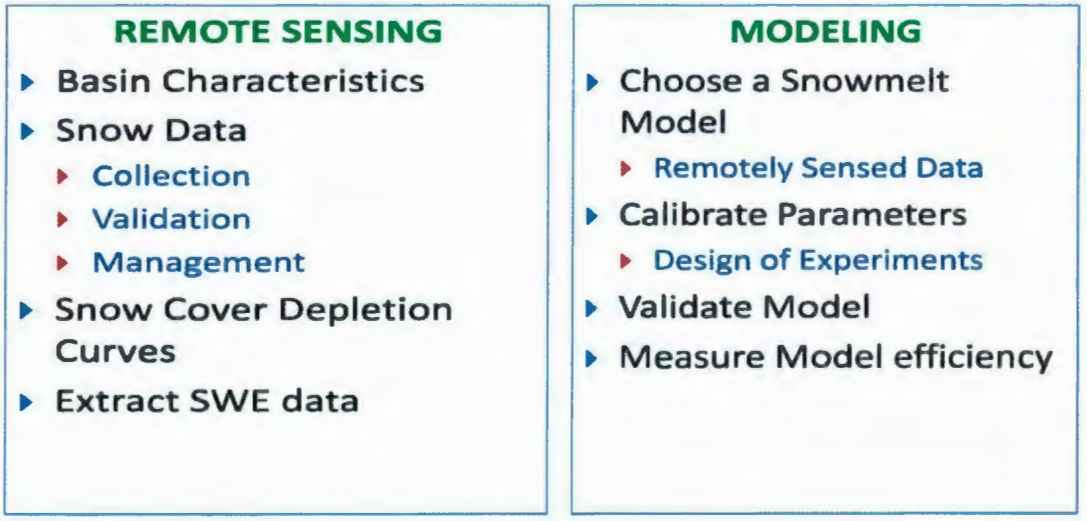


Figure 5.1: Daily Flow Prediction Analysis Divided into Two Sections

The remote sensing section involved four steps: (1) understanding the study area and remote sensing in that area, (2) collecting, validating, and managing the MODIS snow cover data, (3) developing and interpreting snow cover depletion curves, and (4) extracting and assessing SWE data.

Remote sensing is currently the best way to monitor daily snow cover data, based on three main reasons: it is easily accessible, it is reasonably accurate, and it is able to cover large areas within acceptable time periods. The MODIS sensor on Terra was chosen based on its reliability, daily temporal coverage, and ease of access. One major disadvantage was cloud cover. The study area is an extremely cloudy region and given that the MODIS sensor is an optical sensor, cloud cover can significantly affect snow data interpretation. The possibility of misclassifying snow and cloud was a major obstacle. The snow cover data was extracted and sifted through using a 20% cloud cover

threshold. This threshold was used to ensure minimal snow misclassification on account of high cloud cover.

Conventional snow depletion curves were plotted from the percent snow covered area over time. Snowmelt forecasts are most accurately predicted in situations where the snow accumulation and melt periods are well-defined, with minimal disruptions in the accumulation and melt patterns. Ideally, there is relatively little precipitation during the melt (forecast) period (Maidment 1993). Of course, in reality, no depletion curve is perfect. There are many internal processes taking place within a snowpack as it consolidates. The position and shape of each individual crystal changes over time from moisture transport, overlying weight, and wind redistribution. A snowpack's accumulation and melt periods are determined by the snowpack's degree of consolidation. Non-homogeneous snowpacks can develop from rainfall between periods of snow accumulation. The rain can freeze into ice overtop of a snowpack creating layers of snow and ice (Watt 1989). Over such a relatively large study area: 470 km^2, it is impossible to determine this kind of information on the snowpack, and it must be assumed that there is some closeness to a homogenous snowpack.

The second section, on modeling the flows, involved: choosing a snowmelt runoff model, calibrating its parameters, and validating the model. Table 5.1 provides a results summary from the model analysis. These results represent the ability of the SRM model to predict in both the calibration and validation stages for the Upper Humber Basin.

Table 5.1: Results Summary for Daily Flow Predictions using MODIS Snow Cover Data

Optimal Parameter Settings: $a = 1.0\ mm°C^{-1}d^{-1}$ $k = 0.9$ $c_s = 0.2$ $c_r = 0.5$	
Calibration Stage 2000 to 2007	
Nash-Sutcliffe, $E_{overall}$	0.85
Highest $E_{seasonal}$ observed in 2001 and 2005	0.91
Lowest $E_{seasonal}$ observed in 2007	0.64
Difference in Volume, $D_{v,overall}$	2.97%
Highest $D_{v,seasonal}$ observed in 2002	6.40%
Lowest $D_{v,seasonal}$ observed in 2004	0.99%
$(Q_o/Q_m)_{overall}$	1.06 (modeled flows 6% underestimated)
Highest $(Q_o/Q_m)_{seasonal}$ observed in 2007	1.10
Lowest $(Q_o/Q_m)_{seasonal}$ observed in 2000	1.02
Validation Stage 2008 to 2009	
Nash-Sutcliffe, $E_{overall}$	0.81
Highest $E_{seasonal}$ observed in 2008	0.86
Lowest $E_{seasonal}$ observed in 2009	0.75
Difference in Volume, $D_{v,overall}$	4.35%
Highest $D_{v,seasonal}$ observed in 2009	6.20%
Lowest $D_{v,seasonal}$ observed in 2008	2.50%

$(Q_o/Q_m)_{overall}$	1.07 (modeled flows 7% underestimated)
Highest $(Q_o/Q_m)_{seasonal}$ observed in 2008	1.09
Lowest $(Q_o/Q_m)_{seasonal}$ observed in 2009	1.06

From the results, it was observed that the modeled flows fit well with observed flows and that the forecasting period shows promise. The average Nash-Sutcliffe value during the calibration stage was a respectable E = 0.85 and for forecasting (validation stage) maintained its very good fit at E = 0.81. The difference in volume between observed and modeled flows during the validation stage was on average 4.35%. This is a fairly small difference and acceptable for forecasting purposes. Over the 10 year period the difference in volume was always below 6.5%. This minimal change in volume prediction provides considerable security in future predictions. The ratio of observed flow to modeled flow was 1.06 for the calibration stage and 1.07 for the validation stage. The modeled flows are rather consistent with flow underestimation by about 6% to 7%. Also, the small variability of Q_o/Q_m over the 10 years shows great promise for future predictions, regarding the user's expectations for accurate flow prediction.

One of the main drawbacks of the SRM model is that with daily flow predictions there is a one-day lag in the modeled flow rates. This one-day lag is visible from the data set and visually from the plotted 'observed and modeled flows' over 'time'. The modeled flow lag is extremely difficult to adjust for; if the modeled flows were too far ahead of the observed flows one could simply change the time step inputs. A one-day lag means that

the next day's flow information is not being 'transferred' (or calculated) quickly enough. One method that may mitigate the lag would be to predict weekly or monthly flows to average out the lag, but this is ineffective for flood forecasting purposes. A daily flow (or smaller time step, for example: 6-hr or hourly) is a necessary prediction time-step when predicting floods. This allows for proper flood management and effective residential evacuations, if necessary. Although hourly data is collected by the WRMD for the Upper Humber Basin, the information is not conveniently available or reliable enough for proper analysis (i.e., missing data).

No model is perfect and every model experiences some uncertainties in its prediction power. Hydrological processes exhibit substantial variability and cannot be completely accounted for by physical laws. Variability, particularly in flow prediction, is often caused by the natural randomness of driving variables such as precipitation. An incomplete understanding of predicting system outputs from system inputs and errors in parameter estimation can also be a source of flow prediction variability (Maidment 1993). These areas of variability are most likely present in the SRM model applied to the Upper Humber River, NL. There is no way to forecast the exact flow values for any river, the best that can be done is to make close estimates that will aid in flood mitigation.

An ancillary objective was added when snow water equivalent data become available to the study region in January 2010. This SWE data was also obtained via satellite and was processed using the same GIS software used for the MODIS/Terra snow cover data. A

simple modification/substitution to Martinec and Rango's SRM model enables a test on the SWE data to conclude whether or not the new data were useful for the Upper Humber Basin, in terms of improved flow predictions. However, the analysis was unable to proceed past the point of preliminary data retrieval and analysis. The SWE estimates were found to be unreliable for the Upper Humber region. Two reasons why the SWE data are unreliable: (1) many zero value estimates leaving only a few data points per year and (2) for the few SWE estimates retrieved, the data had extremely high variances (many times larger than the estimate itself). Although remote sensing has been successfully used for estimation of snow areal extent, estimation of catchment snow water storage is much more difficult (Maidment, 1993).

Looking to the future, in fall 2010, the WRMD has plans to install a SWE sensor within the Humber River Basin. The readings from an automated SWE monitoring station will help provide added information used as an input for the adapted Martinec-Rango SRM model, Equation [4.2] (adapted for SWE data). The SWE sensor will also be used to improve the GlobSnow SWE product through regression analysis (of remotely sensed SWE estimates with ground sensor SWE data). This added SWE information should help close the gap created between actual SWE and the GlobSnow SWE estimates for the Upper Humber Basin, Newfoundland and Labrador.

5.2 Applications

The methodology developed from this research has a number of promising hydrological applications. The most pertinent hydrological applications are: daily flow predictions (not particularly for the Upper Humber Basin but more likely for larger basins), volume predictions, summer reservoir level measurements, and MODIS snow cover data extraction.

Certainly, for daily flow predictions, parameter estimation must be recalibrated frequently for the Upper Humber River and as historical data is accumulated over time, a better estimate, fit, and understanding of the basin's flows is expected. This SRM method is also a promising model for other basins with similar goals. The SRM model parameters are set to the particular basin being analysed, but the main steps taken to forecast the daily flow rates remain the same. Two essential categories of information necessary to explore are (1) the individual basin characteristics and (2) the available regional data. A perfect example to test this developed methodology is on the Exploits River Basin located in central NL. The Exploits River Basin shows great promise for improved flood forecasting based on the following reasons:

1. Flat topography, hence more accurate SWE estimates (i.e. decreased variability)
2. Large basin area, approximately 10,000 km^2
3. Substantial hourly meteorological data throughout the entire basin
4. Extensive snow sampling records from Abitibi (which is now Nalcor Energy)

This research can also help predict water volumes just as easily as it can predict daily flows. These water volume forecasts help to assess and categorize periods of low, normal, and high flows. Summer reservoir levels are often estimated from volume of water and based on historical data. Quantifying the entire amount of spring runoff, with snowmelt as a major contributing factor, can help predict future summer flow volumes. Hydropower companies are especially interested in this information, especially in the summer, when the lowest flows of the year are routinely observed. The hydropower companies need to plan for the future and ensure that they have the proper volumes of water to create sufficient hydropower to meet the expected demands.

Lastly, this research provides the detailed information on MODIS snow cover data extraction, quality assessment, and interpretation methods. Remote sensing has become the newest technology with real applicability to many hydrological processes. It is becoming the standard for many hydrological applications because it is practical for data mining in remote areas, provides a large variety of information at once, and is capable of handling large amounts of data efficiently and presents them graphically. This MODIS snow cover imagery can also be used qualitatively in flood forecasting to declare the end of the snowmelt season, simply through binary decision making: "snow" or "no snow".

- Chapter 6 -

Conclusion and Recommendations

This final chapter provides conclusions on the research performed for this thesis. This chapter also offers some recommendations for future work on the methodology developed from this research. Recommendations on the uses of remotely sensed snow data, improving Martinec and Rango's SRM for the Upper Humber River, NL, and its potential application to other basins with similar research objectives are discussed.

6.1 Conclusion

This research tested the combination of remotely sensed data and a snowmelt runoff model (SRM) for a sub-watershed in western Newfoundland's Humber Valley. MODIS Terra images were acquired from 2000 to 2009 and processed to extract snow cover data. The snow cover data were used to plot conventional depletion curves. The derived snow cover data obtained from these curves, along with other parameters such as precipitation, watershed area, discharge, and temperature, were input into Martinec's snowmelt runoff model. The four unknown parameters required to run the SRM were optimized using DOE methodology and the recession coefficient k was further refined through Q_{t-1} vs. Q_t plots. This DOE-aided calibration proved statistically significant for three of the four

coefficients, and the fourth coefficient (rain runoff coefficient, c_r) was deemed statistically non-significant at the 5% level.

The SRM was tested by dividing the 10 years of data into two stages: calibration and validation. Calibration data, from 2000 to 2007, was used to optimize and define the empirical coefficients of the model. The Nash-Sutcliffe goodness-of-fit coefficient, E, was used to calibrate these coefficients for the best fit between modeled flows and observed ones. The calibration yielded an optimized average $E_{overall} = 0.85$ over years 2000 to 2007. The second stage, validation, was used for the remaining two years 2008 to 2009, to assess the model's prediction power. The validation provided some mixed results as $E_{2008} = 0.86$ and $E_{2009} = 0.75$. It is difficult to assess the model's prediction power with such a short validation period, but only 10 years of MODIS data are available at the present time. The one-day lag in the modeled flows is difficult to overcome because of the small catchment area and the daily time-step. Lastly, SWE estimates were extracted from AMSR-E/Aqua satellite images, but they were discovered to be unreliable for the study area and no further analysis was performed.

Quantifying snowmelt has been a challenging aspect of hydrology for daily flow modeling, with many uncertainties and difficulty measuring vast watershed areas. This research makes significant contributions to the field of hydrology providing a valuable methodology in adapting remotely sensed snow data to daily flow simulation. The collection, validation, and management of remote sensing snow cover images as well as

the incorporation of these remotely sensed snow cover images into a snowmelt runoff model, will be helpful to local authorities.

6.2 Recommendations

There are some modifications that may be useful for predicting flow rates on the Upper Humber Basin, NL. In the coming years, it is recommended to test these applications further. Recommendations are as follows:

1. To use the MODIS snow cover images for qualitative purposes. This will aid in flood forecasting and indicate the end of the snowmelt season. Satellite imagery can be evaluated as being part of one of the two categories: "snow" or "no snow".

2. To assess a shorter time-step: hourly, 6-hour data, or even half day. For a small basin, such as the Upper Humber Basin, a daily time-step may not be a short enough response time-period. The flow rate may change more rapidly with a sudden change in one or more input parameters. While data is currently being retrieved hourly by the WRMD, they are not conveniently available or reliable enough (i.e. missing data) for extensive and accurate studies.

3. To change the degree-day factor (a) every half month throughout the snowmelt season. Martinec has expressed that 'a' should be treated as a time series rather than a fixed parameter. It is often observed that 'a' increases over the snowmelt

season, especially for heavily forested study areas (Rango and Martinec 1986). This increase in the degree-day factor over the snowmelt season can also be considered as an index of decreasing albedo. The change in 'a' over time is coincident with the increase in snow density over time.

4. To treat the whole watershed as a single unit may be too rough an estimate. It is recommended that the watershed be divided into smaller elevation zones (50 m or 100 m) to improve modeling accuracy.

5. To use snow data collected via satellite to quantify snowmelt runoff for future summer reservoir levels. The purpose of this analysis would be to predict reservoir levels as being either low, normal, or high by measuring flow volumes to provide an indication of water availability. This water availability assessment would be very useful, especially to hydropower companies and government institutions.

6. To test a physical-process based model, or process-based model, is recommended. This process-based model would attempt to mimic the real-world physical processes of the basin. Representations of surface runoff, subsurface flow, snowmelt, evapotranspiration, and channel flow are typical processes mimicked in this model. If modeled accurately, it is likely to capture more of the complex dynamic inner workings of real-world observations affecting runoff, as opposed to the simpler degree-day Martinec-Rango SRM model.

7. To further improve the remotely sensed SWE data over the Upper Humber Basin of Newfoundland and Labrador, for potential improvement in modeling:

    a. To improve the conventional snow depletion curves and upgrade them to the Type III curve by plotting 'SWE' vs. 'Time'. Continued snow surveys by the WRMD along with the installation of the SWE monitoring station will provide useful information. This SWE data can be used in a regression analysis between remotely sensed SWE data and the collected ground work snow data. These Type III curves will help discern additional basin snowmelt information. Most important being the actual volume of water, if all of the snowpack were to melt. This Type III plot also offers information on the likelihood that the degree-day factor alters throughout the season (Rango and van Kawijk 1990).

    b. To substitute SWE estimates in Martinec's adapted SRM, SWE equation [4.2]. This will provide a more accurate snowmelt estimate compared to the snow cover data, which does not offer any snow depth measurements.

8. To implement this process for larger basins may also provide more viable and practical results. It is recommended that the WRMD extend the methodology developed herein to Newfoundland and Labrador's Exploits River Basin. This includes: acquiring MODIS snow cover data, modeling the next day's flow with Martinec and Rango's SRM, monitoring SWE by satellite and through ground

surveys and/or SWE sensors, and modeling the next day's flow with the adapted SRM – SWE equation. There are several reasons why the Exploits River Basin is likely to produce more valuable flow predictions than the Upper Humber Basin. First, its basin area is much larger, approximately 10,000 km^2. Second, the basin has flat topography opposed to the mountainous Upper Humber Basin. This flat terrain will increase the accuracy in the remotely sensed SWE estimates. Third, more hourly data is available throughout the Exploits River Basin. Fourth, extensive snow sampling records are available. This ground snow survey data will provide a better understanding of the amount of snow accumulated and melted annually over the winter/spring seasons. It may also improve GlobSnow SWE product estimates through regression modeling: decreasing the data's variability.

This research was meant to explore the use of satellite snow data to aid in flood forecasting and was tested on the Upper Humber Basin, NL. A second watershed, of interest to the WRMD, where this methodology will be tested, is the Exploits River Basin, NL. Currently, the Newfoundland and Labrador government is not implementing any type of snow data into their flood forecasting models. The methodology developed for this research provides a good understanding of the Upper Humber River's significant flood forecasting parameters as well as a basis for future work.

References

Abbott, K. N. (2009). *A Guide to Mapping Snow Cover using MODIS Data* (Technical Report). Retrieved from Water Resources Management Division Department of Environment and Conservation Government of Newfoundland and Labrador via Keith Abbott

Anderson, M. J. (2005). Trimming the FAT out of Experimental Methods. *Optical Engineering*, June, pages not provided. http:www.statease.com/pubs/doeprimer.pdf

Bettwy, M. (2004). *The Benefits of Hydrologic Predictability*. NASA. Retrieved on February 1, 2010 from www.nasa.gov

Cai, H. (2009). *Flood Forecasting on the Humber River using an Artificial Neural Network Approach*. Unpublished master's thesis, Memorial University of Newfoundland, Canada.

DeWalle, D. R., and Rango, A. (2008). *Principles of Snow Hydrology*. New York: Cambridge University Press.

Hall, D. K., and Martinec, J. (1985). *Remote Sensing of Ice and Snow*. New York: Chapman and Hall.

Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W. (2005). *Geographic Information Systems and Science (2nd ed.)*. Chichester, West Sussex, England: John Wiley and Sons.

Luojus, K., Pulliainen, J., and Derksen, C. (2009). *Snow Water Equivalent (SWE) product guide* (European Space Agency Study Contract Report ESRIN Contract 21703/08/I-EC). Retrieved from Water Resources Management Division Department of Environment and Conservation Government of Newfoundland and Labrador via Keith Abbott

Maidment, D (Ed). (1993). *Handbook of Hydrology*. United States: McGraw-Hill Inc.

Martinec, J. (1975). Snowmelt-Runoff Model for stream flow forecasts. *Nordic Hydrology*. 6(3), 145-154.

Martinec, J., Rango, A., and Major, E. (1983). *The Snowmelt-Runoff Model (SRM) User's Manual*. NASA Reference Publ. 1100, Washington, D.C., USA.

Martinec, J., and Rango, A. (1986). Parameter Values for Snowmelt Runoff Modelling. *Journal of Hydrology*. 84, 197-219.

Martinec, J., Rango, A., and Roberts, R. (2008). *The Snowmelt-Runoff Model (SRM) User's Manual* (updated Edition for Windows). USDA Jornada Experimental Range. New Mexico State University, Las Cruces, NM, USA.

McCuen, R. (1998). *Hydrologic Analysis and Design (2nd ed.)*. Upper Saddle River, NJ: Prentice Hall.

Montgomery, D. C. (2001). *Design and Analysis of Experiments (5th ed.)*. New York, NY: John Wiley and Sons Inc.

National Aeronautics and Space Administration. (1999). *MODIS Specifications* [Fact sheet]. Retrieved from http://modis.gsfc.nasa.gov/about/specifications.php

National Aeronautics and Space Administration WIST. (2010). Data Download. Retrieved on August 2009 from https://wist.echo.nasa.gov

National Aeronautics and Space Administration WIST. (2000). EOSDIS Glossary: Granule. Retrieved on March 8, 2010 from https://wist.echo.nasa.gov

Natural Resources Conservation Service. (2010). Oregon Snow Survey Program. Retrieved on April 4, 2010 from http://www.or.nrcs.usda.gov/snow/index.html

Rango, A., and van Katwijk, V. (1990). Development and Testing of a Snowmelt-Runoff Forecasting Technique. *Water Resources Bulletin: American Water Resources Association*. 26(1), 135-144.

Rees, W. G. (2006). *Remote Sensing of Snow and Ice*. Boca Raton, FL: Taylor and Francis Group.

Riggs, G. A., Hall, D. K., and Salomonson, V. V. (2006). MODIS Snow Products: User Guide to Collection 5. Retrieved from The National Snow and Ice Data Center, MODIS v5: http://nsidc.org/data/docs/daac/modis_v5/dorothy_snow_doc.pdf

ScienceLearn. (2007). Electromagnetic Spectrum (image). Retrieved on March 20, 2010 from http://www.sciencelearn.org.nz

Seidel, K., and Martinec, J. (2004). *Remote Sensing in Snow Hydrology: Runoff Modelling, Effect of Climate Change*. Chichester, UK: Praxis Publishing Ltd.

Statistics Canada. (2006). Population Census for Corner Brooke and Deer Lake. www12.statcan.ca/census-recensement/2006/

The National Snow and Ice Data Center. (2008). MODIS Data. University of Colorado at Boulder. Retreived on March 8, 2010 from http://www.nsidc.org

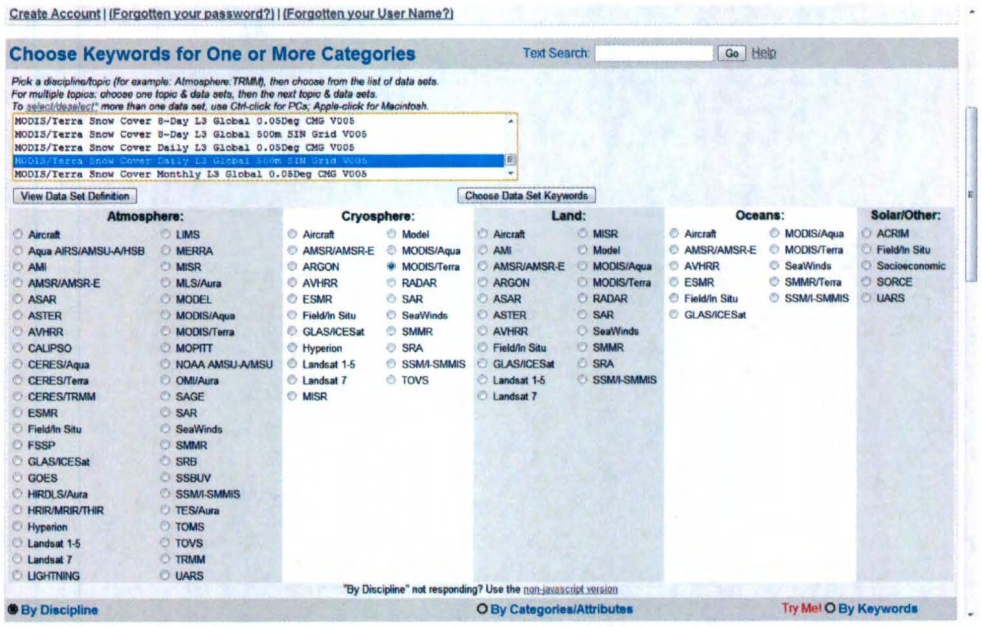
Water Resources Management Division, Department of Environment and Conservation Newfoundland and Labrador Government. *Snow Cover Survey 2009: Upper Humber River Watershed.* (2009). St. John's, NL: Abbott, K.

Watt, W. E (Ed). (1989). *Hydrology of Floods in Canada: A Guide to Planning and Design*. Ottawa, ON: National Research Council of Canada.

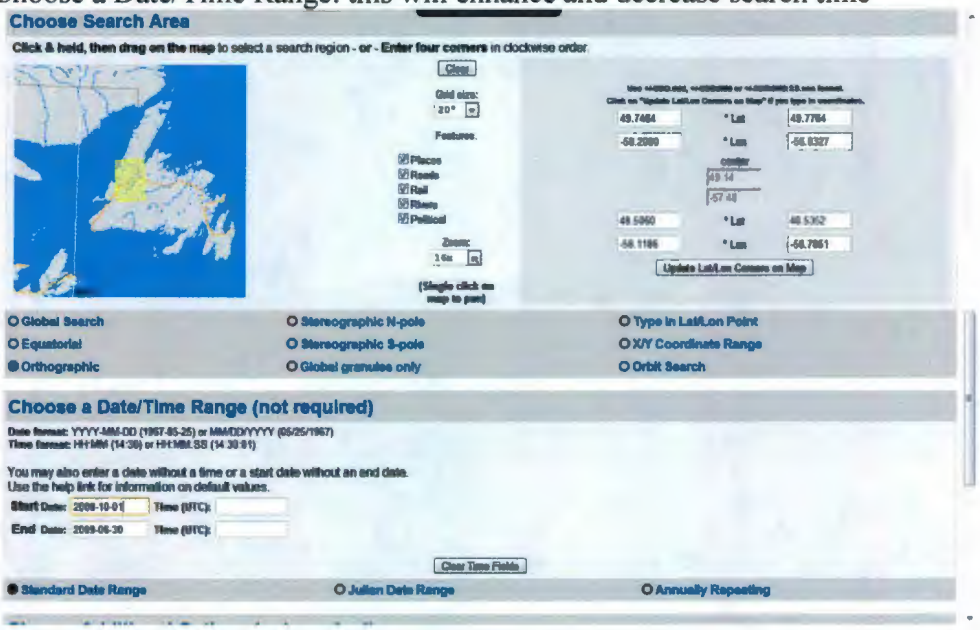
Appendix A

Sequence of Steps to Correct MODIS/Terra Data Download

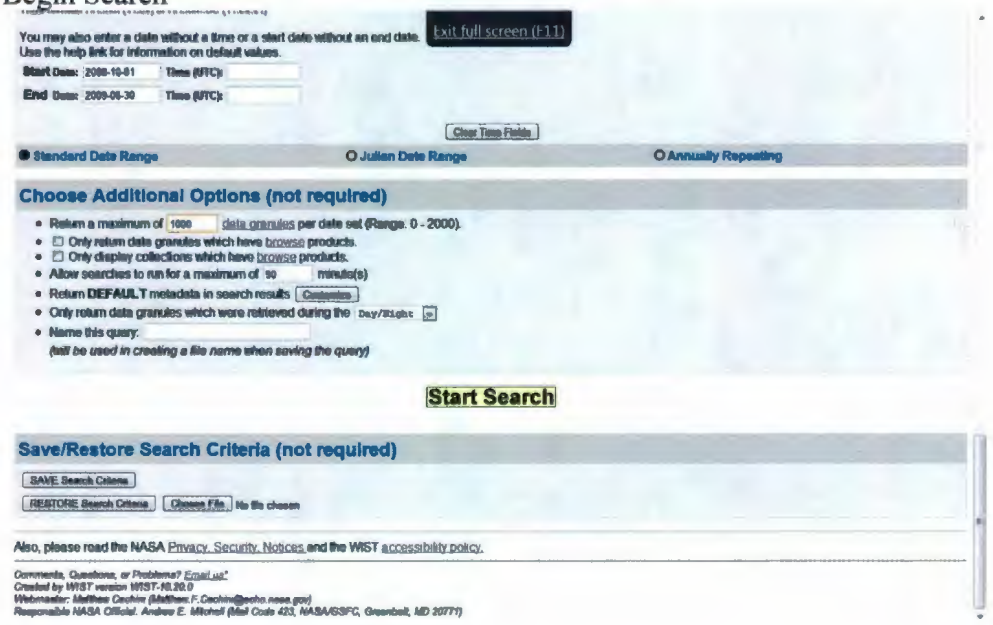
Website: <https://wist.echo.nasa.gov>
WIST: data search page
Choose discipline/topic first: Cryosphere MODIS/Terra
Choose appropriate data set: MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid V005



WIST: data search page
Choose Search Area: be as specific as possible, use zoom.
Choose a Date/Time Range: this will enhance and decrease search time



WIST: Data Search Page
Choose Addition Options: Set Maximum Data Granules per Data Set to 1000
Begin Search

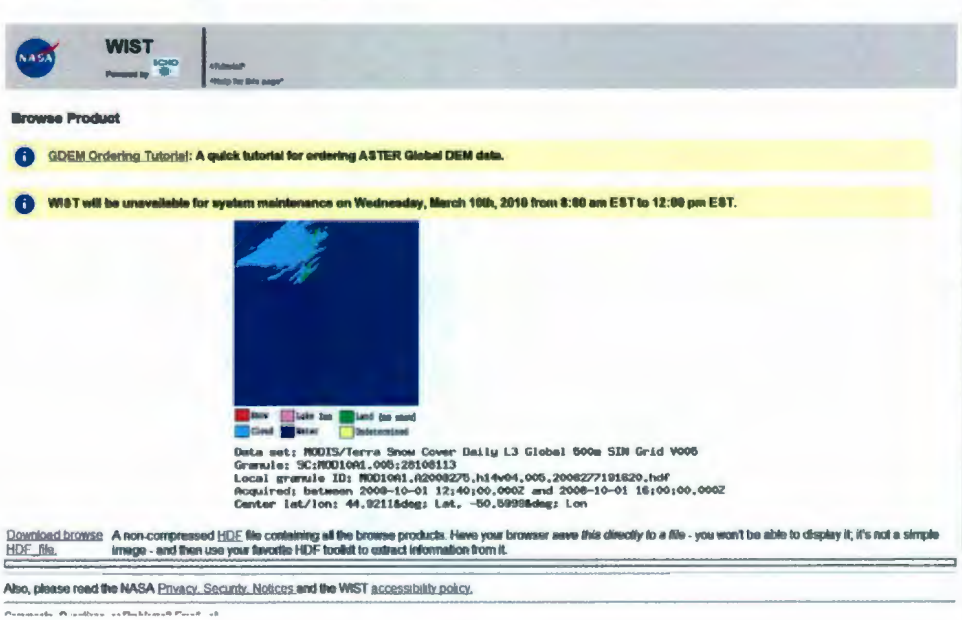


WIST Search Results:

Select	Data Granule ID (Local Granule ID)	Granule Information	On-line Access	Image Quicklook	Special Processing Links	Data Center	Start Date	Stop Date
	SC:MOD10A1.005:28108113 (MOD10A1.A2008275.h14v04.005.2008277191620.hdf)	Attributes	Data* Metadata*	Browse	Guide Document for this product at NSIDC MODIS Snow and Ice Product page at NSIDC	NSIDC_ECS	01 Oct 2008, 12:40:00.000	01 Oct 2008, 16:00:00.000
	SC:MOD10A1.005:28144801 (MOD10A1.A2008276.h14v04.005.2008279062329.hdf)	Attributes	Data* Metadata*	Browse	Guide Document for this product at NSIDC MODIS Snow and Ice Product page at NSIDC	NSIDC_ECS	02 Oct 2008, 13:20:00.000	02 Oct 2008, 15:05:00.000
	SC:MOD10A1.005:28162965 (MOD10A1.A2008277.h14v04.005.2008280062757.hdf)	Attributes	Data* Metadata*	Browse	Guide Document for this product at NSIDC MODIS Snow and Ice Product page at NSIDC	NSIDC_ECS	03 Oct 2008, 14:00:00.000	03 Oct 2008, 15:45:00.000
	SC:MOD10A1.005:28200863 (MOD10A1.A2008278.h14v04.005.2008284163516.hdf)	Attributes	Data* Metadata*	Browse	Guide Document for this product at NSIDC MODIS Snow and Ice Product page at NSIDC	NSIDC_ECS	04 Oct 2008, 13:05:00.000	04 Oct 2008, 16:25:00.000
	SC:MOD10A1.005:28281940 (MOD10A1.A2008279.h14v04.005.2008285152115.hdf)	Attributes	Data* Metadata*	Browse	Guide Document for this product at NSIDC MODIS Snow and Ice Product page at NSIDC	NSIDC_ECS	05 Oct 2008, 13:50:00.000	05 Oct 2008, 15:35:00.000
	SC:MOD10A1.005:28293864 (MOD10A1.A2008280.h14v04.005.2008286124417.hdf)	Attributes	Data* Metadata*	Browse	Guide Document for this product at NSIDC MODIS Snow and Ice Product page at NSIDC	NSIDC_ECS	06 Oct 2008, 12:55:00.000	06 Oct 2008, 16:15:00.000

Image Quicklook from October 1st 2008

Entire island of Newfoundland almost completely covered in cloud



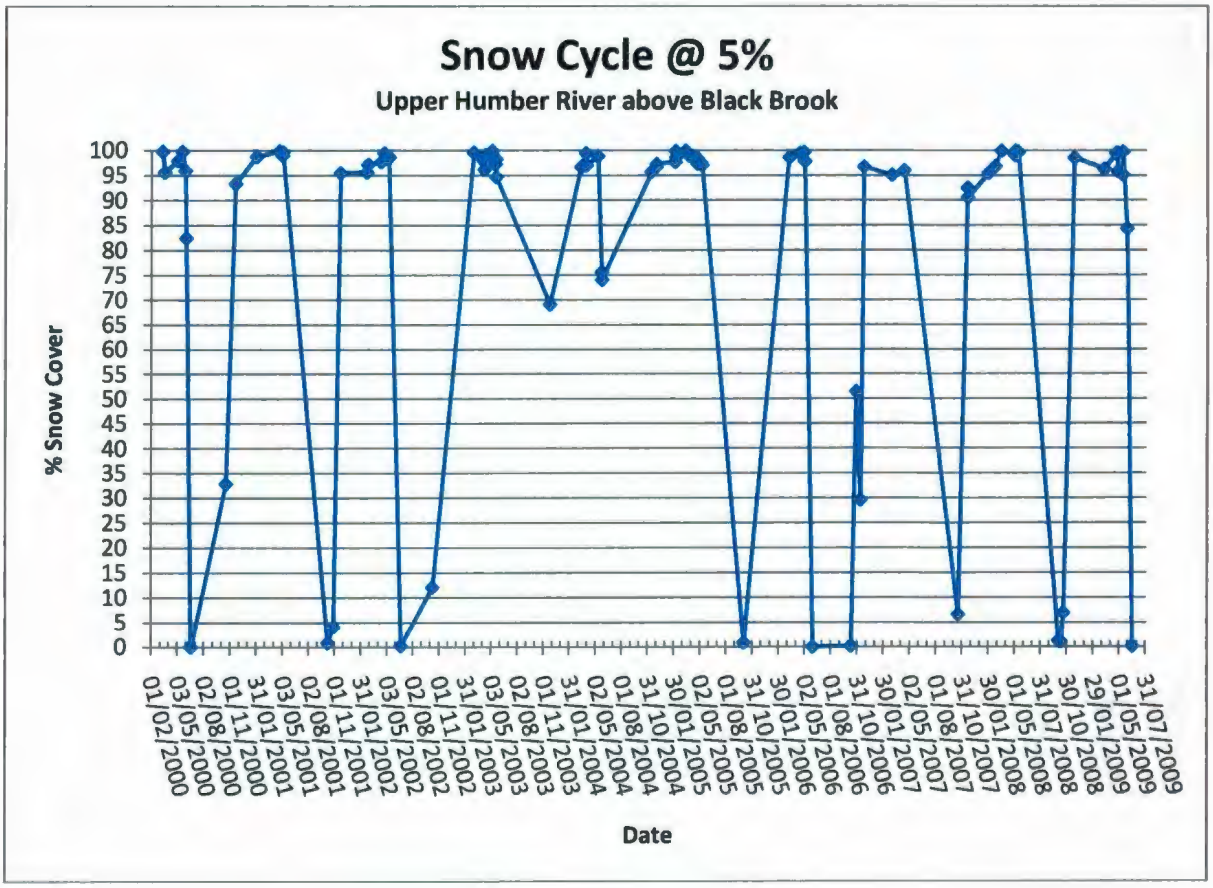
Appendix B

Plot Comparison for Setting Cloud Cover Threshold

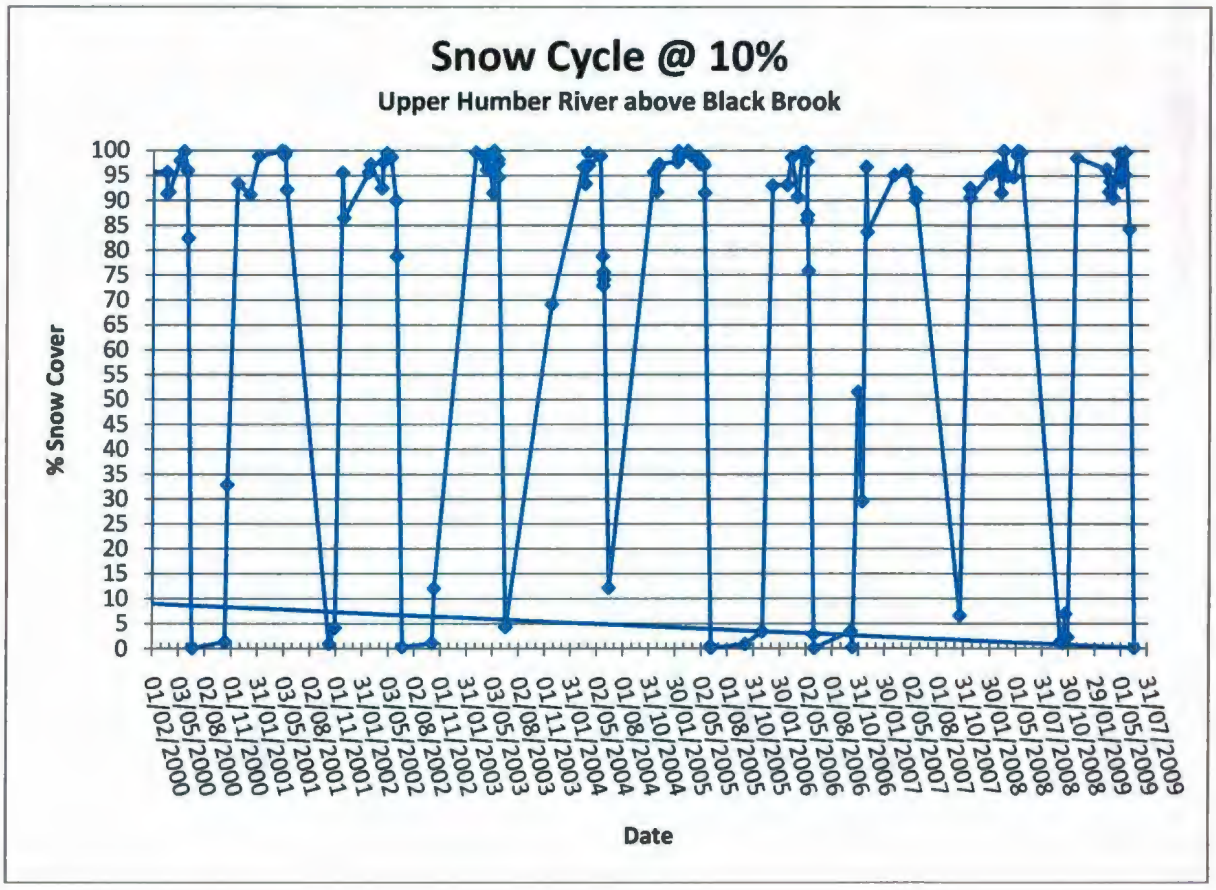
		Ten Year Period from 2000 to 2009										
		'00	'01	'02	'03	'04	'05	'06	'07	'08	'09	% Data
Cloud Cover	5%	10	7	7	9	9	10	9	5	10	8	3.1%
	10%	14	10	11	11	14	14	18	7	14	12	4.6%
	20%	19	14	17	14	21	16	20	15	20	14	6.2%
	30%	21	18	22	19	26	21	25	23	24	17	7.9%

As expected, as the cloud cover threshold increases, the number of snow cover points also increases. This table represents the amount of snow cover data available for a given % of acceptable cloud cover.

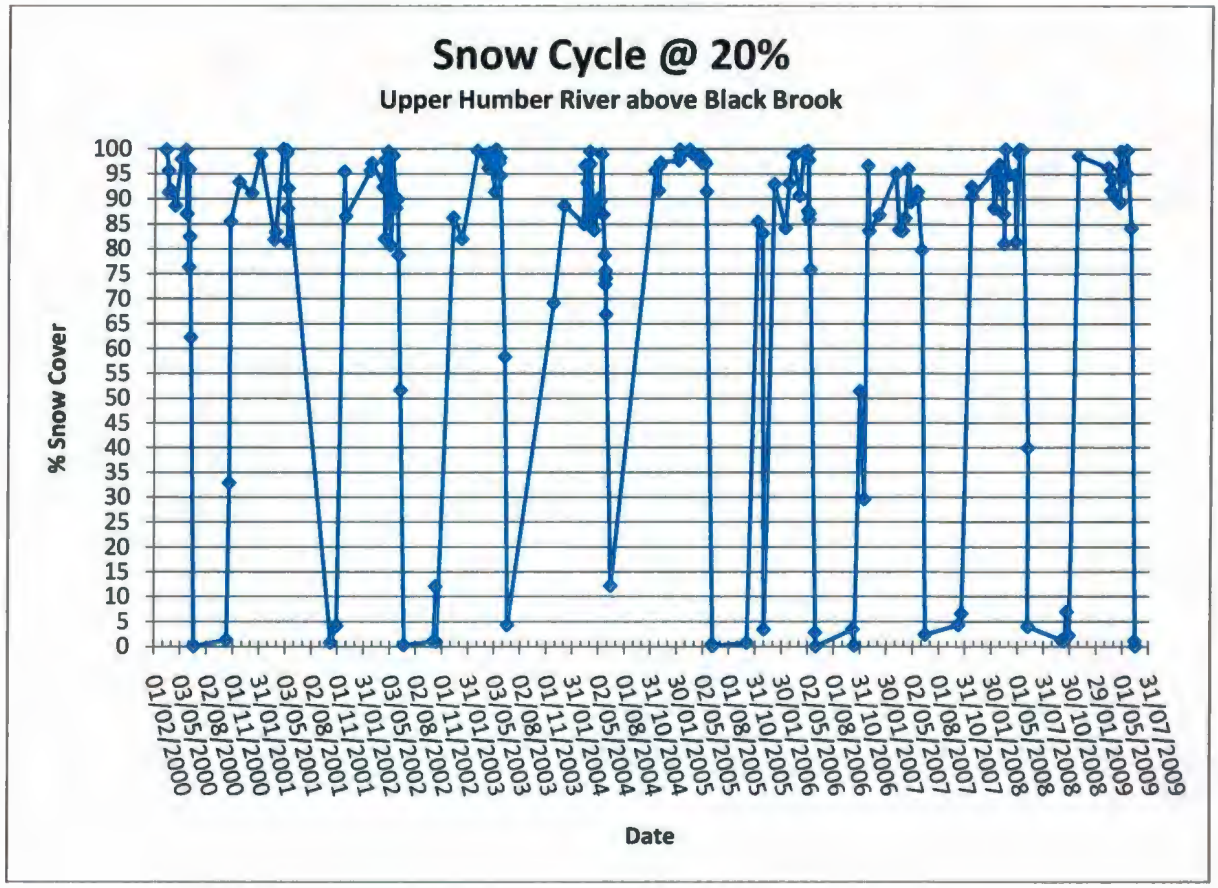
Plot of Snow Cover Data with 5% Cloud Cover Threshold
(Notice how in some years the percent snow cover cycle does not complete itself back to 0% coverage due to a lack of points.)



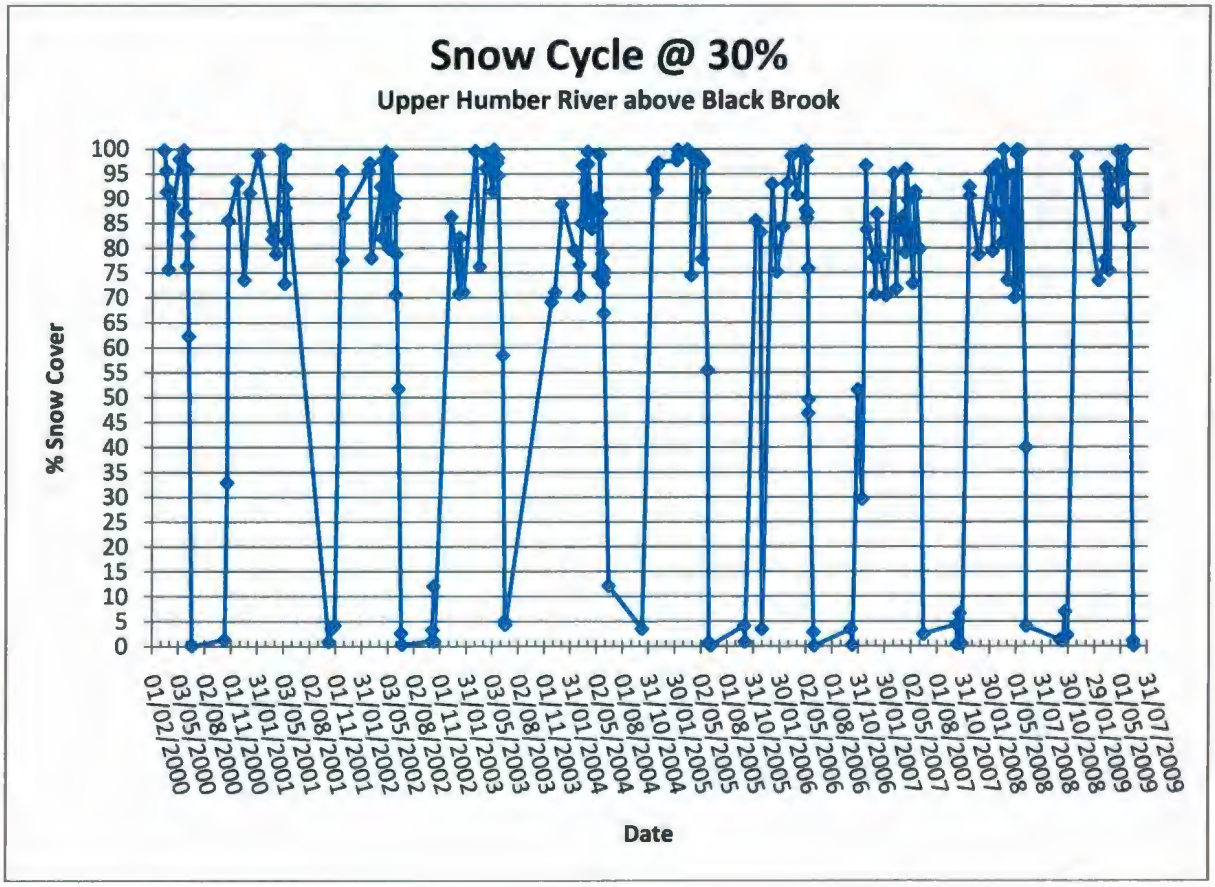
Plot of Snow Cover Data with 10% Cloud Cover Threshold



Plot of Snow Cover Data with 20% Cloud Cover Threshold



Plot of Snow Cover Data with 30% Cloud Cover Threshold



Upon cloud cover threshold analysis from overlapping the plots and discovering when the time and rapidity of decline changed significantly from the previous set threshold, a cloud cover threshold of 20% was used. This 20% threshold provided the proper balance between number of data set points and sufficient accuracy.

Appendix C

EASI Script Used to Automate the Extraction of Snow Cover Information

```
!-------------------------------------------------------------------!
!-------------------------------------------------------------------!
!                                                                   !
! Percent snow cover extraction from multiple NSIDC files [snowcov4.eas]   !
!                                                                   !
!                                                                   !
! This script was written to extract percent snow cover from multiple      !
! National Snow & Ice Data Center product files (*.hdf).                   !
!                                                                   !
! This script assumes that all the input files will all be located         !
! within a given directory, all the files will be of the same format       !
! and that the output directory does not contain any files.                !
!                                                                   !
!-------------------------------------------------------------------!
!-------------------------------------------------------------------!

!-------------------------------------------------------------------
! Define variables
!-------------------------------------------------------------------

!for input and output directory
 local string in_files, out_files

!for directory listing of the input directories
 local mstring dirlist

!for the file format and extension types
 local string type, ext

!file names
 local string waterbit, bn, fn, fn2, fn3, fn4

 local $Z

! FOR loop parameters
 local integer i, j

 local string confirm
```

```
!-------------------------------------------------------------------
! Clear the EASI window and then show the header information
!-------------------------------------------------------------------

 PRINT @(1 ,1,CLREOS)
 print "-------------------------------------------------------------"
 print @reverse,"              Extract Percent Snow Cover                    ",@alloff
 print ""
 print "This script assumes that all the input files will all be located"
 print "within a given directory, the files will be of the same format,"
 print "will be clipped to the same extents, and the output directory will"
 print "not contain any files."
 print ""
 print "-------------------------------------------------------------"

!-------------------------------------------------------------------
! Collect input from user
!-------------------------------------------------------------------
ASKAGAIN: \
 print ""
 print ""
 print "Enter the directory that contains the input files (e.g. C:\snow cover\input):"
 input ">" in_files
 print ""
 print "Enter the directory for the output files (e.g. C:\snow cover\output):"
 input ">" out_files
 print ""
 print "Enter the file format of the files (3-letter file extension; e.g. hdf): "
 input ">" type
 print ""
 print "Enter the path and filename of the PIX file containing the watershed bitmaps:"
 print "(e.g. C:\snow cover\Watershed_Bitmaps2.pix)"
 input ">" waterbit
 print ""
 PRINT @(1 ,1,CLREOS)

!-------------------------------------------------------------------
! Confirm with user to ensure that the parameters are correct
! If they are correct then continue with the script and if they are not
! then run the script over again.
!-------------------------------------------------------------------

 print "-------------------------------------------------------------"
 print ""
 print "The input directory you specified was:"
 print " ", in_files
 print ""
 print "The output directory you specified was:"
```

```
print "", out_files
print ""
print "The file format you specified was:"
print "", type
print ""
print "The path and filename for the Watershed bitmaps you specified was:"
print "", waterbit
print ""
print "-------------------------------------------------------------------"
print "Are these parameters correct? (Y/N)"
print ""
input ">" confirm

if (confirm ~= "y" or confirm ~= "Y" ) then

        !-------------------------------------------------------------
        ! Get the contents of the directory
        !-------------------------------------------------------------

        dirlist = getdirectory(in_files)
        let $Z = "\

        for i = 1 to f$len(dirlist)

        !-------------------------------------------------------------
        ! Extract parts of the filenames
        !-------------------------------------------------------------

        fn = in_files + $Z + dirlist[i]
         ext = getfileextension(fn)
         bn = getfilebasename(fn)
        fn2 = out_files + $Z + bn[10] + bn[11] + bn[12] + bn[13] + "-" + bn[14] + bn[15] +
bn[16]
        fn3 = out_files + $Z + "GH" + $Z + bn[10] + bn[11] + bn[12] + bn[13] + "-" +
bn[14] + bn[15] + bn[16]
        fn4 = out_files + $Z + "UH" + $Z + bn[10] + bn[11] + bn[12] + bn[13] + "-" +
bn[14] + bn[15] + bn[16]
        if (ext ~= type) then

         print ""
         print "Calculate % snow cover for:", bn, ".", ext
         print ""

        !-------------------------------------------------------------
        ! Set up the parameters and execute the FEXPORT command
        !-------------------------------------------------------------

         fili = fn
```

```
filo = fn2 + ".pix"
dbiw = 1,1,1240,890
dbic = 1
dbib =
dbvs =
dblut =
dbpct =
ftype = "PIX"
foptions =

R Fexport

!-----------------------------------------------------------------
! Add bitmaps from Watershed_Bitmaps.pix to the exported PIX file
!-----------------------------------------------------------------

fili = waterbit
filo = fn2 + ".pix"
dbib = 2,3
dbob =
dbiw =
dbow =
report = "OFF"
Monitor = "ON"

R iib


!-----------------------------------------------------------------
! Compute area of landcover classes under watershed masks
!-----------------------------------------------------------------


        FOR j = 2 TO 3 BY 1

        FILE = fn2 + ".pix"
        DBIC = 1
        DBIB = j
        UNITS = "Square Kilometers"

        IF j=2 THEN
                REPORT = fn3 + ".txt"
        ELSE
                REPORT = fn4 + ".txt"
        ENDIF
        MONITOR = "ON"

        r AREAREPORT
```

```
                ENDFOR

            PRINT @(1 ,1,CLREOS)


           endif
          endfor


    else

     goto ASKAGAIN

    endif

    PRINT @(1 ,1,CLREOS)
    print "-----------------------------------------------------------------"

    print ""
    print "The hdf files are stored in the following directory:"
    print " ", out_files
    print ""
    print ""
    print @reverse," Percent snow cover extraction EASI Script Finished ",@alloff
    print ""
    print "-----------------------------------------------------------------"

  return
```

Appendix D

Visual Basic Script: to Import, Amalgamate, and Manage all Daily Individual Output Text Files

A variation of this script was used to grab various data from the text files. This script searched for snow (pixel 200), lake ice (pixel 100), and cloud (pixel 50).

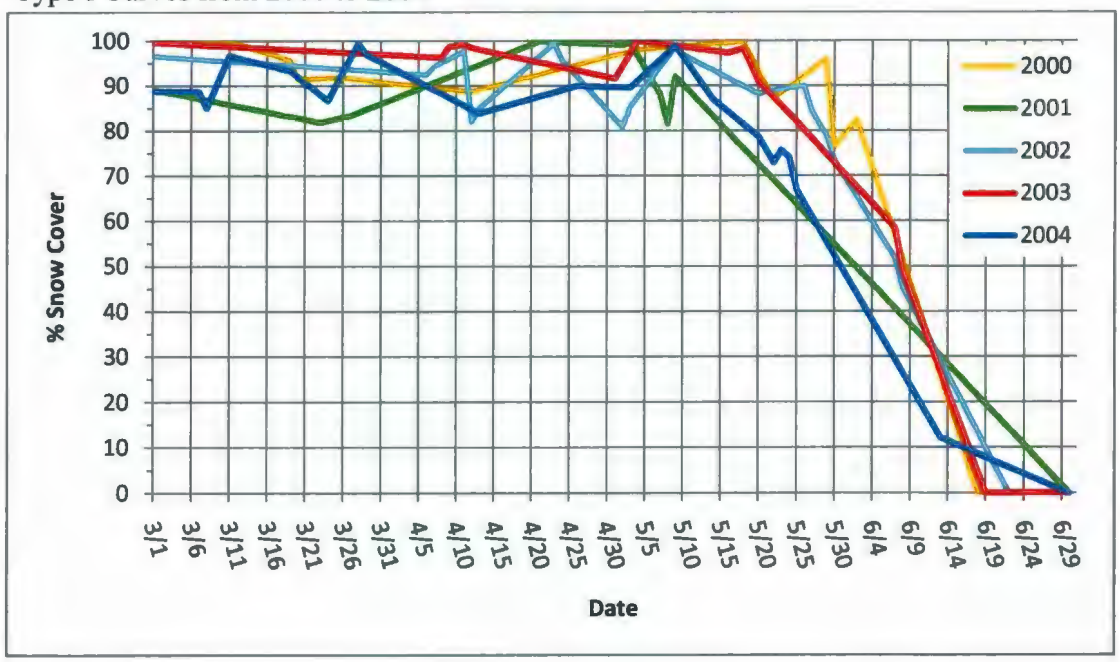
```
Sub test()
Dim myDir As String, fn As String, temp As String, delim As String, a() As String
Dim i As Long, e, n As Long, t As Long, x, m As Object
myDir = "C:\Output Complete\Upper Humber River above Black Brook" '<- change here (folder path)
On Error Resume Next
fn = Dir(myDir & "\200*-*.txt")
ReDim a(1 To 10000, 1 To 100)
Do While fn <> ""
  temp = CreateObject("Scripting.FileSystemObject").OpenTextFile(myDir & "\" & fn).ReadAll
  x = Split(temp, vbCrLf)
  With CreateObject("VBScript.RegExp")
    .Pattern = "^Pixel"
    .IgnoreCase = True
    For Each e In x
      If .test(e) Then flg = True
      If flg Then
        If (InStr(e, "200") = 1) + (InStr(e, "100") = 1) + (InStr(e, "50") = 1) Then
          n = n + 1: a(n, 1) = fn
          .Pattern = "\d+(\.\d+)?"
          .Global = True
          t = 1
          For Each m In .Execute(e)
            t = t + 1: a(n, t) = m.Value
          Next
        End If
      End If
    Next
  End With
```

```
    flg = False
    fn = Dir
Loop
If n > 0 Then
    Sheets(1).Cells(1).Resize(n, 100).Value = a
End If
End Sub
```

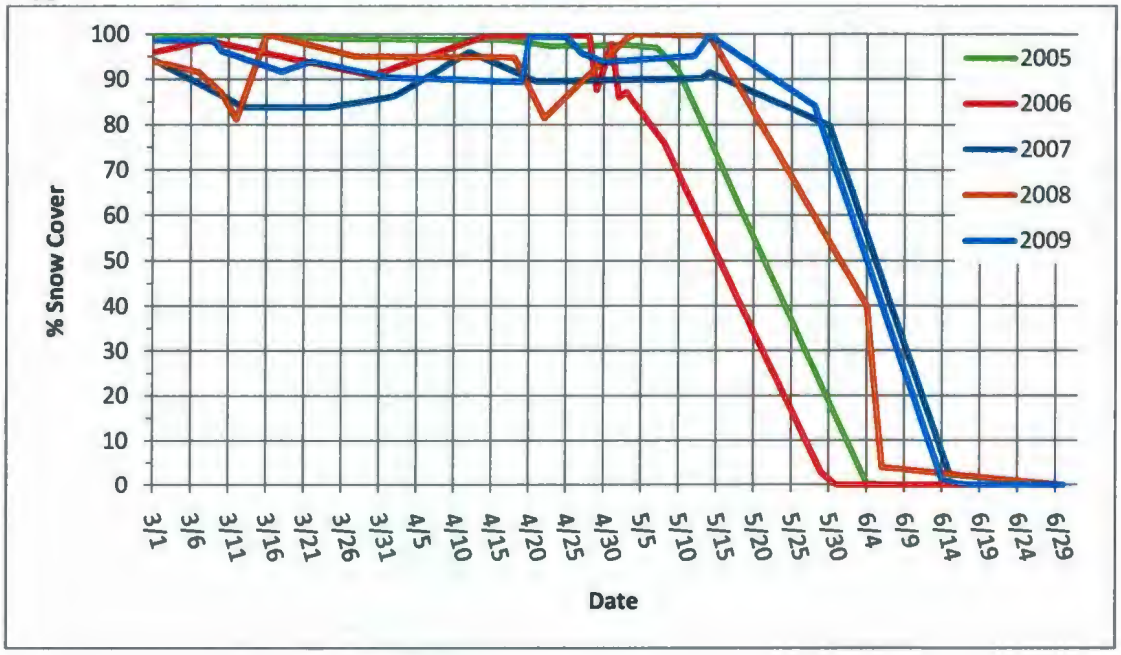
Appendix E

Type I and Type II Depletion Curves For All Snowmelt Periods: 2000 to 2009

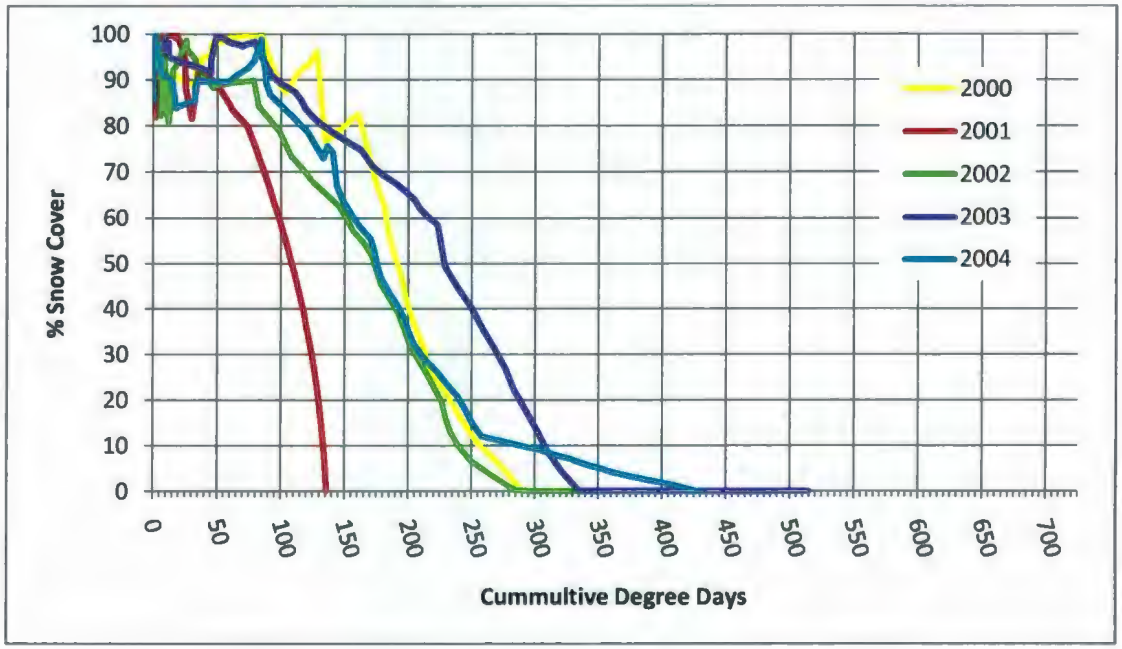
'Type I Curves from 2000 to 2004'



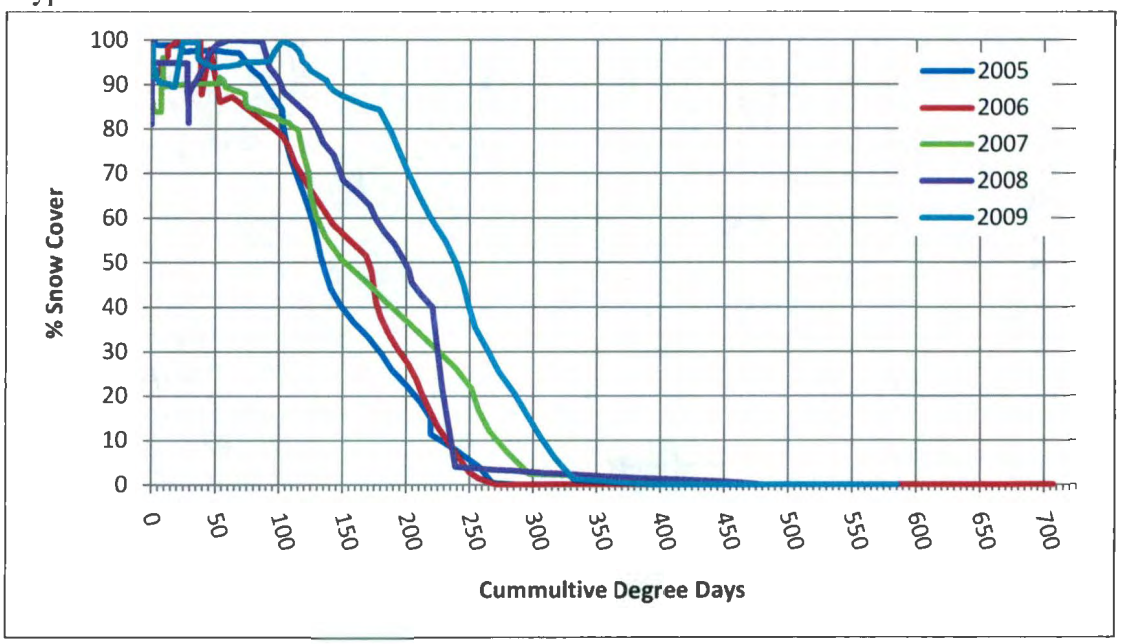
'Type I Curves from 2005 to 2009'



'Type II Curves from 2000 to 2004'



'Type II Curves from 2005 to 2009'



Appendix F

Performing a Design of Experiments Analysis on Four Factors for Martinec's Snowmelt Runoff Model

Preliminary analysis:

DOE 2^4 Parameter Run Order (with optimum Nash-Sutcliffe E trial parameter combination highlighted)

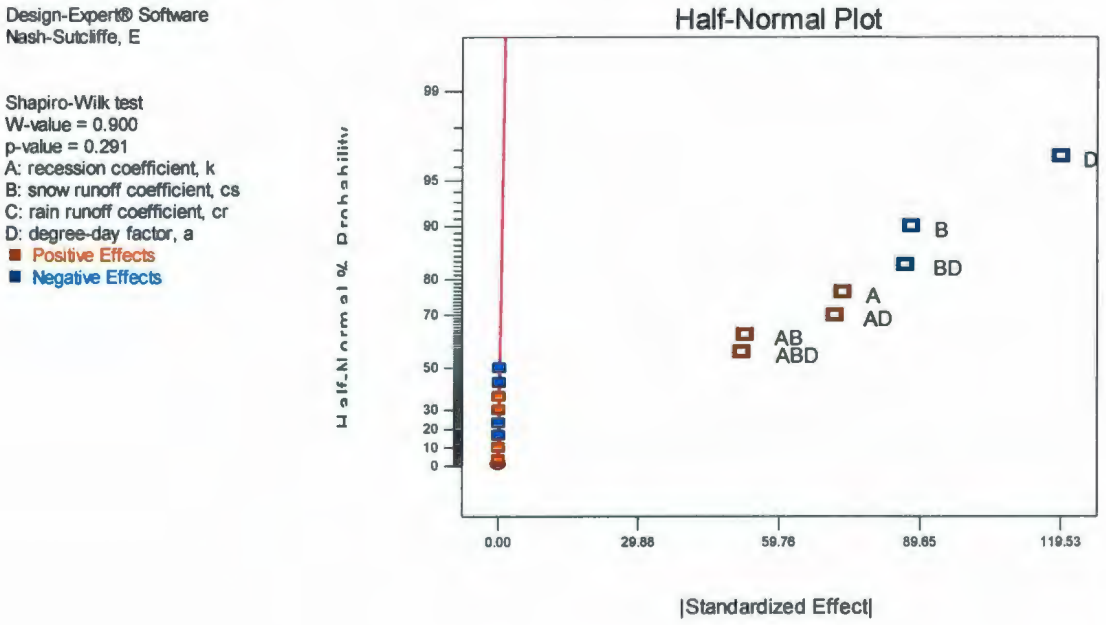
Std	Run	Block	Factor 1 A:recession coefficient, k	Factor 2 B:snow runoff coefficient, cs	Factor 3 C:rain runoff coefficient, cr	Factor 4 D:degree-day factor, a	Response 1 Nash-Sutcliffe, E
3	1	Block 1	0.20	0.70	0.30	1.00	-2.37
5	2	Block 1	0.20	0.30	0.70	1.00	-0.32
9	3	Block 1	0.20	0.30	0.30	8.00	-52.89
7	4	Block 1	0.20	0.70	0.70	1.00	-2.46
6	5	Block 1	0.60	0.30	0.70	1.00	0.6
1	6	Block 1	0.20	0.30	0.30	1.00	-0.3
11	7	Block 1	0.20	0.70	0.30	8.00	-331.15
13	8	Block 1	0.20	0.30	0.70	8.00	-53.45
12	9	Block 1	0.60	0.70	0.30	8.00	-82.28
16	10	Block 1	0.60	0.70	0.70	8.00	-82.64
14	11	Block 1	0.60	0.30	0.70	8.00	-12.76
10	12	Block 1	0.60	0.30	0.30	8.00	-12.62
2	13	Block 1	0.60	0.30	0.30	1.00	0.59
15	14	Block 1	0.20	0.70	0.70	8.00	-332.6
8	15	Block 1	0.60	0.70	0.70	1.00	0.04
4	16	Block 1	0.60	0.70	0.30	1.00	0.06

Parameter Settings:

Factors	Low	High
Recession Coefficient, k (A)	0.2	0.6
Snow Runoff Coefficient, Cs (B)	0.3	0.7
Rain Runoff Coefficient, Cr (C)	0.3	0.7
Degree-Day Factor, a (D)	1.0	8.0

DOE Analysis - Estimating the factor effects and determining which effects appear important:

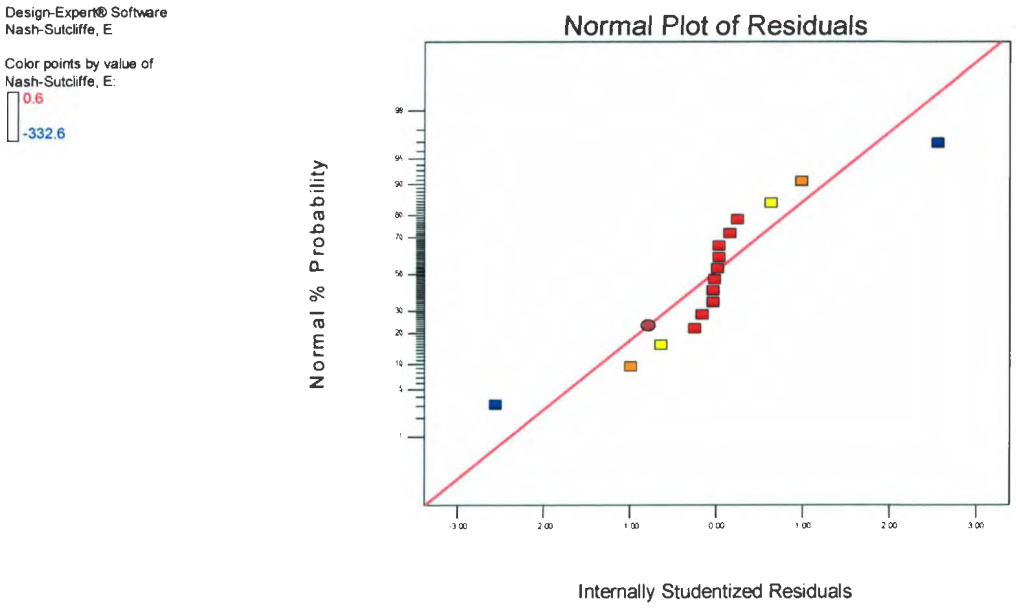
Using Design-Expert 7.1.3 the effects were analyzed. Below is the "Half-Normal Plot" with the effects that appear important. These are factors: A, B, D, AB, AD, BD, and ABD.



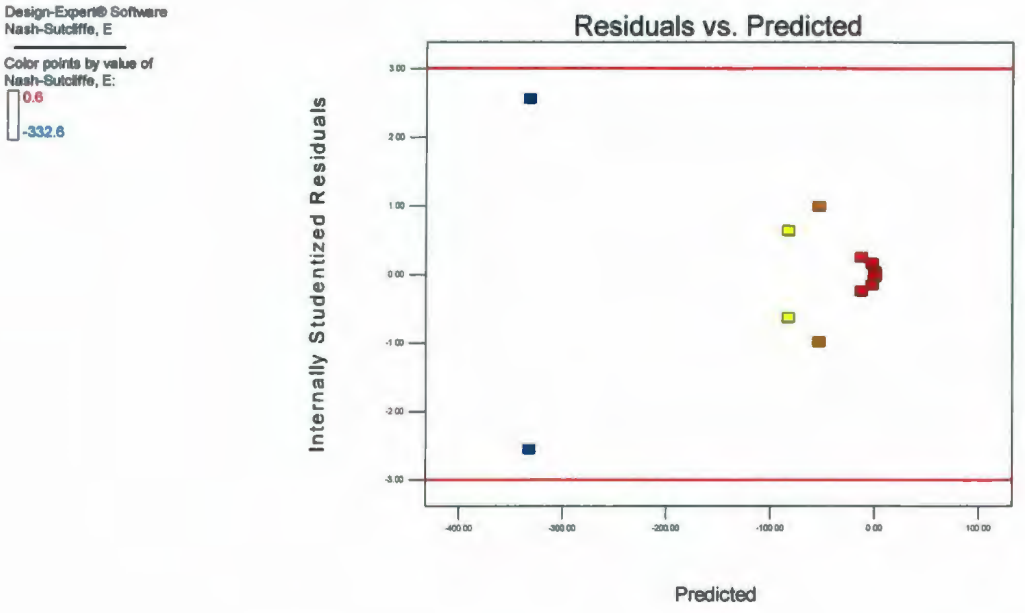
Performing ANOVA:

Checking the assumptions of ANOVA: (1) Normality, (2) Constant variance, and (3) independence:

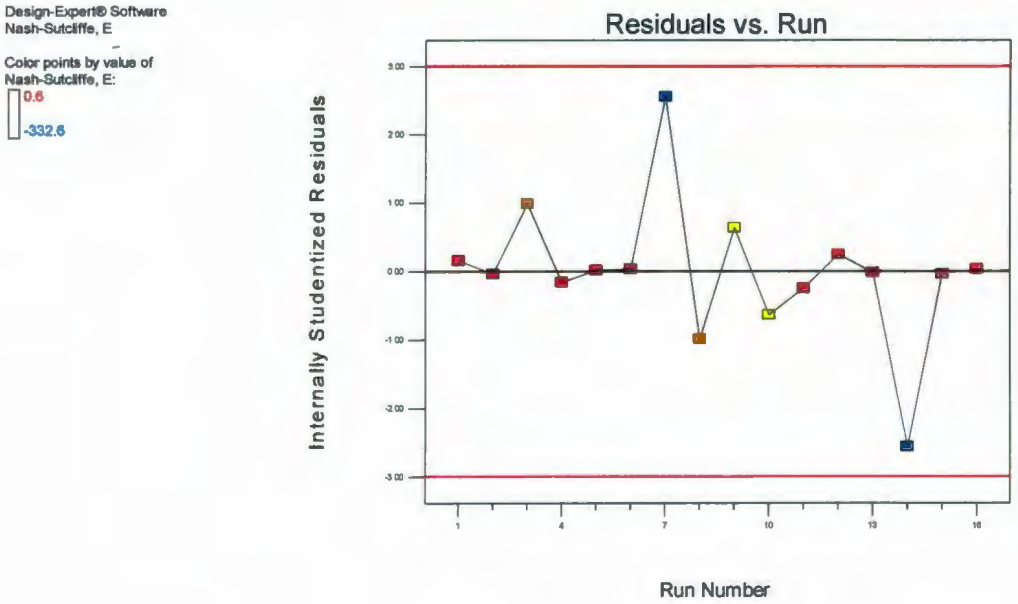
First, the "Normal Plot of Residuals" looks fairly normal. The data points roughly follow the straight line.



Second, the "Residuals vs. Predicted" plot **does not** show a constant variance, there is a funnel shape with the data. This means a transformation should be tested. It should be noted that the lack of points may cause this pattern to look more severe.

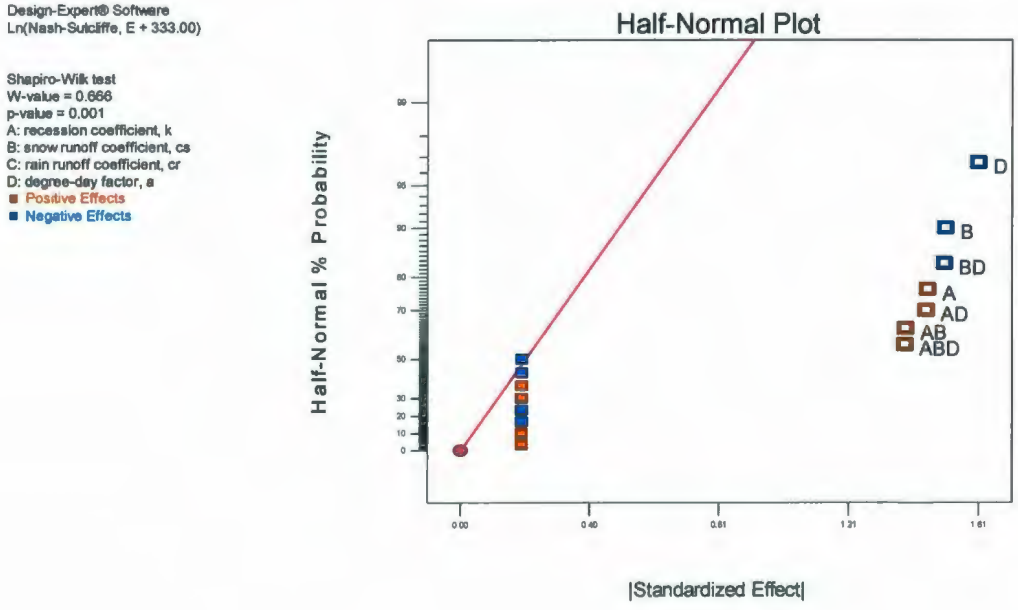


Third, the "Residuals vs. Run" plot does not show any pattern, indicating independence.



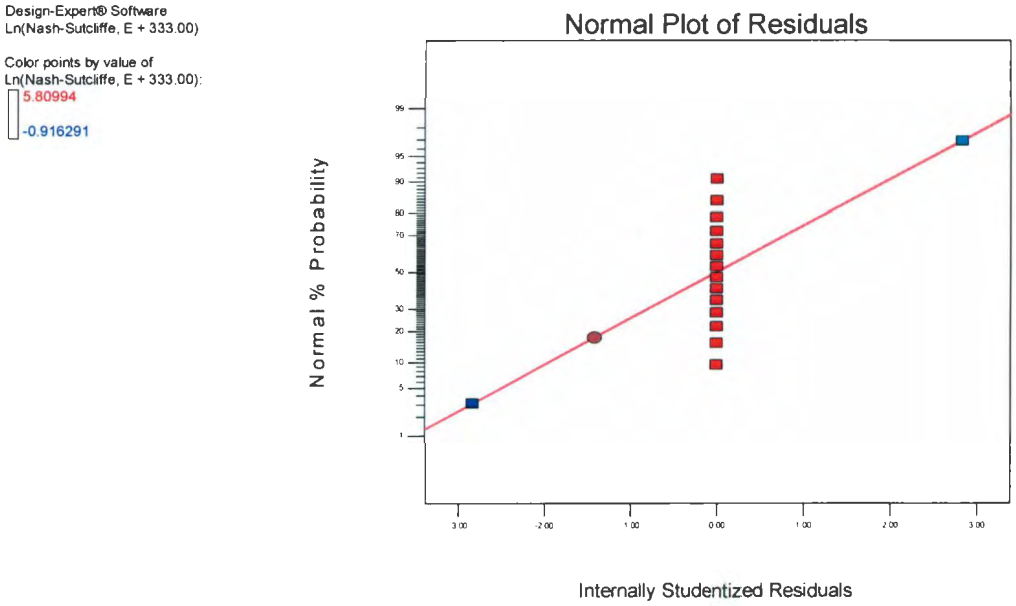
Trying a **NATURAL LOG TRANSFORM**: Note: a constant k = 333 needed to be added to ensure all responses were greater than zero.

Below is the "Half-Normal Plot" with the effects that appear important. These are the same factors as above: A, B, D, AB, AD, BD, and ABD.

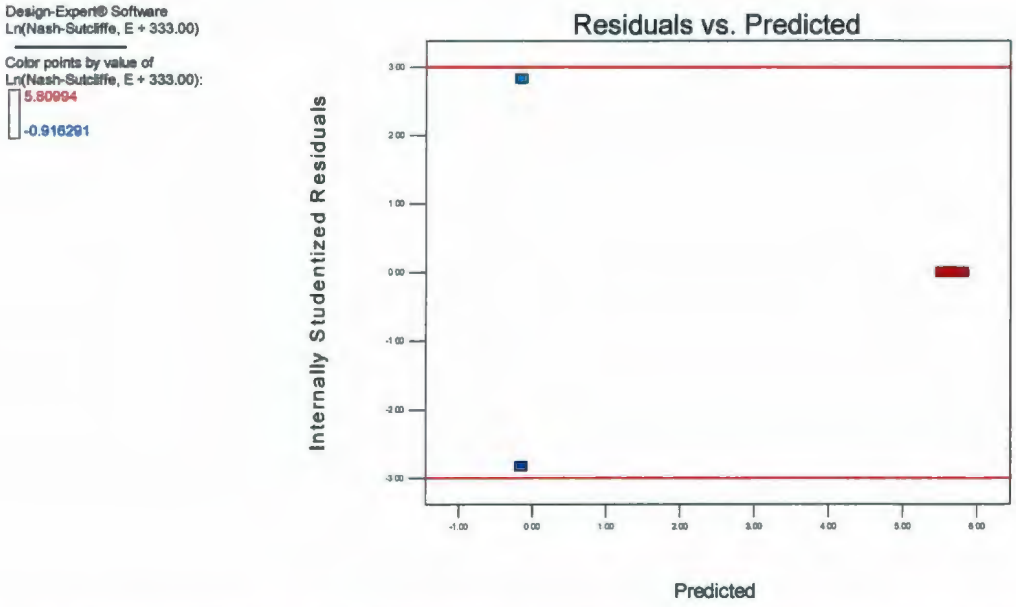


Performing ANOVA: Checking the assumptions of ANOVA: (1) Normality, (2) Constant variance, and (3) independence:

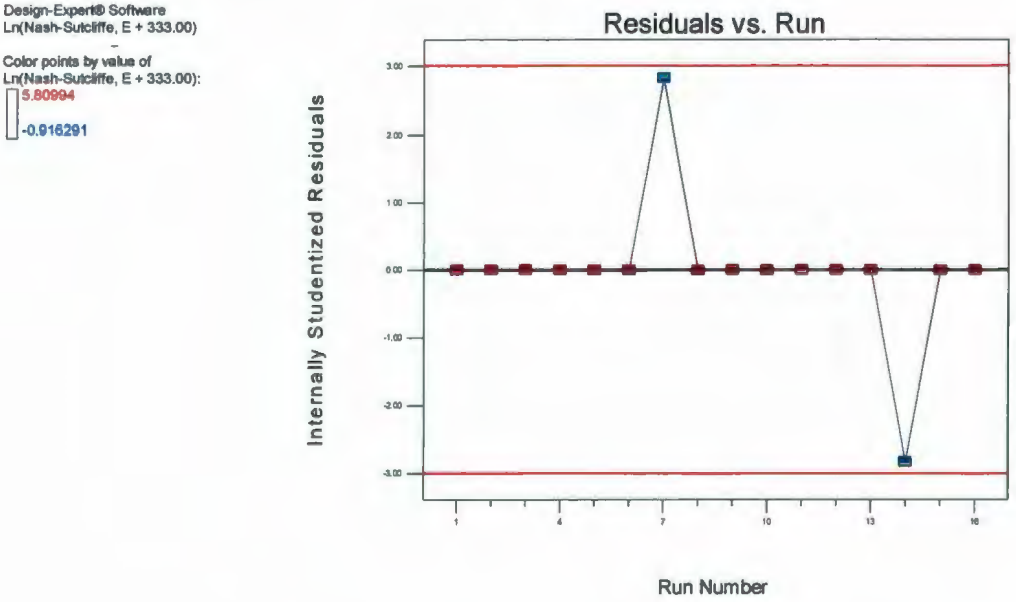
First, the "Normal Plot of Residuals" **does not** look normal.



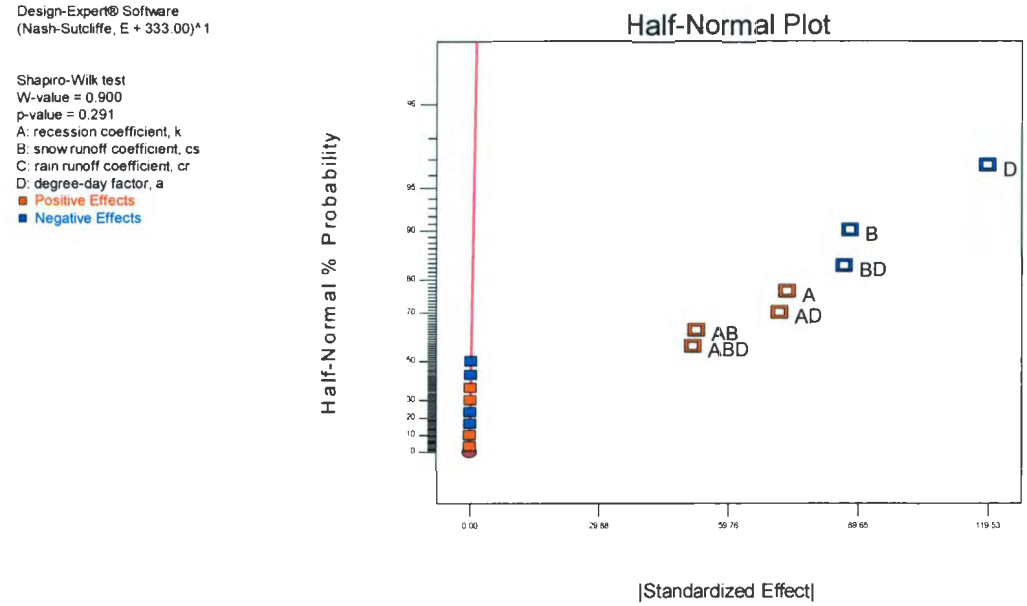
Second, the "Residuals vs. Predicted" plot **does not** show a constant variance, there is a severe funnel shape with the data. This transformation did not help the data.



Third, the "Residuals vs. Run" even seems to show a pattern, something is not right with this natural log transform.

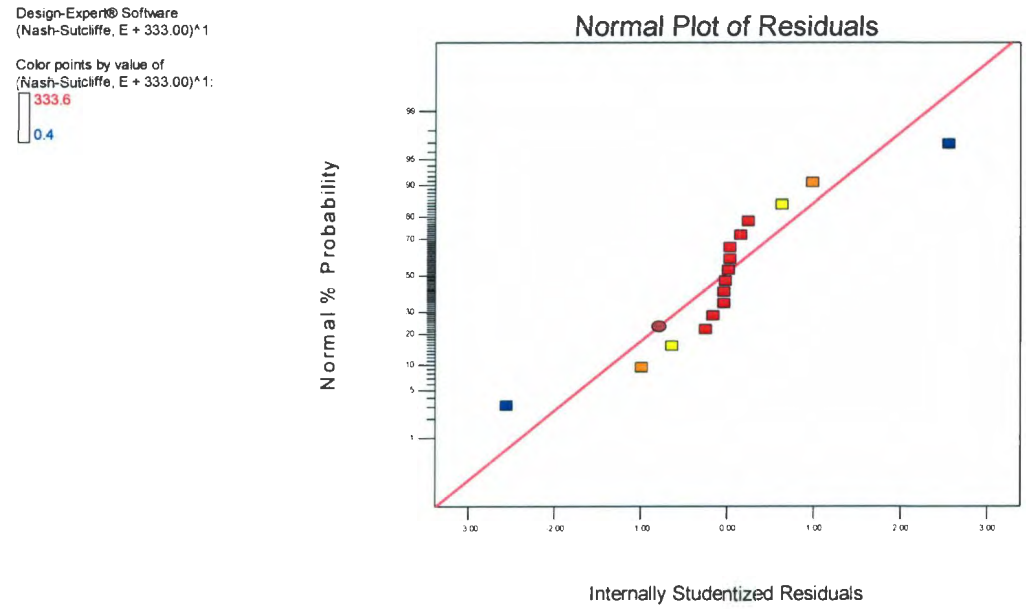


Trying a **POWER TRANSFORM**: Note: a constant k = 333 needed to be added to ensure all responses were greater than zero. Lamda = 1. Below is the "Half-Normal Plot" with the effects that appear important. These are the same factors as above: A, B, D, AB, AD, BD, and ABD.

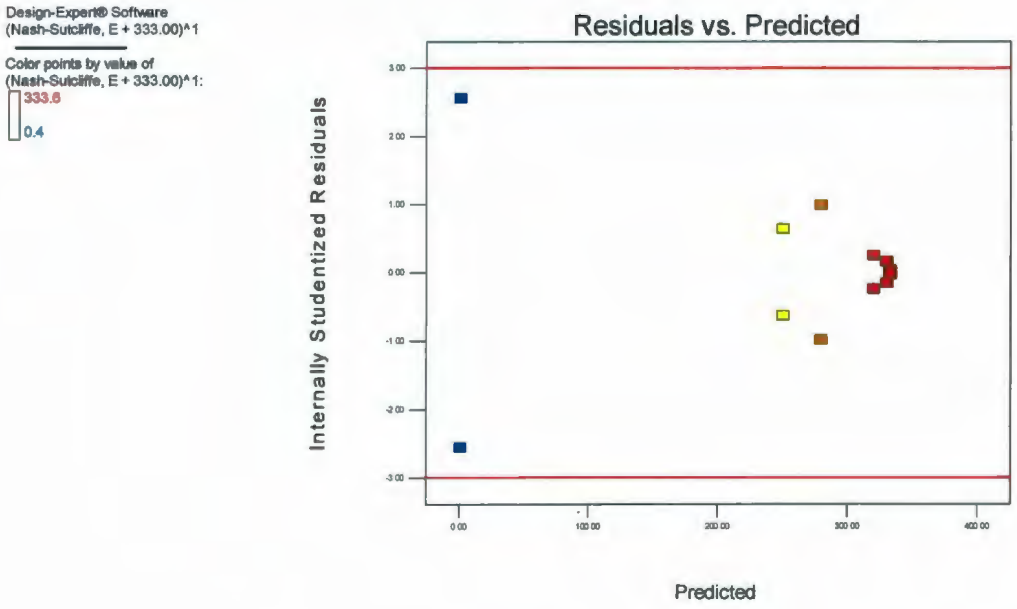


Performing ANOVA: Checking the assumptions of ANOVA: (1) Normality, (2) Constant variance, and (3) independence:

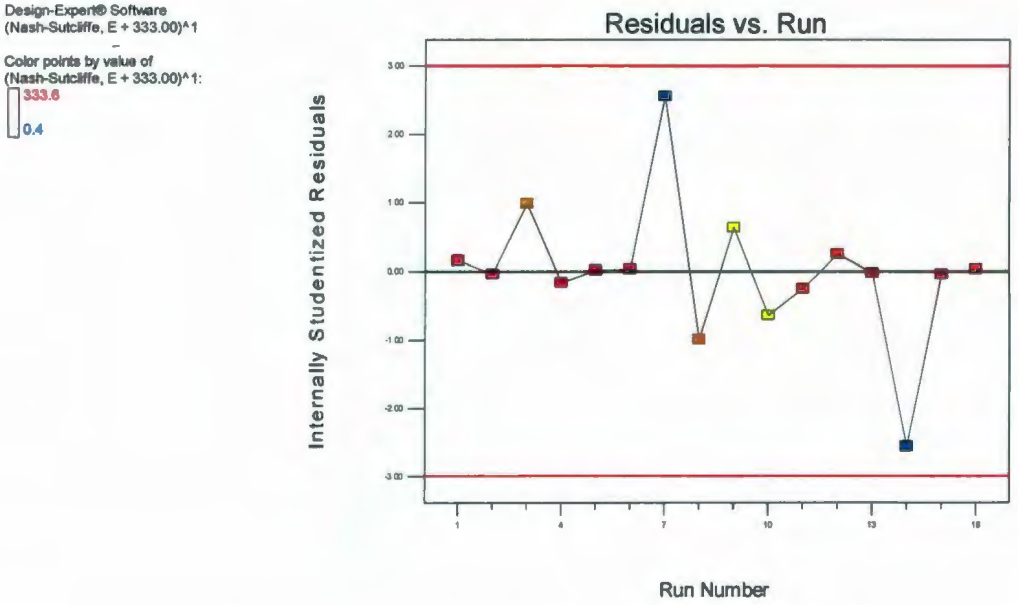
First, the "Normal Plot of Residuals" looks normal - about as normal as the data with no transformation earlier.



Second, the "Residuals vs. Predicted" plot **does not** show a constant variance, there is a funnel shape with the data. This transformation did not help the data- this residuals vs. predicted plot looks very similar to the original plot from the data with no transform.



Third, the "Residuals vs. Run" plot does not show any pattern, indicating independence.



The two transforms: natural log and power, did not improve the data in any way. We will choose the original no transformation to the data.

The following information provided by Design-Expert shows that the chosen terms to be added to our model are all significant with p-values much less than 0.05 (α=5%).

Response 1 Nash-Sutcliffe, E
ANOVA for selected factorial model
Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F-Value	p-value	
Model	1.817E+005	7	25960.40	1.614E+005		< 0.0001
B-snow runoff coefficient, cs	*30822.19*	*1*	*30822.19*	*1.916E+005*		*< 0.0001*
D-degree-day factor, a	*57148.49*	*1*	*57148.49*	*3.552E+005*		*< 0.0001*
AB	*11077.04*	*1*	*11077.04*	*68846.90*		*< 0.0001*
AD	*20524.14*	*1*	*20524.14*	*1.276E+005*		*< 0.0001*
BD	*29898.73*	*1*	*29898.73*	*1.858E+005*		*< 0.0001*
ABD	*10751.10*	*1*	*10751.10*	*66821.10*		*< 0.0001*
Residual	1.29	8	0.16			
Cor Total	1.817E+005	15				

The adjusted R^2 value is 1.000, which means that 100% of the model's total variability is represented by the factors in the model. The "adjusted" part means that it is adjusted for the number of terms in the model.

Here is the effects list, which provides information on the model variables and their % contribution in predicting the response variable:

	Term	Effect	SumSqr	% Contribtn
Require	Intercept			
Model	**A, k**	**73.3163**	**21501.1**	**11.8317**
Model	**B-cs**	**-87.7813**	**30822.2**	**16.961**
Error	C-cr	-0.32875	0.432306	0.000237892
Model	**D-a**	**-119.529**	**57148.5**	**31.4479**
Model	**AB**	**52.6238**	**11077**	**6.09553**
Error	AC	0.20125	0.162006	8.91496E-005
Model	**AD**	**71.6312**	**20524.1**	**11.2941**
Error	BC	-0.15125	0.0915063	5.03545E-005
Model	**BD**	**-86.4563**	**29898.7**	**16.4528**
Error	CD	-0.29875	0.357006	0.000196455
Error	ABC	0.08875	0.0315062	1.73374E-005
Model	**ABD**	**51.8437**	**10751.1**	**5.91617**
Error	ACD	0.17625	0.124256	6.83763E-005
Error	BCD	-0.12625	0.0637563	3.50841E-005
Error	ABCD	0.07875	0.0248063	1.36505E-005
	Lenth's ME	0.631399		
	Lenth's SME	1.28183		

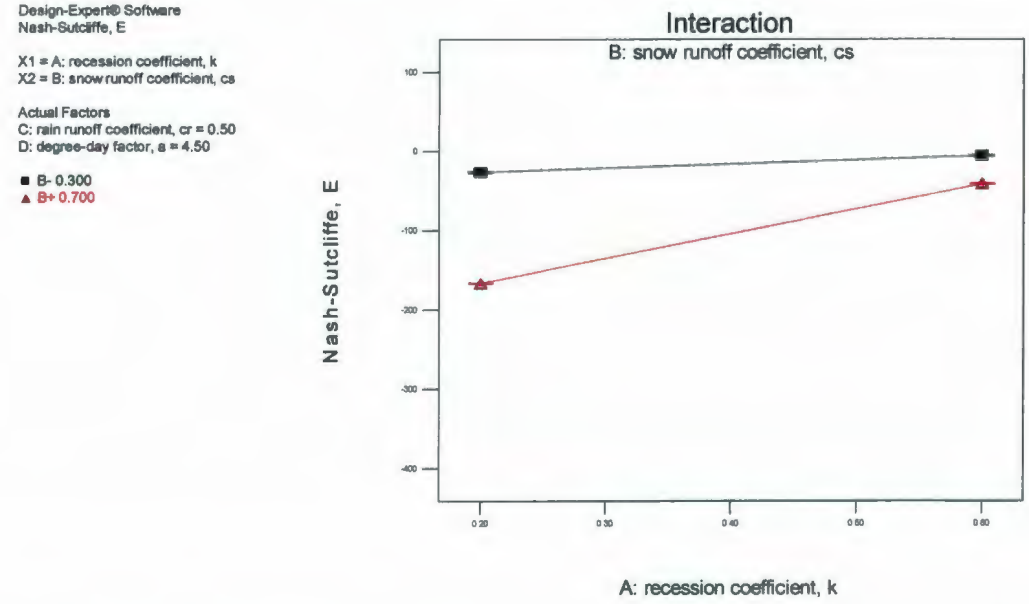
The final equation in coded terms is:

Nash-Sutcliffe = -60.28 + 36.66 A – 43.89 B – 59.76 D + 26.31 AB + 35.82 AD – 43.23 BD + 25.92 ABD

From this equation the user can conclude that the recession coefficient (factor A), snow runoff coefficient (factor B), and degree-day factor (factor D), all play significant roles in the model used to predict the Nash-Sutcliffe model efficiency coefficient (used to assess

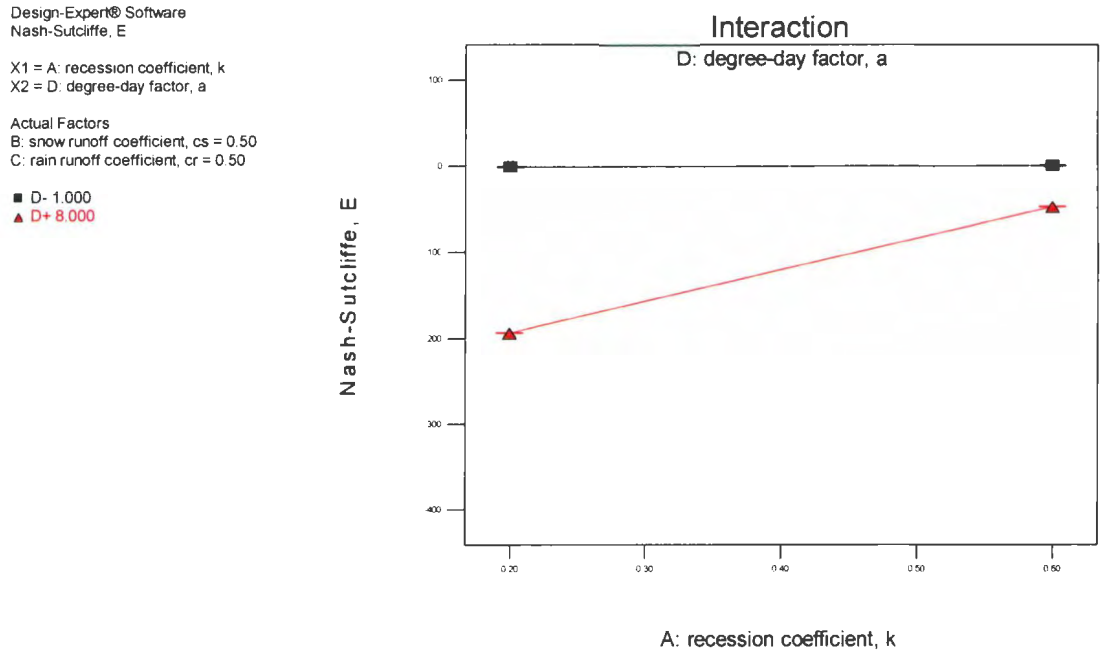
the predictive power of hydrological models). The rain runoff coefficient (factor C) is not significant for measuring this response. It should, however, be noted that there are three significant two-factor interactions between factors A and B, factors A and D, and factors B and D. These interactions can be visualized in the plots to follow.

Interaction plot between factors A and B: recession coefficient and snow runoff coefficient:



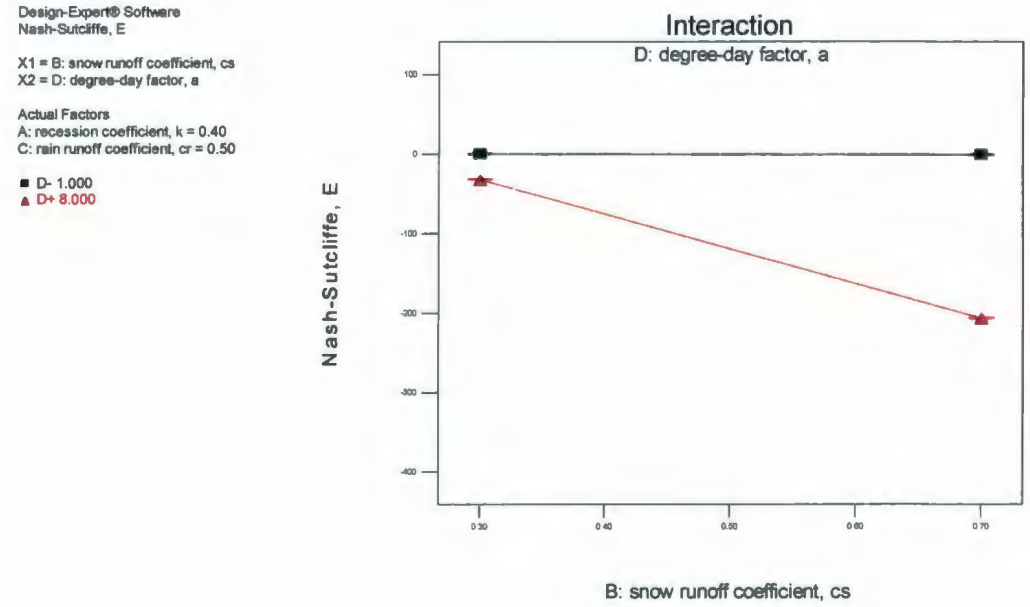
As the recession coefficient increases from 0.20 to 0.60 the Nash-Sutcliffe coefficient increases to its desired value of 1.0. A snow runoff coefficient of 0.30 as opposed to 0.70 seems more desirable for a higher predicitive power for the snowmelt runoff model proposed by Martinec.

Interaction plot between factors A and D: recession coefficient and degree-day factor:



As the recession coefficient increases from 0.20 to 0.60 and the degree-day factor is high (8.0) the Nash-Sutcliffe coefficient increases at a steep rate to it's desired value of 1.0. It is quite obvious that a degree-day factor of 1.0 is much better for predicting the flow compared to a = 8.0. When 'a' is 1.0 the variation of the recession coefficient from 0.20 to 0.60 has very little effect and the Nash-Sutcliffe coefficient stays very close to 1.0.

Interaction plot between factors B and D: snow runoff coefficient and degree-day factor:



As the snow runff coefficient increases from 0.3 to 0.7 and the degree-day factor is set at 8.0 the Nash-Sutcliffe response variable decreases dramatically, which is undesirable. When 'a' is set to 1.0 there is much less variability when the snow runoff coefficient changes from 0.3 to 0.7. Again it is obvious that a high degree-day factor of 8.0 is undesirable.

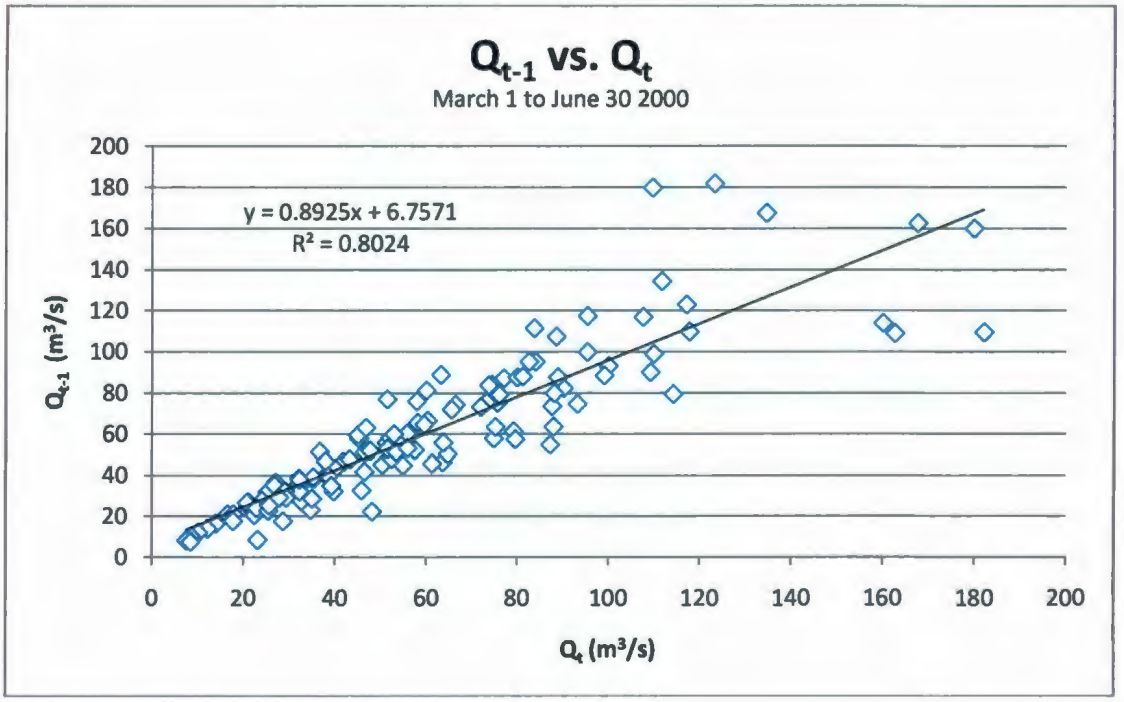
It is assumed (from the sparcity of effects principle) that the effects of three-factor interactions are negligible (close to zero).

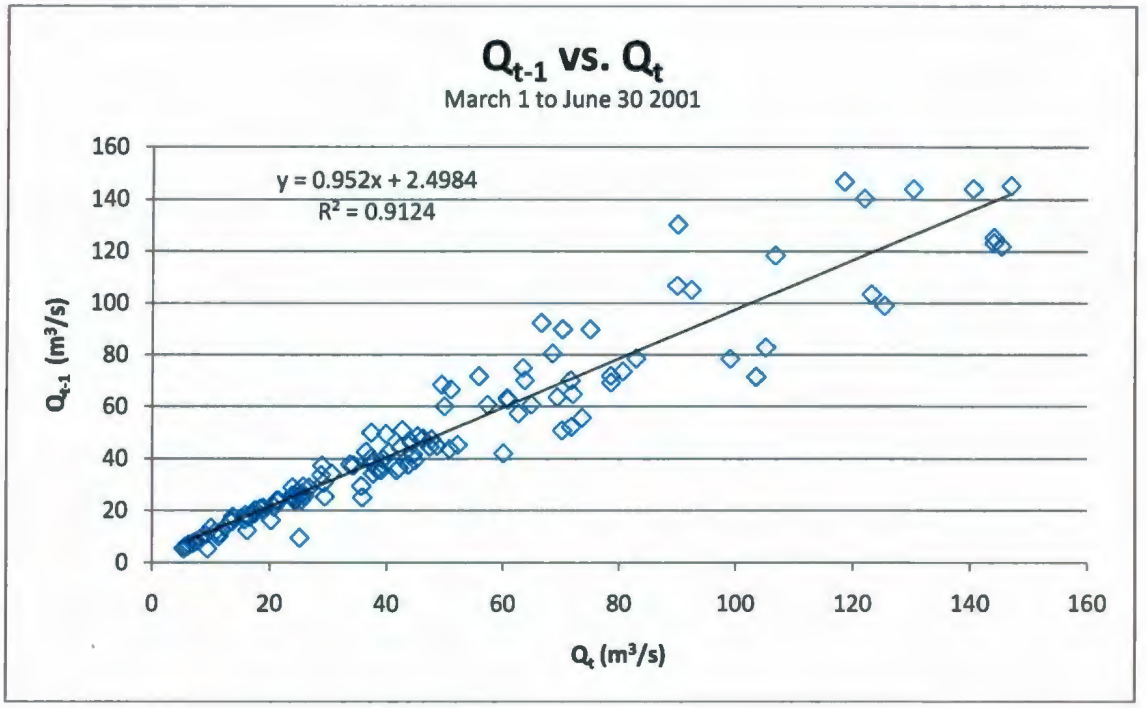
Recommendations: Try another DOE experiment with refined ranges for factors to assess improvement. The rain runoff coefficient physically seems like an important factor and it will be kept for the next trial. It may have been overshadowed by the large effects that the degree-day factor had on the response. From the DOE analysis it can be concluded that for predicting flow in the snowmelt season of 2002 from March 1st to June 30th for the watershed of the Upper Humber River above Black Brook:

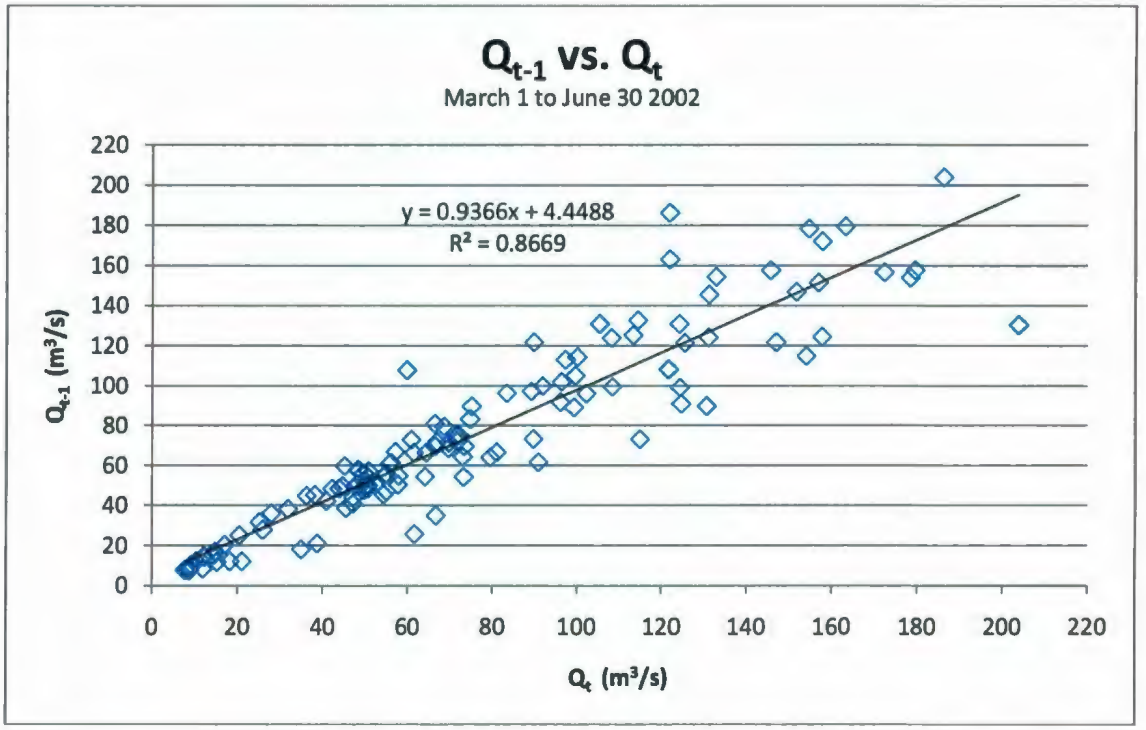
- Factor A, recession coefficient: 0.6 was better than 0.2
- Factor B, snow runoff coefficient: 0.3 was better than 0.7
- Factor C, rain runoff coefficient: no conclusion, deemed non-significant in this model
- Factor D, degree-day factor: 1.0 was much better than 8.0

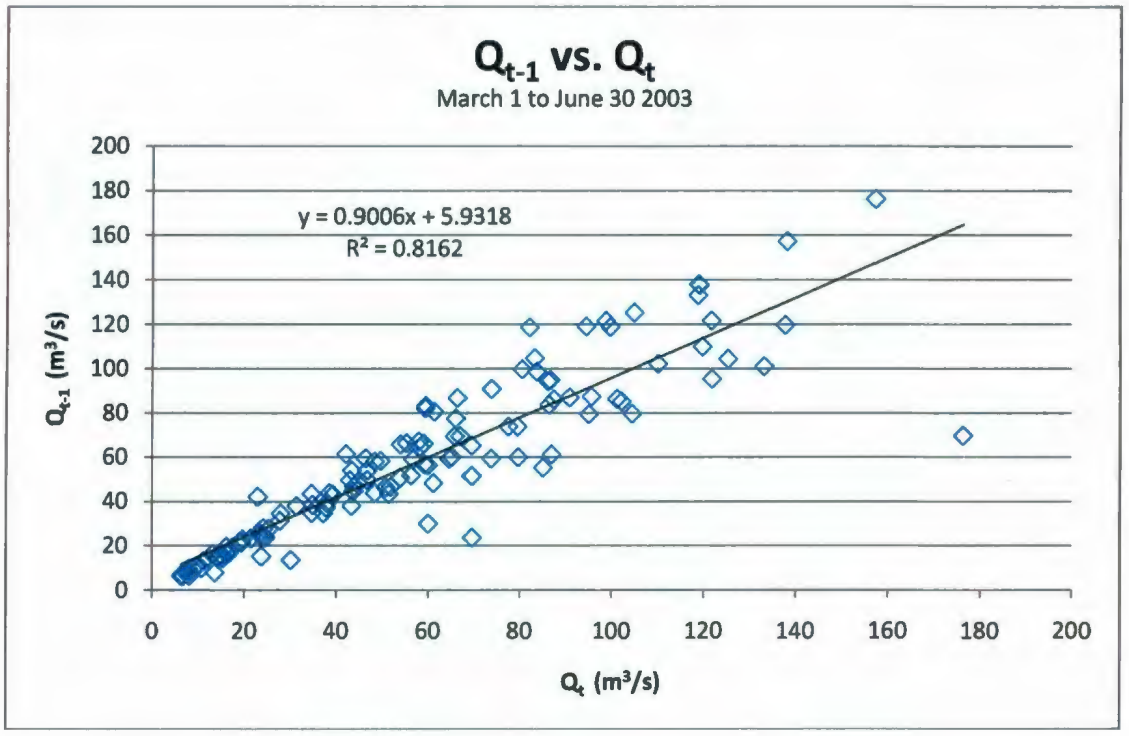
Appendix G

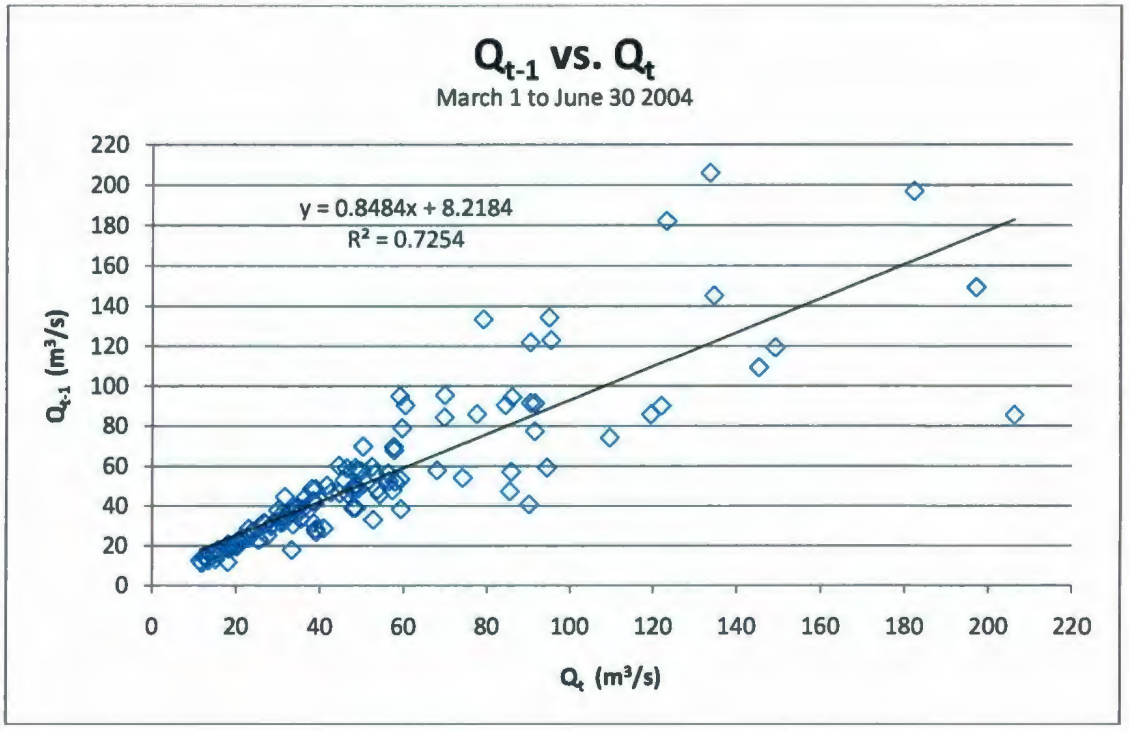
Determination of Recession Coefficient k-value from Q_{t-1} vs. Q_t plots for Calibration Years 2000 to 2007 during Snowmelt Season March 1 to June 30

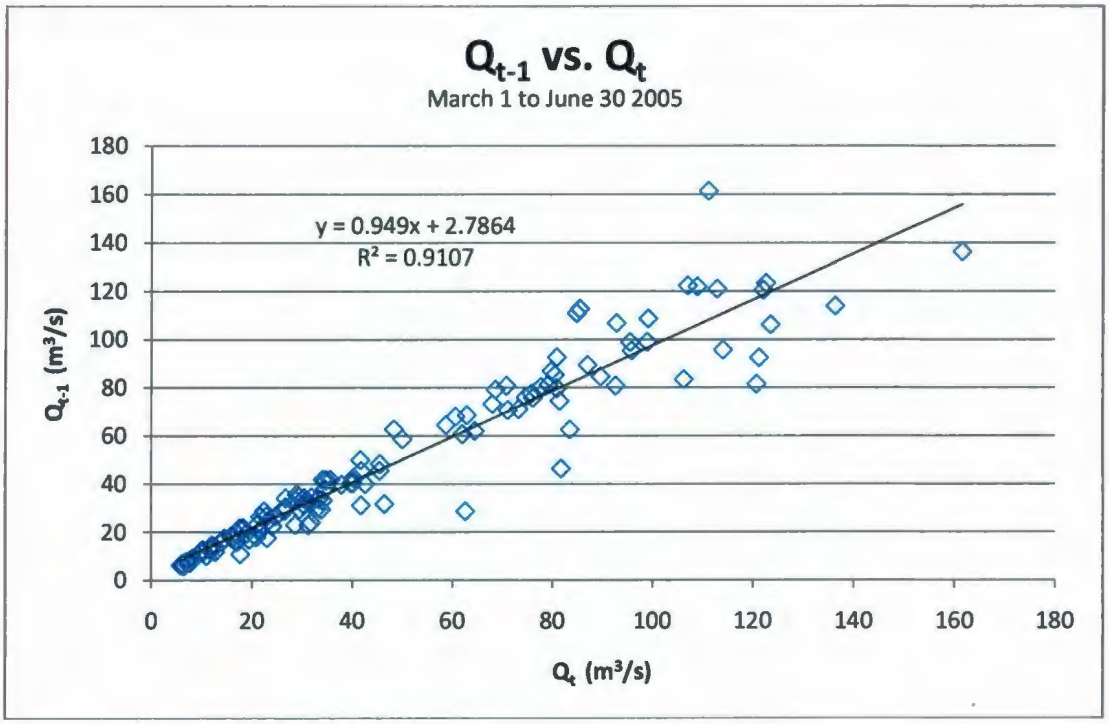


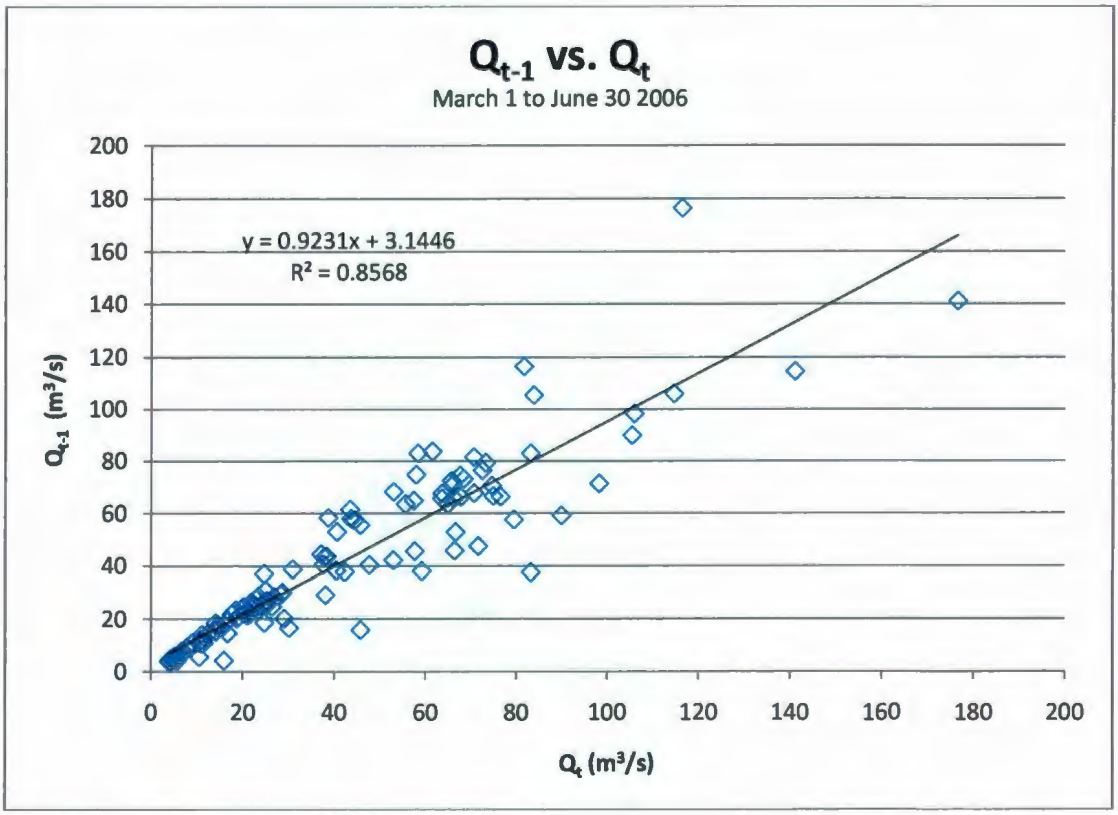

Q_{t-1} vs. Q_t
March 1 to June 30 2001
y = 0.952x + 2.4984
R^2 = 0.9124
Q_{t-1} (m^3/s)
Q_t (m^3/s)
0
20
40
60
80
100
120
140
160

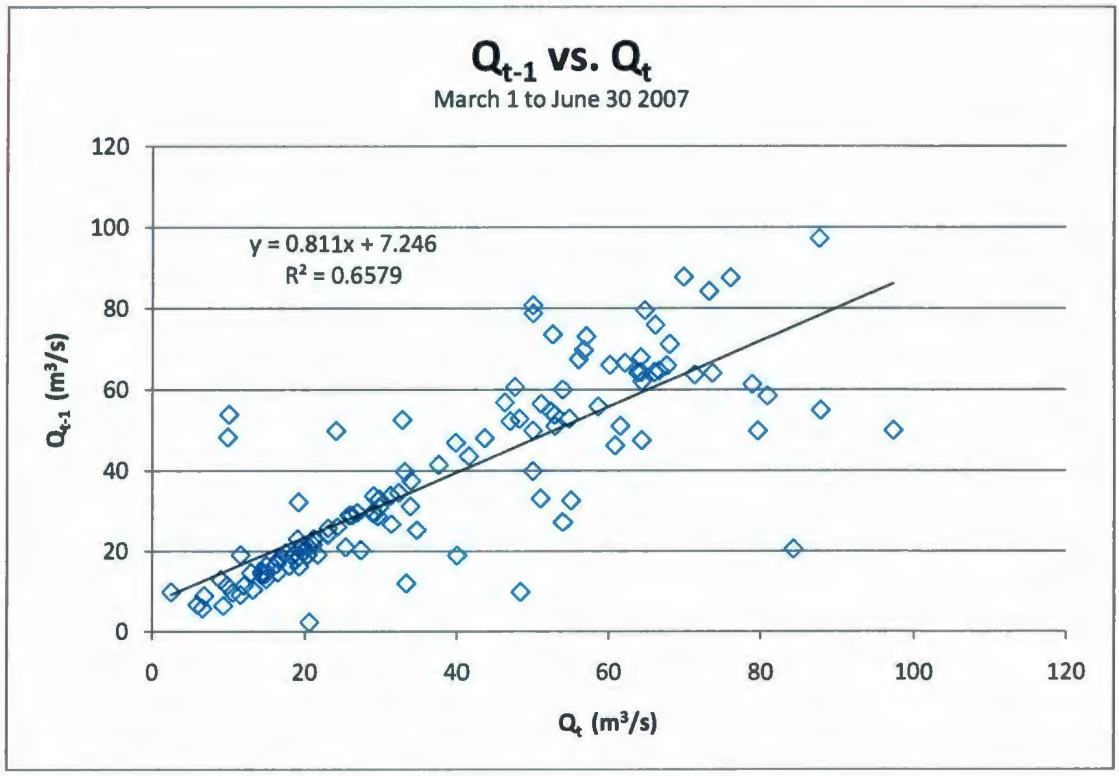

Q_{t-1} vs. Q_t
March 1 to June 30 2002
y = 0.9366x + 4.4488
R^2 = 0.8669
Q_{t-1} (m^3/s)
Q_t (m^3/s)
0
20
40
60
80
100
120
140
160
180
200
220


Q_{t-1} vs. Q_t
March 1 to June 30 2003
y = 0.9006x + 5.9318
R^2 = 0.8162
Q_{t-1} (m^3/s)
Q_t (m^3/s)
0
20
40
60
80
100
120
140
160
180
200


Q_{t-1} vs. Q_t
March 1 to June 30 2004
y = 0.8484x + 8.2184
R^2 = 0.7254
Q_{t-1} (m^3/s)
Q_t (m^3/s)
0
20
40
60
80
100
120
140
160
180
200
220


Q_{t-1} vs. Q_t
March 1 to June 30 2005
y = 0.949x + 2.7864
R² = 0.9107
Q_{t-1} (m³/s)
Q_t (m³/s)
0
20
40
60
80
100
120
140
160
180

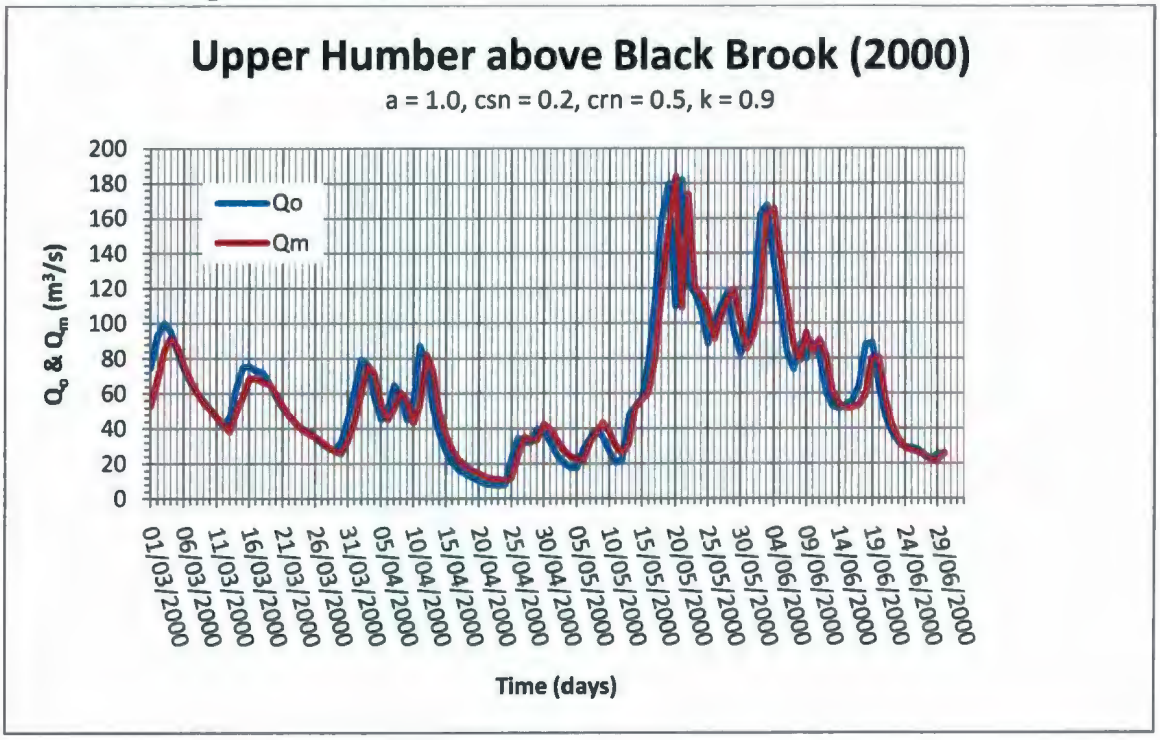

Q_{t-1} vs. Q_t
March 1 to June 30 2006
y = 0.9231x + 3.1446
R² = 0.8568
Q_{t-1} (m³/s)
Q_t (m³/s)
0
20
40
60
80
100
120
140
160
180
200

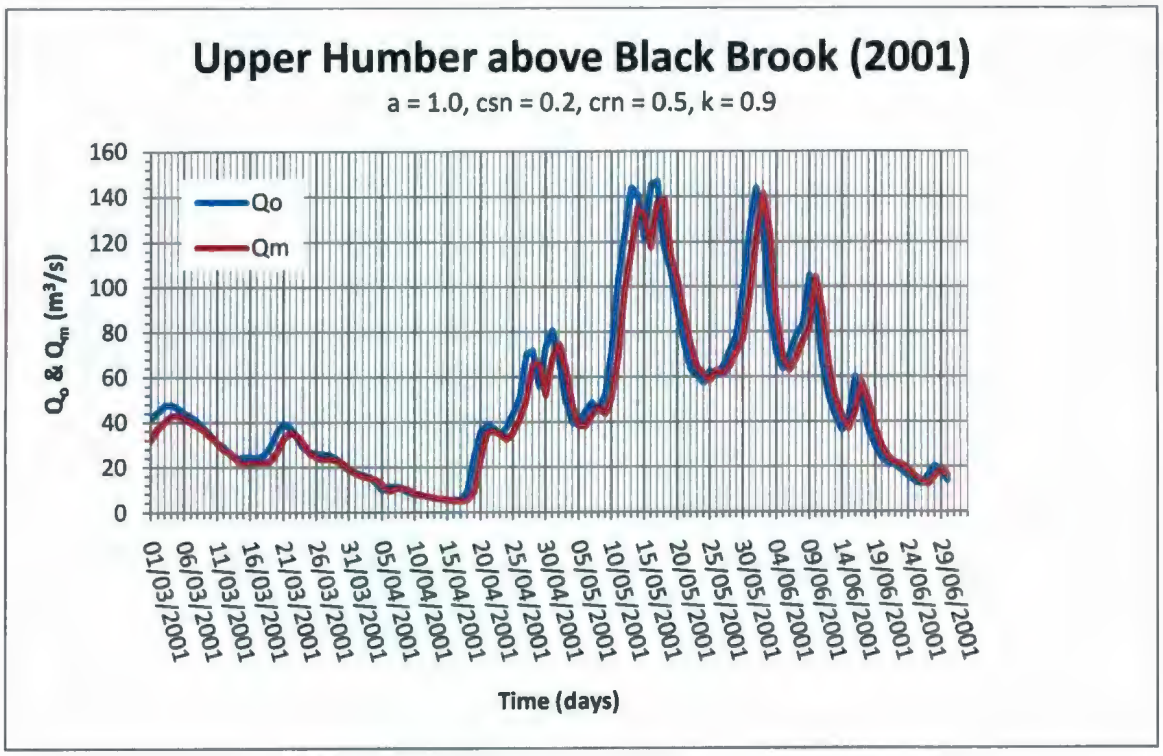

Q_{t-1} vs. Q_t
March 1 to June 30 2007
y = 0.811x + 7.246
R^2 = 0.6579
Q_{t-1} (m^3/s)
Q_t (m^3/s)
0
20
40
60
80
100
120

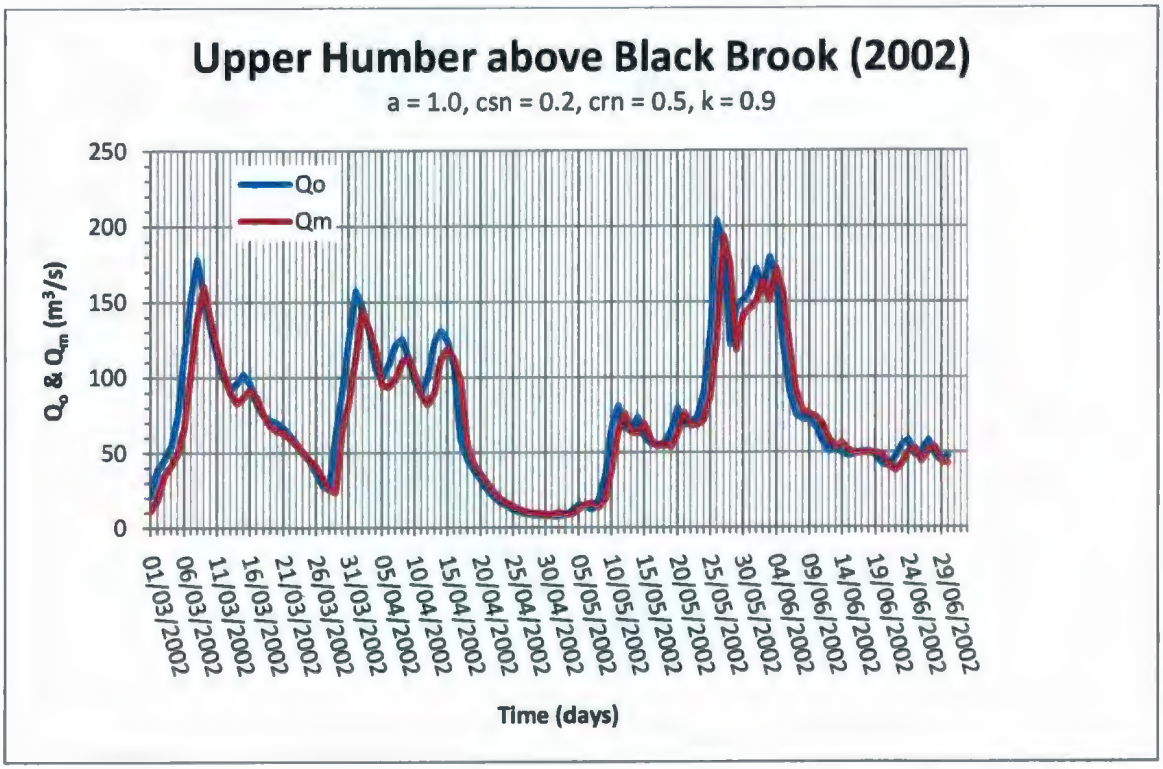
Appendix H

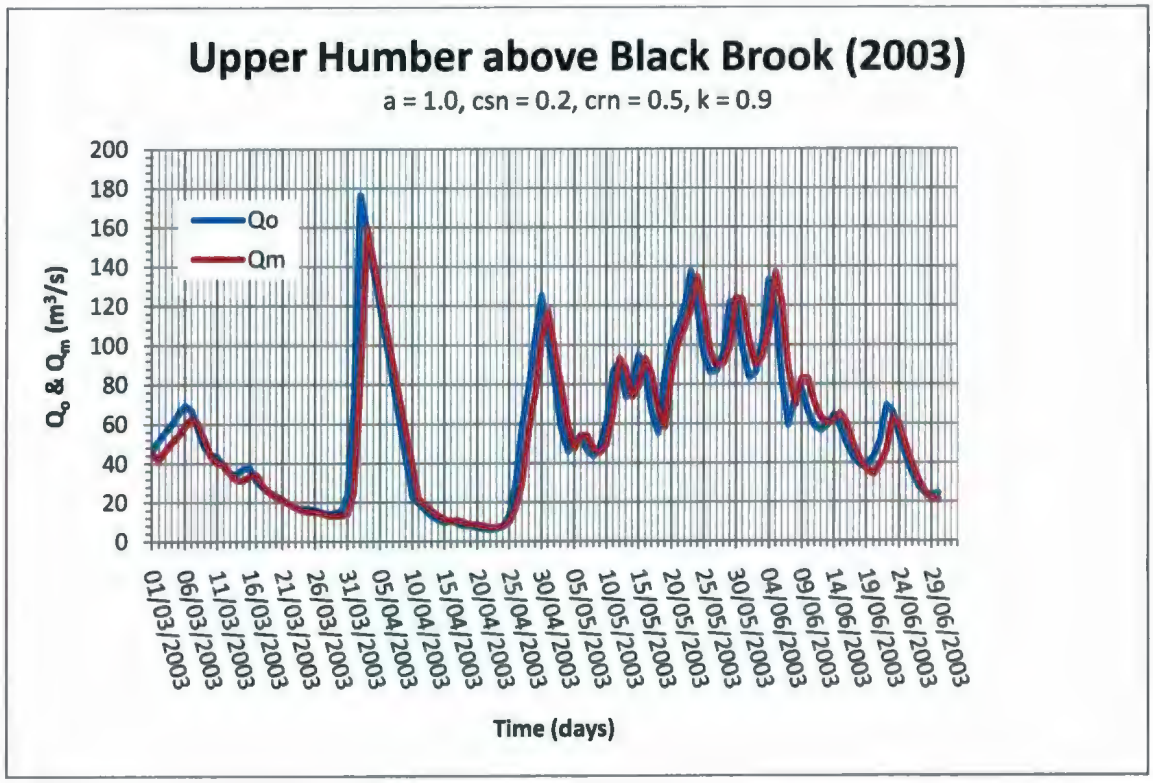
All $Q_{modeled}$ and $Q_{observed}$ vs. Time Plots from 2000 to 2009

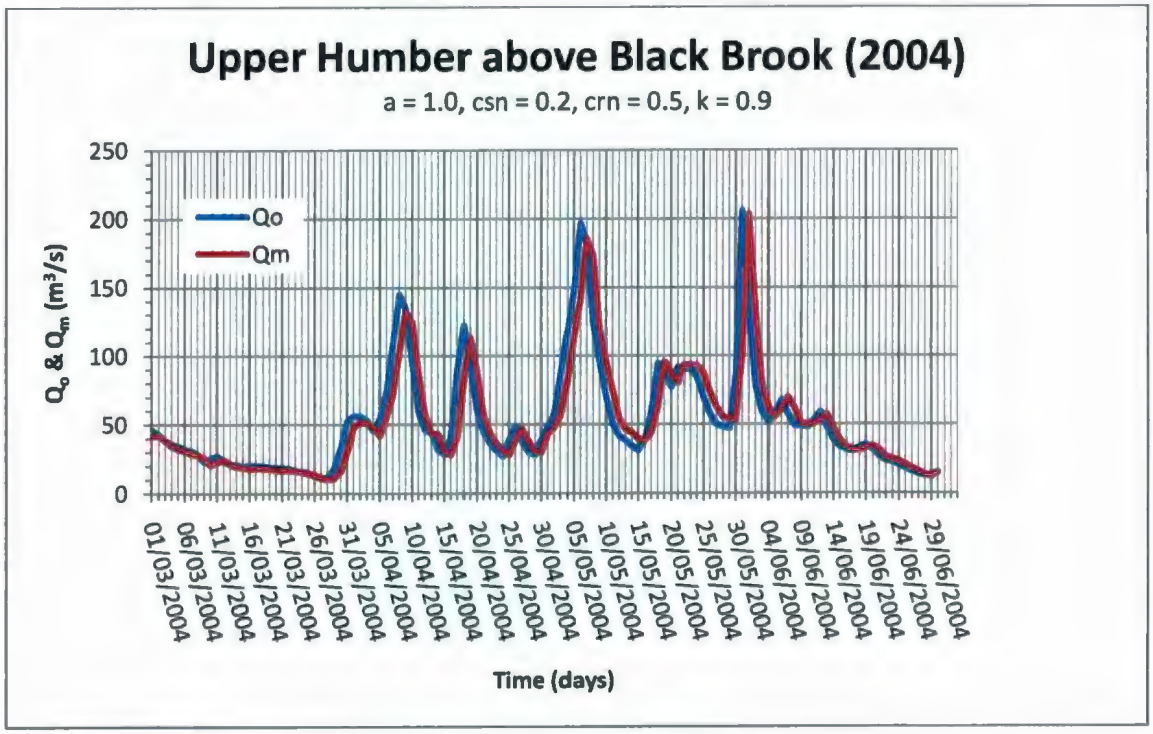
Calibration Stage 2000 to 2007:

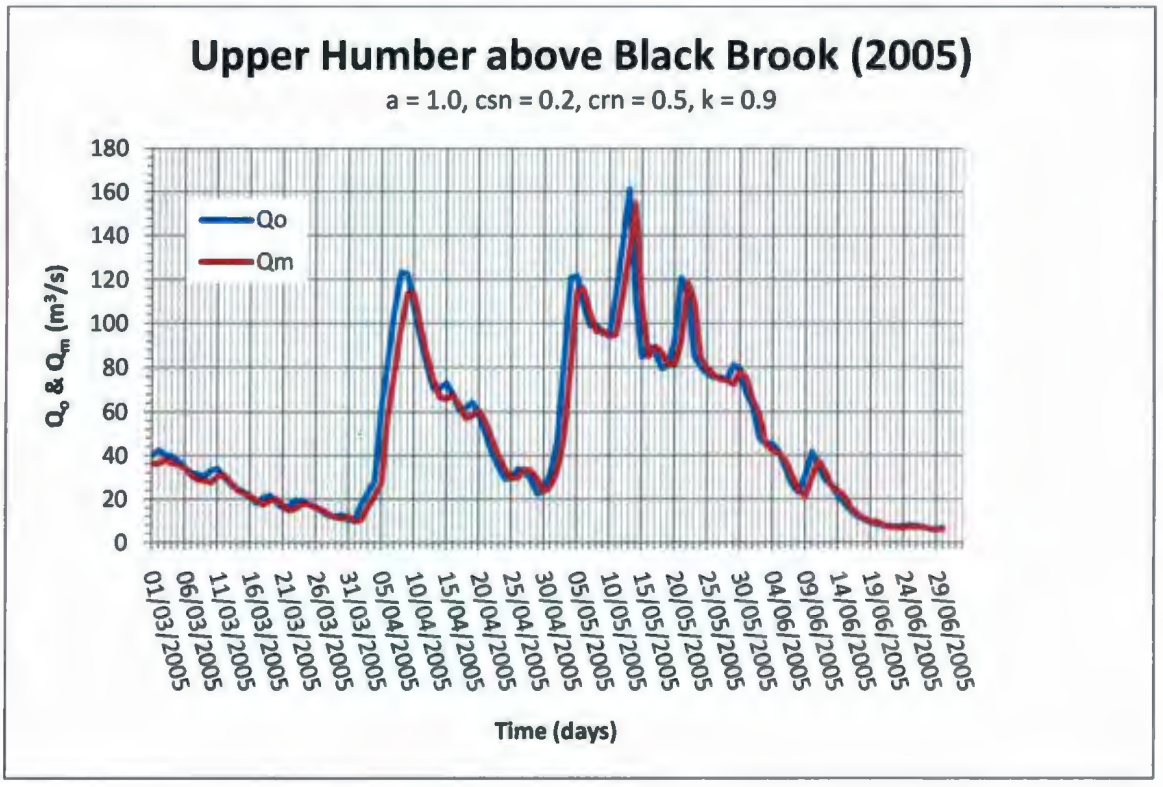


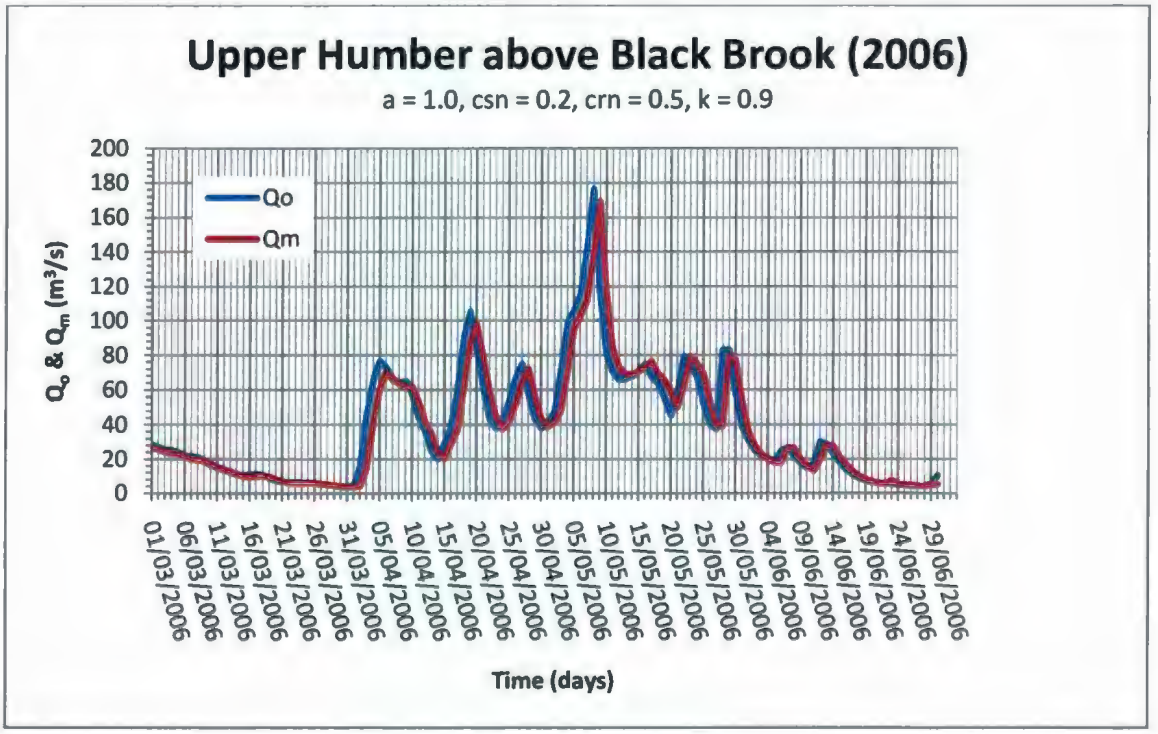

Upper Humber above Black Brook (2001)
a = 1.0, csn = 0.2, crn = 0.5, k = 0.9
Qo
Qm
Q_o & Q_m (m^3/s)
0
20
40
60
80
100
120
140
160
01/03/2001
06/03/2001
11/03/2001
16/03/2001
21/03/2001
26/03/2001
31/03/2001
05/04/2001
10/04/2001
15/04/2001
20/04/2001
25/04/2001
30/04/2001
05/05/2001
10/05/2001
15/05/2001
20/05/2001
25/05/2001
30/05/2001
04/06/2001
09/06/2001
14/06/2001
19/06/2001
24/06/2001
29/06/2001
Time (days)

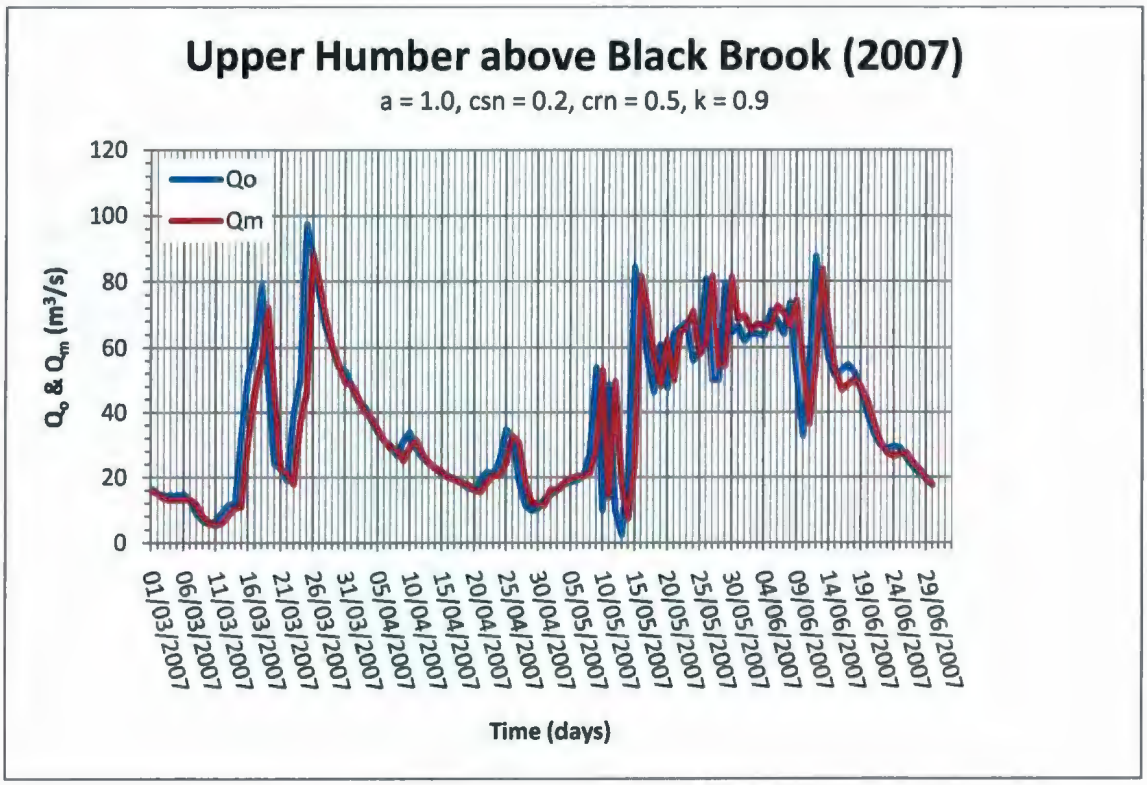

Upper Humber above Black Brook (2002)
a = 1.0, csn = 0.2, crn = 0.5, k = 0.9
Qo
Qm
Q_o & Q_m (m^3/s)
0
50
100
150
200
250
01/03/2002
06/03/2002
11/03/2002
16/03/2002
21/03/2002
26/03/2002
31/03/2002
05/04/2002
10/04/2002
15/04/2002
20/04/2002
25/04/2002
30/04/2002
05/05/2002
10/05/2002
15/05/2002
20/05/2002
25/05/2002
30/05/2002
04/06/2002
09/06/2002
14/06/2002
19/06/2002
24/06/2002
29/06/2002
Time (days)


Upper Humber above Black Brook (2003)
a = 1.0, csn = 0.2, crn = 0.5, k = 0.9
Qo
Qm
Q_o & Q_m (m^3/s)
Time (days)


Upper Humber above Black Brook (2004)
a = 1.0, csn = 0.2, crn = 0.5, k = 0.9
Qo
Qm
Q_o & Q_m (m^3/s)
Time (days)


Upper Humber above Black Brook (2005)
a = 1.0, csn = 0.2, crn = 0.5, k = 0.9
Qo
Qm
Qo & Qm (m³/s)
Time (days)


Upper Humber above Black Brook (2006)
a = 1.0, csn = 0.2, crn = 0.5, k = 0.9
Qo
Qm
Qo & Qm (m³/s)
Time (days)


Upper Humber above Black Brook (2007)
a = 1.0, csn = 0.2, crn = 0.5, k = 0.9
Qo
Qm
Q_o & Q_m (m^3/s)
120
100
80
60
40
20
0
01/03/2007
06/03/2007
11/03/2007
16/03/2007
21/03/2007
26/03/2007
31/03/2007
05/04/2007
10/04/2007
15/04/2007
20/04/2007
25/04/2007
30/04/2007
05/05/2007
10/05/2007
15/05/2007
20/05/2007
25/05/2007
30/05/2007
04/06/2007
09/06/2007
14/06/2007
19/06/2007
24/06/2007
29/06/2007
Time (days)

Validation Stage 2008 to 2009: