ABSTRACT

Title of Thesis: MAPPING SNOW COVER IN SIBERIA USING GIS AND REMOTE SENSING

Aditya Saini, Master of Science, 2003

Thesis directed by: Associate Professor Kaye L. Brubaker

Department of Civil and Environmental Engineering

The seasonal snowpack dynamics of the Siberian mountains and plains play a critical role in the freshwater fluxes of northern rivers into the Arctic Ocean. This study is part of a larger project seeking to quantify the freshwater discharge of the Ob River (Russia) to the Arctic, using satellite remote sensing and hydrologic modeling. Subareas of the Tom River watershed (a tributary to the Ob) are delineated using digital elevation and land cover data within a Geographical Information System (GIS). Daily satellite snow images for 1981 to 1985 are analyzed with the GIS to give time series of percent snow cover for each subarea. These time series are compared with model generated snow cover visually and statistically, allowing spatially distributed evaluation of the model's snow sub-routines. This study will contribute to improved models and better understanding of important and climate-sensitive snow processes and freshwater supply in Arctic river basins.

MAPPING SNOW COVER IN SIBERIA USING GIS AND REMOTE SENSING

by

Aditya Saini

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2003

Advisory Committee:

Associate Professor Kaye L. Brubaker, Chair
Associate Professor Glenn E. Moglen
Professor Richard H. McCuen

ACKNOWLEDGEMENTS

Numerous people have contributed directly or indirectly to the completion of this research. First and foremost, I would like to thank my advisor, Dr. Kaye Brubaker, for her ideas, direction and exceptional support throughout the project. Dr. Moglen's assistance in the early stages of the project with my on-and-off GIS queries and Dr. McCuen's help in the final stages with the statistical analysis is also greatly acknowledged. I would also like to thank Jeremy Stoll from NASA GSFC for his timeliness in sharing common data. My room-mates and lab-mates also deserve recognition for the periodic but valuable discussions on various useful topics. I would also like to thank Ms. Jothsna Rao for her patience and unwavering support. Finally, I would like to thank my parents and brother, who are half way across the world in India. This would have never been possible without their love, guidance and support.

The funding for this research was provided by the National Science Foundation (NSF). Their support is greatly acknowledged.

TABLE OF CONTENTS

List of Abbreviations

1. AVHRR: Advanced Very High Resolution Radiometer
2. AVIRIS : Airborne Visible/Infrared Imaging Spectrometer
3. EASE Grid: Equal Area Scalable Earth-Grid
4. ftp or FTP: file transfer protocol
5. GIS: Geographic Information System
6. GISMO: Graphical Interface for Subsetting, Mapping, and Ordering
7. GLCF : Global Land Cover Facility
8. GSFC: Goddard Space Flight Center
9. HRU: Hydrological Response Unit
10. HSB : Hydrologic Science Branch
11. MODIS : Moderate Resolution Imaging Spectroradiometer
12. NASA: National Aeronautics and Space Administration
13. NOAA: National Oceanic and Atmospheric Administration
14. NSIDC: National Snow and Ice Data Center
15. PRMS : Precipitation Runoff Modeling System
16. SCA: Snowcover Area
17. SSM/I: Special Sensor Microwave/Imager
18. SWE : Snow Water Equivalent
19. TM : Thematic Mapper
20. UMIACS : University of Maryland Institute for Advanced Computer Studies
21. USGS: United States Geological Survey

CHAPTER ONE

INTRODUCTION

1.1 PROBLEM STATEMENT

A key component in both the hydrology of northern Eurasia and the global climate system is the Siberian snowpack, which covers a vast region of Eurasia. Snowpack makes up a large fresh water reservoir that is rapidly released when the temperature rises in the spring. Unfortunately, the contribution of snow in northern Siberia and the high mountainous rim of this project's study area, the Ob river basin in Russia, is largely unquantified. This is principally due to the absence of snow course and precipitation gages in such remote areas. Uncertainties in individual water balance components in the Arctic are especially large when estimates are made for periods shorter than a year. The poor accuracy is explained by the scarcity of measuring stations on the Arctic land and by specific difficulties associated with measurements in very cold climates.

Remote sensing appears to be the most feasible alternative for estimating snow data in these remote regions. Advances in remote sensing and GIS technologies provide an opportunity to further our understanding of the effects of snow depth and snow cover on the water balance of the Arctic region. One of the uses of data thus acquired and analyzed is the calibration and validation of spatially distributed models that are used for making hydrological forecasts for such remote areas.

1.2 RESEARCH OBJECTIVES

The goal of this study was to examine the feasibility of using Geographic Information Systems (GIS) and remote sensing to acquire and process spatial data from regions where actual ground data are not easily available. If successful, these efforts will provide useful data for hydrologic modeling and forecasting in Siberia and similar remote regions.

The first research objective was to search and acquire appropriate spatial snowcover data for the target region (the Ob river basin, Russia) and time (1981 – 1985) from remote sensing data sources that record and manage data in a projection and scale that best suits the needs of the project. The utility of a GIS software is demonstrated in processing and analyzing the remotely sensed satellite data and in the production of data elements for the final product of this study: a set of daily snowcover area (SCA) time-series.

The second objective was to attempt the use of generated SCA time-series to calibrate and validate a spatially distributed hydrological model for further use in quantifying hydrological processes and making accurate hydrological forecasts for regions similar to the one analyzed in this research. The data provided by remote sensing provide the opportunity to test the performance of spatially distributed models and improve their representation of hydrologic processes within river basins.

The final objective was to demonstrate the ability of remote sensing and GIS to assist in decision-making about hydrological forecasting in regions as remote as Siberia, where actual ground data are scarcely available. This project will help future researchers

realize the power of GIS and remote sensing to study water balance and other hydrological considerations for remote areas on the globe.

1.3 CONTEXT OF THE STUDY

This thesis is a sub-project of a larger study being conducted by Brubaker et al. (2001) on the fresh water fluxes of the Ob river basin, Russia, under the sponsorship of the NASA Earth Sciences Enterprise. The thesis was carried out by the University of Maryland team of the project headed by Dr. Kaye Brubaker, and was supplemented by periodic sharing of results and views with the counterpart team at National Air and Space Administration's (NASA) Goddard Space Flight Center (GSFC) Hydrological Science Branch (HSB), led by Dr. Alfred Chang and Dr. Michael Jasinski. Both teams share keen interest in physical hydrology and in the use of state-of-the-art technology to make better hydrological forecasts for regions in the world that are largely out of the domain of regular and reliable ground data.

1.4 DESCRIPTION OF THE STUDY AREA

Figures 1-1 and 1-2 show maps of the region this project aims to study. This project focuses on the Tom River basin in northern Eurasia (Figure 1-2). The Tom River is one of the better gaged catchments in the Ob River basin, which is our larger area of focus. On the territory of the basin are situated six subjects of Russian Federation: Kemerovo, Tomsk, and Novosibirsk regions, Hakasia and Altay republics, and Altay's region. The territories of Kemerovo and Tomsk secure more than 90% of the whole area of the basin. The region is home to around three million people (Suhov, 1992).

Figure 1-1. The Ob and Enisei river basins map
(Source: Wellesley College, 2003)

For the Ob river, large areas are frequently flooded as a result of the south-to-north melt progression. These floods have a severe impact on human activities in the area as well as on the peat and permafrost regions, which sequester significant amounts of

carbon from the global system. Smith et al. (2000) claim that the drying of these peat bogs may result in a significant input of CO_2 into the atmosphere. A decrease in flooding may accelerate the drying of the peat bogs

Figure 1-2. Close-up of the Tom River basin region from Figure 1-1.

1.5 RESEARCH IMPLICATIONS

As mentioned earlier, the goal of this study is to demonstrate the utility of remote sensing and GIS as tools for mapping snowcover in Siberia. The SCA time-series generated as a result of the analysis are intended for use in the calibration of the Precipitation Runoff Modeling System (PRMS, Section 3.1), which is used by the NASA counterparts of this project to quantify water-budget components and make hydrological forecasts in the northern Eurasian region. The effects of image resolution (or pixel size) of the remotely sensed satellite products on the accuracy of the SCA time-series generated are analyzed. Theories are suggested for the imperfect matching of satellite

generated SCA time-series with the corresponding SCA prediction by NASA's PRMS model. It is hoped that results from this project will prove beneficial to NASA-HSB and other researchers using spatially distributed hydrological models in regions with characteristics similar to northern Eurasia.

CHAPTER TWO

PREVIOUS RESEARCH

2.1 OVERVIEW

This chapter discusses the context and importance of carrying out the current project. The following sections talk about recent and relevant work in mapping snow covered areas, especially regions of importance in Siberia because such areas are concerned directly with this research. The current state of the art in spatially distributed hydrological models and the popular means of calibrating and validating them are also discussed. The model (PRMS) and other datasets and tools used in this study are discussed in greater detail in Chapter 3. This chapter largely focuses on the need for this research project.

2.2 MAPPING SNOW COVERED AREAS

Snow is an important, though highly variable, Earth surface cover. Its presence affects physical, chemical and biological processes and has important economic and societal impacts. For example, in many areas of the world, a significant proportion of the freshwater available for consumption, irrigation and hydroelectric generation is supplied from snowpack runoff. Because of its importance, accurate monitoring of snowcover extent is an important research goal in the science of Earth systems (Klein et al., 1998).

Traditional snow survey methods are costly and resource intensive. The most efficient means of monitoring snowcover extent is by remote monitoring from satellites. Fortunately, the physical properties of snow make it highly amenable to monitoring via

remote sensing. Snow has the highest albedo of any natural surface and can reflect significant amounts of solar radiation that would otherwise be absorbed by the earth's surface. Some common properties of snow recorded on a daily and even hourly basis are as follows: Snowcovered Area (SCA), surface albedo, Snow Water Equivalent (SWE), snow depth and snow grain size. Albedo is the fraction of incident electromagnetic radiation reflected by a surface. SWE is the weight of snow equivalent to inches of water at a site.

During the past 35 years, much important information on Northern Hemisphere snow extent has been provided by the National Oceanic and Atmospheric Administration (NOAA) weekly snow extent charts derived from visible-band polar orbiting and geo-stationary imagery (CIRES, 2002). This product represents the longest single time-series of any geophysical product obtained from satellite and is available from the National Snow and Ice Data Center (NSIDC) as the *Northern Hemisphere EASE-Grid Weekly Snow Cover and Sea Ice Extent*.

A number of satellite sensors collect snow data every day. The most noteworthy is the Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS was launched in 1998 aboard the first NASA Earth Observing System (EOS) platform. MODIS is designed to give a global view of the earth at least every other day (Klein et al., 1998). Unfortunately, the unprecedented coverage and resolution provided by MODIS were not available during the time period of this study (1980s). Other notable sensors include the Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave Imager (SSM/I), Advanced Very High Resolution Radiometer (AVHRR), Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Landsat Thematic Mapper (TM),

RADARSAT, etc. This project uses snowcover data from the NSIDC data catalog, and that data was recorded by the AVHRR sensor. The “Surface Type Mask” category of products from that catalog were recorded by the SSM/I sensor.

2.3 SPATIALLY DISTRIBUTED MODELS

Watersheds may be modeled by a lumped model using basin average input data and producing total basin streamflow. Such a model may produce reasonable results but because of the distributed nature of hydrological properties such as soil type, slope and land use, the model cannot be expected to accurately represent conditions within the watershed.

Watersheds are complex systems that consist of a number of sub watersheds, such as forested, agricultural, mountainous, water-covered, urban-developed, or wetland sub watersheds. Each of these sub watersheds can be further divided into a number of lower order sub watersheds and so on until the lowest order sub watersheds are obtained (Chen and Wang 1996). This research subdivides the main Tom R. watershed into five sub watersheds. Each of those sub watersheds are further divided into four Hydrological Response Units (HRUs) each based on land cover classification over the watershed (Section 4.3.3).

The spatial variability of snowmelt processes has received increasing attention in recent years (Bloschl et al. 1991, Kirnbauer et al. 1994). The hydrological model used by the NASA team in this study, PRMS, takes in numerous inputs before returning forecasted values of snowcover area, runoff, snow depth, etc. Some of those input parameters include: precipitation, solar radiation, daily maximum temperature,

infiltration, etc. Calibrating such spatially distributed models is usually not easy because of the many parameters involved and because of the distribution in space. A slight change in one parameter may cause an unexpected degree of change in the output. Intensive subjective optimization is required to calibrate such models distributed over larger areas. This is unlike lumped hydrological models, most of which are still calibrated by simply attempting to match modeled to measured outflow at stream gages. Such models may also require subjective optimization for proper calibration, but they are usually relatively easier to manage.

A spatially distributed model may also consider each pixel in the watershed as a spatially distributed unit, making the model more complex and the output potentially more accurate. This study groups together pixels with similar land uses in each sub watershed to obtain the lowest order sub watersheds: HRUs.

2.4 CHANG ALGORITHM

Dr. Alfred Chang, a senior member of the NASA HSB team involved with this project, devised an equation for snow depth that uses brightness temperatures as input. Brightness Temperature, T_b, is defined as the temperature of a blackbody that emits the same intensity as the measured radiation. It is a key concept in remote sensing. The 'Chang Equation' is as follows:

$$SD = 1.59 (18_{hv} – 37_{hv}) / (1 - f)$$

Here SD is the snow depth, f is the fraction of forest cover as viewed in a satellite image, 18_{hv} and 37_{hv} are brightness temperatures in the 18 GHz and 37GHz bands respectively.

This equation also involves a factor (1.59) which is different for different regions (such as tundra, taiga, etc) and different snow classes.

Microwave remote sensing offers great promise for future applications to snow hydrology. With the availability of satellite microwave data (SMMR and SSM/I), Chang et al (1982) have developed an algorithm for estimating snow water equivalent for dry snow and mapped the depth and global extent of snow cover (Chang et al, 1987).

2.5 SCA v/s SWE

The Snow Cover Area (SCA) fraction of a given region at any point in time is the ratio of the area covered by snow, irrespective of its depth, to the total area of the region. Snow Water Equivalent (SWE) is the amount of liquid water that is contained in a volume of solid snow. As mentioned earlier, this research focuses on the SCA time series in southern Siberia.

Snow can readily be identified and mapped with the visible bands of satellite imagery because of its high reflectance in comparison to non-snow areas. Although snow can be detected at longer wavelengths, i.e., in the near infrared region, the contrast between snow and non-snow areas is considerably reduced compared to the visible region of the spectrum. However, the contrast between clouds and snow is greater in the infrared region and serves as a useful discriminator between clouds and snow (Dozier, 1984). Thermal data are perhaps the least useful of the common remote sensing products for measuring snow and its properties; but they can be useful for helping identify snow/no-snow boundaries and discriminating between clouds and snow. This project deals with a

snow/no snow (SCA) dataset which is generated from original observations of brightness temperatures by the SSM/I sensor.

During the snow melt season, a gradual depletion in both SCA and SWE is expected as the spatially variable snowpack is depleted by spatially varying energy supply throughout the watershed. A sudden drop in either SCA or SWE would imply the sudden delivery of all the water mass contained in the snowpack to downstream areas of the watershed. This, however, is rare; the snowpack is observed to disappear gradually over time, thereby causing smooth sigmoidal (S-shaped) traces of SCA and SWE in time (these time series are known as depletion curves).

CHAPTER THREE

DATASETS AND TOOLS

3.1 OVERVIEW

The focus of this chapter is to describe the role of various tools and datasets chosen and used in this research. This chapter will explain in detail why certain datasets, resolutions and softwares were chosen; how they compare with some other options; their special features, limitations, etc

Section 3.2 discusses the chosen hydrological model to deal with spatial snow data. The following sections address the snow data resource utilized, the capabilities and usefulness of ArcView and Arc/Info, the EASE grid projection, elevation and land cover datasets.

3.2 PRECIPITATION RUNOFF MODELING SYSTEM (PRMS)

As mentioned in Section 1.5, the hydrological model had been chosen and was being calibrated by the collaborators at NASA GSFC before the involvement of the University of Maryland in this research. Although the model is not the focus of this thesis, a description is in order to clarify the modeling context of the snow cover mapping work presented here.

The Precipitation Runoff Modeling System (PRMS) is a modular-design modeling system developed to evaluate the impacts of various combinations of precipitation, climate, and land use on surface-water runoff, sediment yields, and general basin hydrology (Leavesley et al. 1983). Basin response to rainfall and snowmelt can be

simulated on various combinations of land use to evaluate water-balance relationships, flood peaks and volumes, sediment yields, etc. The modular design of PRMS provides a flexible modeling capability and an adaptable modeling system for both management and research applications.

PRMS is designed to function either as a lumped or distributed-parameter type model. PRMS components are designed with the concept of sub-dividing a watershed into units on the basis of characteristics such as slope, vegetation, soil type, precipitation, etc. Each sub-division is considered homogenous with respect to its hydrologic response and is therefore called a Hydrologic Response Unit (HRU). PRMS takes numerous inputs, including descriptive data on the physiography, vegetation, soils and hydrologic characteristics of each HRU (infiltration capacity, solar radiation, snow depth, etc). The output includes hydrologic variables such as surface runoff, ground water flow, streamflow hydrographs, snow-covered area, etc.

The collaborators at NASA had been working on calibrating this model before the start of this project by the University of Maryland team. Their goal was to calibrate the model using measured streamflow and snowcover area (SCA) data over the Ob R. basin for the years 1981-1985, and then use the calibrated model to make forecasts for the same region for the years 1986 to 1990. They decided to focus on the Tom R. watershed, which is the better gaged sub-basin in the Ob R. watershed, for more reliable results towards their goal. The Tom R. watershed had already been sub-divided into five sub areas, and each subarea was further sub-divided into four HRUs (except for one which was divided into three HRUs) based on land cover over the area. Many of their parameters such as elevation and areas of the subareas were based on data from paper maps. This study's

digital sub-division of the Tom R. watershed into its subareas and then HRUs will be explained in detail in Section 3.7.

It is because of the need for simultaneous processing of the various units or HRUs of the Tom R. watershed that a spatially distributed model like PRMS suited the needs of the NASA team. The model outputs SCA time series for each of the nineteen NASA team HRUs for the specified time range. These series will be compared with the corresponding SCA time series generated by using satellite data over the same region and for the same time in detail in Chapter 5.

3.3 SNOW DATA (NSIDC)

The National Snow and Ice Data Center (NSIDC) is part of the University of Colorado Cooperative Institute for Research in Environmental Sciences, and is affiliated with the National Oceanic and Atmospheric Administration (NOAA) National Geophysical Data Center. Established by NOAA as a national information and referral center in support of polar and cryospheric research, NSIDC archives and distributes snow and ice data in both digital and paper map formats. They also maintain information about snow cover, avalanches, glaciers, ice sheets, freshwater ice, sea ice, ground ice, permafrost, atmospheric ice, paleoglaciology, and ice cores.

The NSIDC ‘snow and ice data’ catalog has extensive data archives on snow and ice properties such as ablation, albedo, snow cover, snow depth, snow water equivalent (SWE), etc. These archives are generated by a vast range of sources, including remote sensing satellite sensors such as AVHRR (Advanced Very High Resolution Radiometer)

and MODIS (MODerate resolution Imaging Spectroradiometer), surface observations from NOAA, CLPX (Cold Land Processes Experiments), etc.

3.3.1 Surface Type Mask

Of the various types of data available at NSIDC the dataset selected as most beneficial to the project is *Surface Type Mask*, one of the products from the AVHRR Polar Pathfinder Twice-Daily 5 km EASE-Grid Composites. These composites are a collection of products for both poles, consisting of gridded and calibrated satellite channel data and derived parameters. The polar AVHRR data makes this dataset suitable for the Ob region, which lies in the higher latitudes (North of 45^{o} N).

Surface Type Mask is a good way to generate Snowcover Area (SCA) time series for a region. This dataset has information only on the surface properties of the earth. It does not contain information on the actual snow depth and other properties of snow.

Table 3-1. Numeric codes for surface data types

Code	Surface Data Type
10	Open water
20-29	Predominately first-year ice
30-39	Predominately multiyear ice
40	Bare land
50	Snow-covered land
60	Ice sheet

Table 3-1 shows the list of numeric codes for surface data types in the Surface Type Mask images. Ranges for first-year and multiyear ice types indicate the predominant ice type and the total ice concentration in tens of percent (Fowler et al.,

2000). For example, a value of 24 means that first-year ice predominates and that the total ice concentration is between 41 and 50 percent.

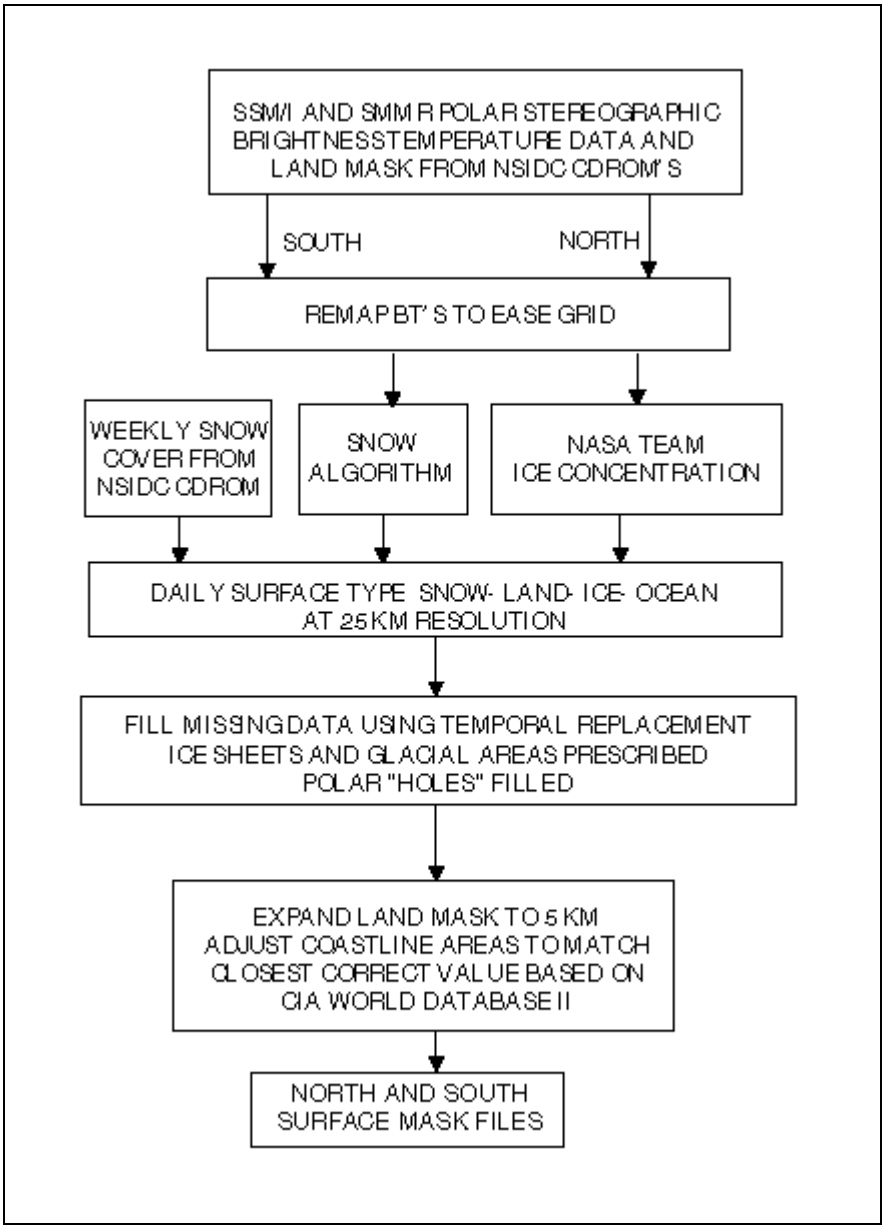


Figure 3-1. Flowchart showing the generation of Surface Type Mask from Passive Microwave data (Source: NSIDC website)

Figure 3-1 is a flowchart that illustrates the steps in generating a Surface Type Mask. Special Sensor Microwave Imager (SSM/I) brightness temperatures were initially divided into land and ocean ('land' also includes ice sheets). Over oceans, a NASA Team Algorithm was used to generate first-year and multiyear ice concentrations (Fowler et al., 2000). All ocean pixels with an estimated ice concentration greater than zero percent were flagged as sea ice. Areas that consist of at least 50 percent multiyear ice were assigned the multiyear flag. The multiyear ice estimate is subject to considerable error and uncertainty, particularly in the snow melt season. Over land, snow cover is detected with Goodison's (1989) SSM/I algorithm.

A sample Surface Type Mask image (as viewed in ArcView) is displayed in Figure 3-2. Procedures to download this data series from the NSIDC website will be explained in Section 3.3.3 and the steps required to convert that data into images viewable in ArcView are detailed in Section 4.3.

Figure 3-2. Sample surface type mask in EASE grid (polar azimuthal projection). The image shows northern Europe. The **North Pole** is Top-Left corner of image. (resolution is 25km x 25 km, text notations were added in ArcView)

3.3.2 GISMO

The GISMO (Graphical Interface for Subsetting, Mapping, and Ordering) is the NSIDC search engine for their gridded datasets by collection, parameter (channel) and date. It proved to be a very handy tool to order specific datasets from their archives. It allows the user to select a geographical area (the Ob basin in this case) from an interactive digital map and specify dates and the product (Surface Type Mask for this study). GISMO was used several times to obtain datasets at different locations in the northern hemisphere. This helped in accurately selecting and georeferencing NSIDC data with other datasets used in the project from other sources.

Data obtained from GISMO was usually in the form *"ftp-site-compressed"*, free of charge and usually available 2-4 days after placing an order. Each order of data would also include a common header file (*.hdr*) for the entire series of images over an area. These images were easily made ArcView-compatible by making a few adjustments, finer details of which are explained in Section 4.4.1.

3.3.3 AVHRR and SSM/I data acquisition and characteristics

The AVHRR data set at NSIDC consists of twice-daily composites (approximately 0400 and 1400 hrs for the Northern Hemisphere). Products are provided in the EASE grid (Equal Area Scalable Earth grid, explained in Section 3.4). This data archive is part of the NASA and NOAA *Pathfinder Program* and is designed to provide time series data on a global scale from a gamut of remote sensing products. The AVHRR Polar Pathfinder Twice-Daily 5 km EASE-Grid Composites are designed to facilitate regional studies and climate studies, and were therefore ideal for this project. 5km x 5km

resolution AVHRR data was selected for the study. Other resolution options were 25km x 25 km and 1.25km x 1.25km. The 5km resolution was chosen because it could be easily resampled to 1km resolution to match other datasets (elevation, landcover) used in the project

Spatial coverage of this dataset extends from 48.4^{o}N to 90^{o}N latitude, and from 53.2^{o}S to 90^{o}S latitude. Table 3-2 gives the values of corner pixels in the Northern Hemisphere. The Northern Hemisphere grid is 1805 pixels wide by 1805 pixels high, centered on the North Pole. Temporal coverage is from 24 July 1981 to 31 December 2000, (Appendix D). Surface Type Mask data were obtained for the Ob region from 24 July 1981 to 31 December 1985. These dates were chosen because the NASA team used PRMS model to get their SCA time series within the same date range (1980-1985), and the goal of this project was to perform a comparative analysis between the two time series. It must be mentioned here that data was not available for 110 days (6.82%) out of 1611 days in that date range.

Table 3-2. Corner pixel values in spatial coverage of **northern** hemisphere

Corner	**Center of corner pixel**	**Outer edge of corner pixel**
Upper left	29.74956 N, 135.00000 W	29.71269 N, 135.00000 W
Upper right	29.74956 N, 135.00000 E	29.71269 N, 135.00000 E
Lower left	29.74956 N, 45.00000 W	29.71269 N, 45.00000 W
Lower right	29.74956 N, 45.00000 E	29.71269 N, 45.00000 E

Source: EASE-grid information at NSIDC (2003)

Lastly in this section the spectral range of the AVHRR sensor is tabulated in Table 3-3. The spectral signatures of snow and water are displayed in Figure 3-3 to give

an idea of what channels in the AVHRR sensor capture higher proportions of reflectance from these features. The data for these curves were obtained from the online Spectral Library maintained by NASA (.DATE – add to REFERENCES) Figure 3-3 indicates that there is a significant difference in reflectance of fine snow and granular snow in channel 3 of the AVHRR sensor, and some perceivable difference in channels 4 and 5. This information will be useful to draw conclusions in the results of this research (Chapter 5).

Table 3-3. Wavelengths for channels of the AVHRR sensor.

Channel	**Wavelength (micrometer)**	
Channel 1	0.58 to 0.68	visible
Channel 2	0.725 to 1.05	reflected infrared
Channel 3	3.55 to 3.92	reflected/thermal infrared
Channel 4	10.3 to 11.3	thermal infrared
Channel 5	11.5 to 12.5	thermal infrared

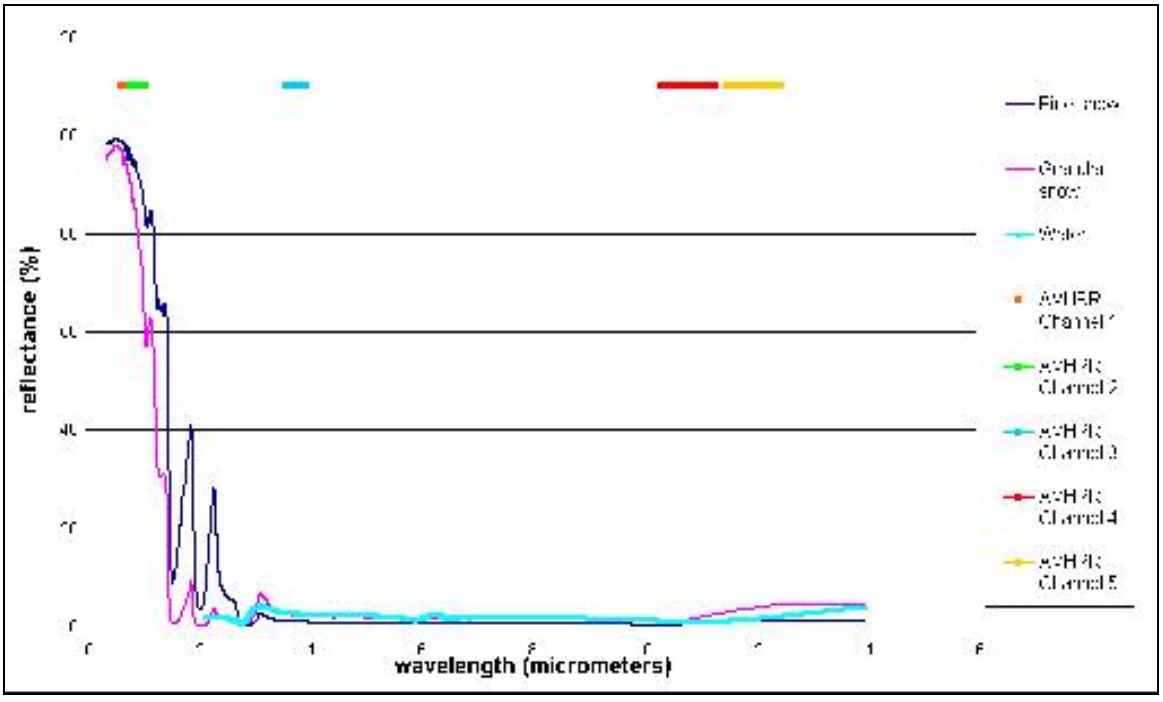
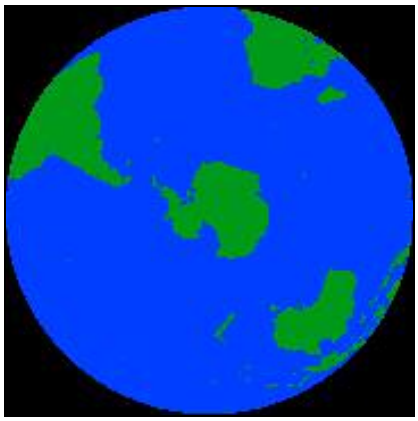


Figure 3-3. Spectral reflectance of snow and water (NASA, 2003).

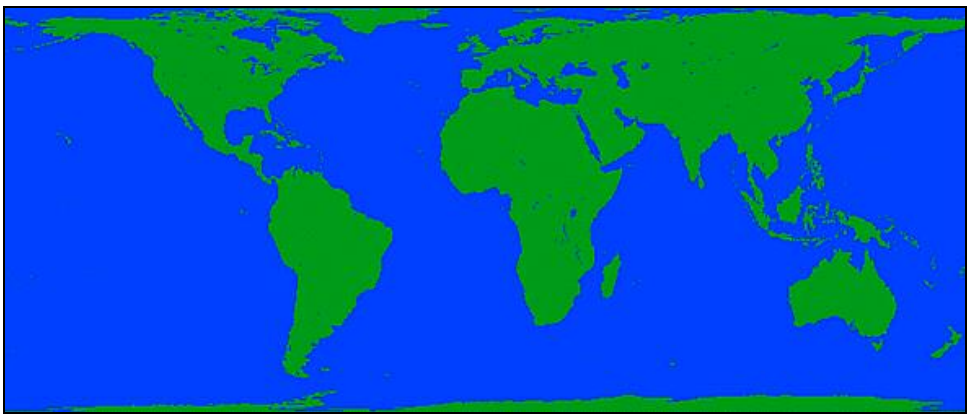
3.4 EASE GRID

The Equal-Area Scalable Earth Grid (EASE-Grid) consists of a set of three equal-area projections. These three EASE-Grid projections comprise two azimuthal equal-area projections, for the Northern or Southern hemisphere, respectively, and a global cylindrical equal-area projection. They are displayed in Figures 3-4a, 3-4b and 3-4c. These projections are based on a spherical model of the Earth with radius R = 6371.228 km, and are based on a philosophy of digital mapping and gridding definitions that was developed at the National Snow and Ice Data Center in Boulder, CO (Brodzik, 2003).

(a) Equal-area azimuthal map (North)



(b) Equal-area azimuthal map (South)



(c) The cylindrical equal-area map

Figure 3-4. The three EASE-grid projections. (Source: NSIDC 2003)

Each projection has different properties and thus is suited for different uses. Two of the most important characteristics of maps are whether they are conformal or equal-area. No map projection is both and some are neither (Knowles, 1993). On equal-area maps, a small circle placed anywhere on the map will always *cover the same amount of area* on the globe. In contrast, on conformal maps, angles within a small area are reproduced accurately, so a small circle on the globe will *look* like a small circle on the map. One popular map is the cylindrical equidistant map or the "latitude-longitude" grid (Mercator projection). This map is neither equal-area or conformal and, therefore, suffers from both areal and shape distortions.

EASE-grid is intended to be a versatile tool for users of global-scale gridded data. especially remotely sensed data. The present project is largely a comparative analysis of snow cover area (SCA) time series from two different sources (satellite remote sensing and a hydrologic model), in a large region situated between latitudes 45^{o} N to 60^{o} N and occupying an area of approximately 64,000 km^2. An equal-area projection facilitates in the computation of areas for each sub-watersheds by merely counting the number of grid cells they occupy and multiplying by a constant area per pixel. Also, EASE grid minimizes the amount of distortion over the hemispheric scale this study is dealing with.

3.5 GTOPO DATA (DIGITAL ELEVATION MODELS)

Digital Elevation Models (or DEMs) were the starting point of the GIS based spatial data analysis in this project. The DEM dataset allowed the GIS to delineate the Tom river watershed and subsequent subareas and HRUs. The DEM data obtained was

from the USGS (2003). It is part of the “Global 30-Arc-second elevation dataset (GTOPO30)”.

GTOPO30 is a global data set covering the full extent of latitude from 90 degrees south to 90 degrees north, and the full extent of longitude from 180 degrees west to 180 degrees east. The horizontal grid spacing is 30-arc seconds (0.008333333333333 degrees), resulting in a DEM having dimensions of 21,600 rows and 43,200 columns. The horizontal coordinate system is decimal degrees of latitude and longitude referenced to World Geodetic System 1984 (WGS84). The vertical units represent elevation in meters above mean sea level. The elevation values range from -407 to 8,752 meters (USGS, 2003).

GTOPO30 has been divided into 33 smaller pieces, or tiles. The area from 60 degrees south latitude to 90 degrees north latitude and from 180 degrees west longitude to 180 degrees east longitude is covered by 27 tiles, with each tile covering 50 degrees of latitude and 40 degrees of longitude. Antarctica (90 degrees south latitude to 60 degrees south latitude and 180 degrees west longitude to 180 degrees east longitude) is covered by 6 tiles, with each tile covering 30 degrees of latitude and 60 degrees of longitude. The tiles names refer to the longitude and latitude of the upper-left (northwest) corner of the tile. For example, the coordinates of the upper-left corner of tile E020N40 are 20 degrees east longitude and 40 degrees north latitude. Fig 3-5 displays each of the GTOPO30 tiles as shown on the USGS GTOPO30 website. There is no overlap among the tiles so the global data set may be assembled by simply abutting the adjacent tiles.

GTOPO30 was derived from several raster and vector sources of topographic information. Some of these sources include:

a) Digital Terrain Elevation data

b) Digital chart of the world

c) USGS digital elevation models

d) Army Map Service maps

and others. (USGS, 2003)

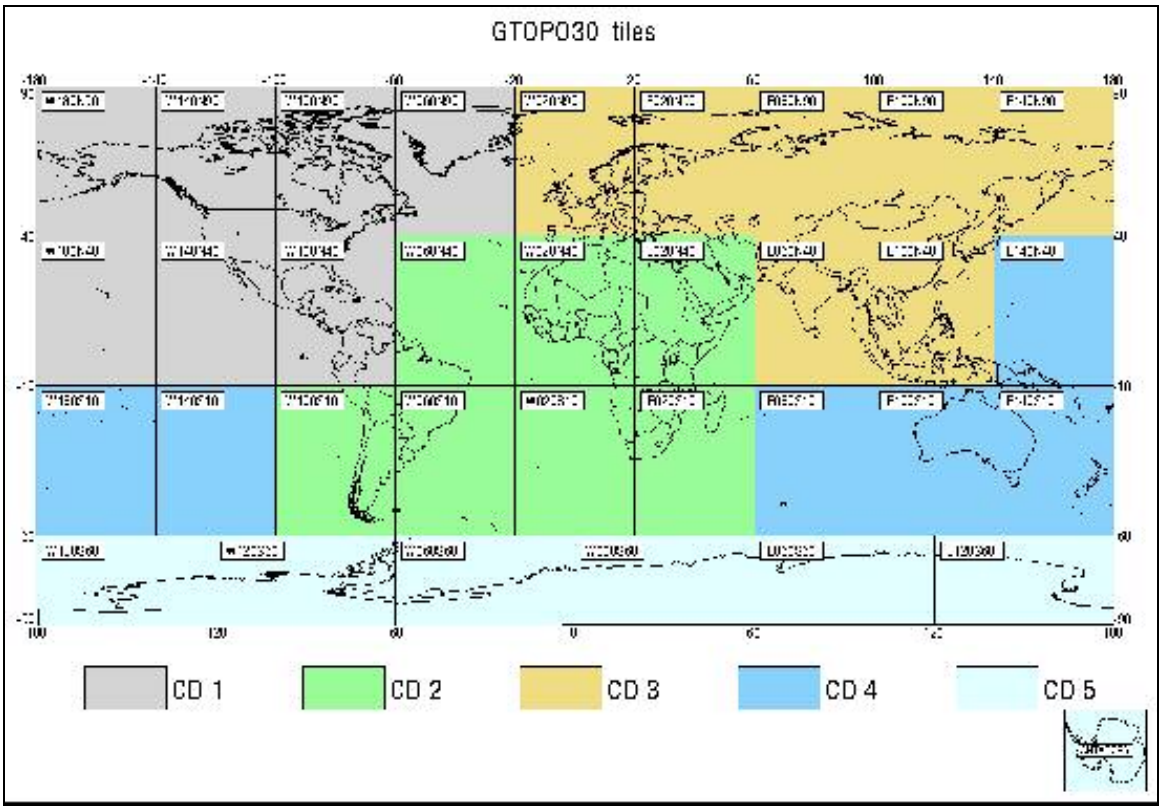


Figure 3-5. GTOPO30 tiles as shown on the USGS website (Source: USGS, 2003)

Data for each tile are provided in a set of 8 files (USGS, 2003). The files are named with the tile name and a file name extension indicating the contents of the file. Table 3-3 shows the extensions used in the 8 files. Further details on actually

downloading data from the GTOPO30 tiles website and converting the data for GIS use are provided in Section 4.2.1 of this thesis.

Table 3-4. Extensions used in files for a single GTOPO30 tile.

Extension	Contents
DEM	digital elevation model data
HDR	header file for DEM
DMW	world file
STX	statistics file
PRJ	projection information file
GIF	shaded relief image
SRC	source map
SCH	header file for source map

This project utilized two of the GTOPO30 tiles, namely: E020N90 and E060N90. The Tom and Ob rivers watersheds lie entirely in these two tiles. GTOPO30 tiles will be discussed in detail again in Sections 4.2.1 and 4.2.2, which also explain how the DEM tiles were handled in the GIS and also how the appropriate region was selected from the big tiles and thereby processed.

3.6 LAND COVER DATA (GLCF)

Another key element required for this research study was a reliable and accurate land cover dataset for the Tom River region. Land cover data was essential because it was to be used to determine the twenty Hydrological Response Units (or HRUs) from among the five subareas. This procedure will be explained in detail in Section 3.7.1.

Land cover data for this research were obtained from the Global Land Cover Facility (GLCF) maintained by the University of Maryland Institute for Advanced Computer Studies (UMIACS). GLCF develops and distributes remotely sensed satellite data and products and claims to be the largest free source of LANDSAT data. They are concerned with land dynamics from the local to global scales. GLCF is a funded member of NASA's Earth Science Information Partnership (Hansen et al., 2000).

The GLCF data catalog has remotely sensed satellite data from AVHRR, LANDSAT, MODIS and other sensors, and also from other land cover data sources. They initially aimed to develop a coarse resolution, global land cover dataset from satellite data for use in climate models. AVHRR data were resampled to a spatial resolution of one degree by one degree and used to carry out a conventional, supervised classification of global land cover(Hansen et al., 2000).

The landcover data acquired from GLCF used in this project comes from multiple sources. Figure 3-6 displays a preview of these data for Europe and Asia. These data are in the Lat-Long or Geographical projection. In addition to this GLCF also maintains landcover data in Goode's Homolosine projection. The data were not recorded by some satellite sensor at some single particular instant. Instead, the final product is the result of data recorded over 14 years (1981-1994), providing ability to test the stability of the classification algorithms (Hansen et al. 2000).

Data in Lat-Long projection for the Eurasian dataset were downloaded from the GLCF website. Thereby, the target area i.e., the Ob R. basin was cropped out. The cropped out landcover dataset lies between 46^{o}N to 68^{o}N and 58^{o}E to 93^{o}E. The data have 14 different landcover categories as displayed in Table 3-5. These 14 categories

were later grouped together into four main landcover types to match those used by the NASA team to delineate HRUs within the subareas.

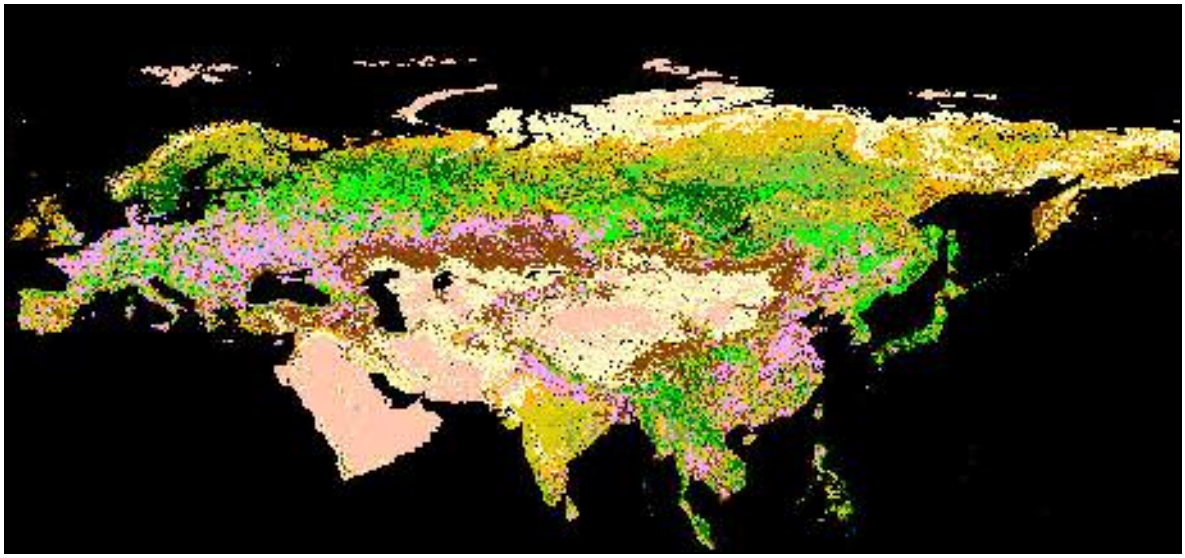


Figure 3-6. GLCF Landcover data for Eurasia (Source: GLCF website)

Table 3-5. Landcover categories for data obtained from GLCF

Class values	
0	Water (and Goode's interrupted space)
1	Evergreen Needleleaf Forest
2	Evergreen Broadleaf Forest
3	Deciduous Needleleaf Forest
4	Deciduous Broadleaf Forest
5	Mixed Forest
6	Woodland
7	Wooded Grassland
8	Closed Shrubland
9	Open Shrubland
10	Grassland
11	Cropland
12	Bare Ground
13	Urban and Built-up

The landcover data were supplied free of charge in a compact disc in the *.tif* format. This format is pre-georeferenced and is GIS-compatible and was therefore imported into the GIS relatively easily. The projection was changed to EASE-grid in the GIS.

3.7 GEOGRAPHIC INFORMATION SYSTEM (GIS)

GIS is a system of hardware and software used for storage, retrieval, mapping, and analysis of geographic data. GIS differs from CAD and other graphical computer applications in that all spatial data is stored in a coordinate system and is geographically referenced to a map projection in an earth coordinate system.

GIS proved to be a vital tool for this project because of the way it manages spatially distributed data. This study involves various types of data layers that had to be overlaid to get desired results (e.g.: land cover over watershed to determine HRUs, snow cover over HRUs to obtain the snow/ no snow time series). Environmental Systems Research Institute's (ESRI) ArcView (a popular GIS software) was selected for all these operations because of its efficiency and reliability to perform such tasks. Another advantage is the impressive use of colors to display subsequent map-calculation results.

This project uses different types of spatial data from different sources. Fortunately all of them could be made GIS-compatible by making slight adjustments at some stage of their analysis. All of those data layers must then be converted to the same units and projection and be accurately georeferenced to get proper results. ESRI's Arc/Info (another popular GIS software) facilitates in the conversion, selection, reprojection, etc of gridded data to make them ArcView compatible. ArcView and Arc/Info are discussed in greater detail in the following sections.

3.7.1 ArcView

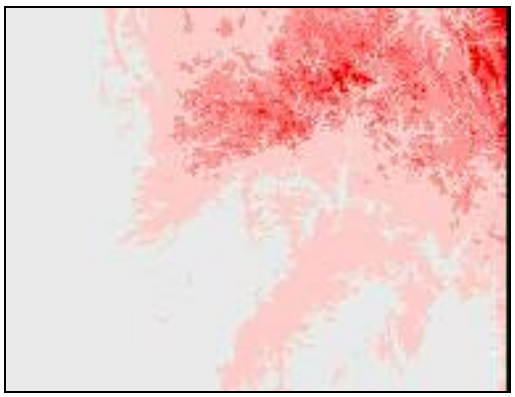
The main GIS software used in this research is ArcView version 3.0b for UNIX. It has been the backbone of this project. This section will explain in detail the three major and essential tasks performed efficiently by ArcView:

a) Delineation of Tom watershed and five subareas

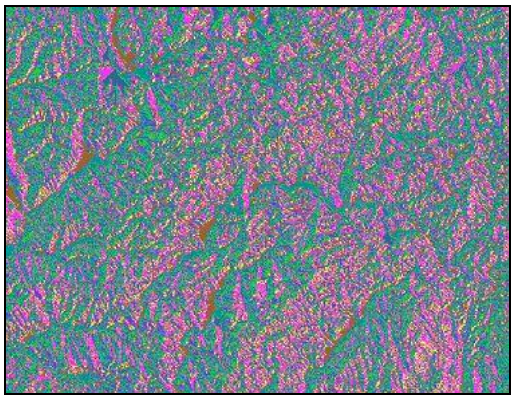
Once the digital elevation models (DEMs) are loaded into ArcView, it is desired to have the exact boundaries of the Tom river watershed. The *CE465Hydro* extension in ArcView (Moglen, 2003) performs this task with great ease. The exact geographic coordinates of the outlet point of the watershed were supplied by the NASA team. GIS Hydro initially resolves pits and flats in the original DEM and produces new themes (spatial data layers) showing *flow direction* for each pixel and *flow accumulation* that quantifies the drainage area at each location in the area defined by the DEM. It then determines the watershed using the outlet point and following the flow direction upstream along all streams and upland pixels that contribute to this outlet. Eventually all pixels that contribute their flow to this outlet point are delineated as a watershed. Figure 3-7 shows all of the above phases from the raw DEM to the final delineated watershed.

The same process was used to delineate the five subareas (A, B, C, D and E). They are shown in Figure 4-13 in Section 4.3.2 of this report. The Tom river watershed matched quite closely with the watershed delineated by the NASA team. The subareas did not match as closely but they were not expected to match perfectly because the NASA team used hand-drawn paper maps to delineate their watersheds, where the GIS relied on the DEM data. There is bound to be a certain amount of error (human and

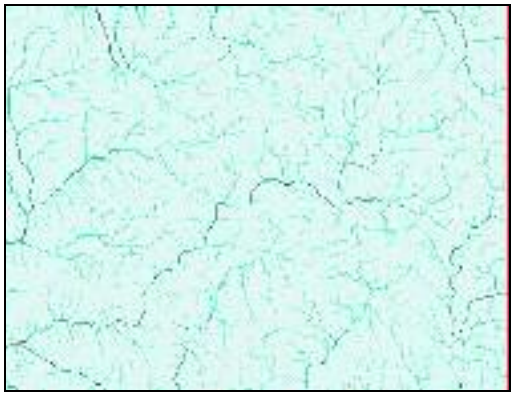
instrument) in both techniques. Table 4-2 compares the areas of the subareas to give the reader an idea of the magnitude of the errors.



(a) DEM before analysis



(b) Flow directions (each pixel has one of 8 possible flow directions)



c) Flow accumulation

(d) Tom R. watershed
(Pink dot represents the outlet point)

Figure 3-7. Different stages in ArcView analysis of the DEMs. The figures 3-7a to 3-7d represent the exact same region in EASE grid projection.

b) Determining HRUs and creating 'masks' for each

On obtaining the five subareas and loading the land cover theme (spatial data layer), the hydrological response units (or HRUs) needed to be determined (Section

4.3.3). The GIS was used to create twenty new themes by taking the intersection of four main land cover types (as specified by the NASA team) and the five subareas. They are shown in Figure 4-14. A 'mask' of each of these 20 HRUs is then created. A mask of an HRU theme is a theme that has a value of "1" for all cells (pixels) in that HRU and "No Data" (or "Null") for the rest. It must be clarified here that the pixels outside the HRU (or any other feature being masked) do not hold a value of '0'(zero), but "Null". Figure 3-8 demonstrates an example of grid masks.

1	1	1	2	2
1	1	2	2	2
1	1	2	2	2
1	2	2	2	3
1	2	2	3	3

(a) Sample grid showing portions of 3 watersheds, each denoted by a different number.

1	1	1	N	N
1	1	N	N	N
1	1	N	N	N
1	N	N	N	N
1	N	N	N	N

Mask 1

N	N	N	2	2
N	N	2	2	2
N	N	2	2	2
N	2	2	2	N
N	2	2	N	N

Mask 2

N	N	N	N	N
N	N	N	N	N
N	N	N	N	N
N	N	N	N	3
N	N	N	3	3

Mask 3

(b) Individual masks of each of three watersheds created from Fig.(a)
(N denotes "Null", NOT zero)

Figure 3-8. Example on creating individual masks from raster data

c) Map calculation

A map calculation is a mathematical operation performed on an entire theme (data layer). Each pixel value in one grid is operated upon by a mathematical function by the corresponding pixel value in the second grid to give a resultant third grid. For example, a

map calculation on two sample grids A and B may be C = A x B. It is demonstrated in Figure 3-9. Map calculations may be performed on a single grid also, e.g. $C = Log_{10} (A)$.

A =

1	1	2	2	N
1	1	2	2	N
1	1	2	2	N
1	1	2	2	N
1	1	2	2	N

B =

1	1	1	1	1
1	1	1	1	1
2	2	2	2	2
2	2	2	2	2
N	N	N	N	N

C = A x B =

1	1	2	2	N
1	1	2	2	N
2	2	4	4	N
2	2	4	4	N
N	N	N	N	N

Figure 3-9: Sample ‘map calculation’. Note that map calculation is an element-by-element, not a matrix operation (‘N’ denotes “No Data”).

Masks are helpful in map calculation. On ‘map multiplying’ a theme (say snow cover area) with a HRU mask, the resultant theme would be that of snow cover area (SCA) values exclusively for that HRU. All pixels outside the mask would still contain the “Null” value in the resultant theme. This feature has been extensively used in this project to determine the snow cover area for each day for each HRU, thereby producing the SCA time series. Map calculations (multiplications) of the 20 HRU masks with the snow cover area (SCA) theme for each of the 1511 days for which SCA data were available yielded the SCA time series.

3.7.2 Arc/Info

ArcInfo (Arc version 7.1.2) was used in this project. ArcInfo successfully supplemented ArcView by operating on big spatial data grids before they were loaded in

ArcView. The major tasks performed by ArcInfo in this study were the merging, cropping, reprojection and resampling of grids. ArcInfo uses simple commands like “merge”, “set window”, “project”, etc to perform these easy but essential operations. Their applications in the Ob R. project are explained in more detail in Sections 4.2.1 and 4.2.2 of this thesis.

Both ArcView and ArcInfo were used in the GLUE UNIX workspace provided by the University of Maryland. Both GIS softwares were obtained from the UMD GLUE UNIX system using the ‘tap’ command to access these specialized softwares.

CHAPTER FOUR

METHODS

4.1 OVERVIEW

This chapter provides detailed information on the step-by-step procedures for each major individual step of the project. The following sections deal with:

a) Digital Elevation Model acquisition and processing: The DEM dataset (GTOPO30 tiles) has been described in Section 3.5. The sections in this chapter will explain the step-by-step algorithm of downloading data from the USGS website, making appropriate changes to the dataset so it can be prepared for analysis in ArcView, and, finally, the actual processing of that data.

b) Delineation of the main Tom river watershed, the subareas and the hydrologic response units (HRUs): This part of the chapter will deal with the ArcView functions and tools (such as CE465Hydro extension) that processed the DEMs into relevant watershed and subareas.

c) Acquisition and processing of remotely sensed snow cover images from NSIDC: Again, relevant information about the NSIDC dataset has been included in Section 3.3. This chapter will deal with the creation of snow cover area (SCA) images and their intersection with the HRU masks.

d) Generation of final time series of snow cover area (SCA) for each HRU: Using the map multiplication of SCA images and HRU masks for each day with available data, to count pixels of snow/no snow in each HRU to estimate fractional snow-covered area.

4.2 ELEVATION (DEM)

The digital elevation model (DEM) data were obtained from US Geological Survey's 'Land Processes Distributed Active Archive Center' (USGS 2003). GIS software packages ArcView and Arc/Info were used in the processing of DEMs and for watershed delineation.

4.2.1 Acquiring data and converting it to ArcView-compatible format

The DEM data obtained from USGS (2003) are part of the 'Global 30-Arc-second elevation dataset (GTOPO30)'. GTOPO30 is a global digital elevation model with a horizontal grid spacing of 30 arc-seconds (approximately 1 kilometer at the equator). More information on the dataset is given in Section 3.5.

GTOPO30 has been divided into tiles. Each tile encompasses an area spanned by 40 degrees longitude and 50 degrees latitude, except for areas south of 60^{o}S i.e., Antarctica (Figure 4-1). Two adjacent tiles (E020N90 and E060N90) were selected because the Ob watershed (and also the Tom River watershed) lies within both tiles.

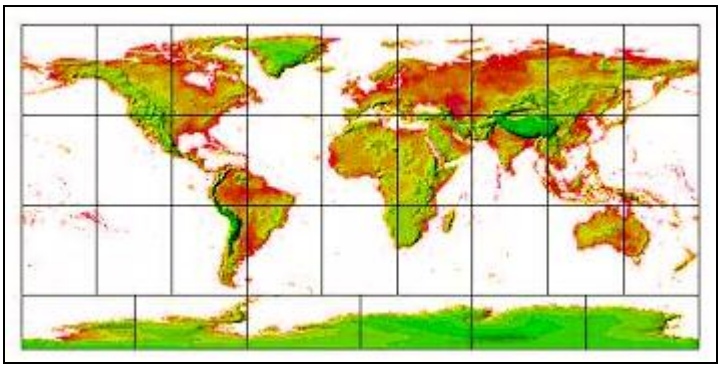


Figure 4-1. GTOPO30 tiles (Source: USGS, 2003)

Each tile was downloaded as a set of 8 files including a 57.6 MB raster file, for example: *e020n90.dem*, consisting of 4800 pixels longitude and 6000 pixels latitude, where each pixel value is given by 16 bits. The DEMs are 16-bit signed integer data in a simple binary raster. The binary file contains only pixel values. Along with each binary data file, a header file was also downloaded. The header file is a text file (Fig. 4-4) that gives georeferencing information for the binary image.

Figure 4-2. Tile E020N90

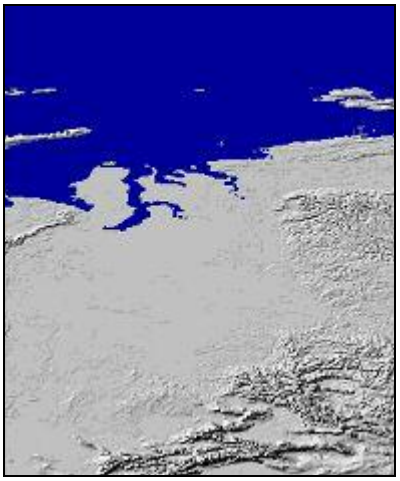


Figure 4-3. Tile E060N90

```
BYTEORDER       M
LAYOUT          BIL
NROWS           6000
NCOLS           4800
NBANDS          1
NBITS           16
BANDROWBYTES        9600
TOTALROWBYTES       9600
BANDGAPBYTES        0
NODATA          -9999
ULXMAP          20.00416666666667
ULYMAP          89.99583333333334
XDIM            0.00833333333333
YDIM            0.00833333333333
```

Figure 4-4. sample header file (for *e020n90.dem*) (Source: USGS, 2003)

Each of the terms used in the sample header file shown in Fig. 4-4 are explained as follows:

BYTEORDER: byte order in which image pixel values are stored.

M = Motorola byte order (most significant byte first)

LAYOUT: organization of the bands in the file.

BIL = band interleaved by line (note: the DEM is a single band image)

NROWS and NCOLS: number of rows and columns respectively.

NBANDS: number of spectral bands in the image (1 for a DEM)

NBITS: number of bits per pixel (16 for a DEM, 8 bits = 1 byte).

BANDROWBYTES: number of bytes per band per row (twice the number of columns for a 16-bit DEM)

TOTALROWBYTES: total number of bytes of data per row (twice the number of columns for a single band 16-bit DEM)

BANDGAPBYTES: the number of bytes between bands in a BSQ format image (0 for a DEM)

NODATA: value used for masking purposes

ULXMAP: longitude of the center of the upper-left pixel (decimal degrees)

ULYMAP: latitude of the center of the upper-left pixel (decimal degrees)

XDIM: x dimension of a pixel in geographic units (decimal degrees)

YDIM: y dimension of a pixel in geographic units (decimal degrees)

After downloading a GTOPO tile, the extension of the image file was changed from *.DEM* to *.bil* because *.DEM* has a different meaning to ArcView, whereas *.bil* indicates a

'band interleaved by line' image, which indicates a binary image file. The filename was also changed to lowercase because of restrictions in the UNIX ArcInfo software. The *.bil* and the header file (*.hdr*) were then transferred (via FTP) to the UNIX workspace provided by University of Maryland (GLUE). The header file must be present in the same file directory as the *.bil* file.

The *.bil* file is then converted to 'grid' data in ArcInfo by using the '*imagegrid*' command. The '*imagegrid*' command simply converts an image file into a raster (or grid) dataset, e.g.,

Grid > **imagegrid** *filename.bil grid1*

Thus, '*grid1*' now contains the GTOPO30 tile as a georeferenced, raster digital elevation dataset. '*grid1*' is stored as a directory of files in the Arc database.

The DEM data are stored in a 16-bit binary format. The '*imagegrid*' command does not support conversion of signed image data, therefore the negative 16-bit DEM values will not be interpreted correctly. After running '*imagegrid*', an easy fix is accomplished using the '*con*' command in Grid ($32768 = 2^{15}$ and $65536 = 2^{16}$, the example below is for a 16-bit binary format):

out_grid = **con**(*grid1* >= 32768, *grid1* - 65536, *grid1*)

The converted grid now has the negative values properly represented, and the statistics of the grid match those listed in the .STX file. The ocean mask values (-9999) are then set to NODATA using the '*setnull*' command, thereby *masking out* just the land surface in the image.

4.2.2 Processing the DEMs using GIS

After the two individual GTOPO30 tiles are imported into the ArcInfo / ArcView database, they must be merged into a single dataset because the Ob watershed lies partly in both tiles. This is achieved by using the simple *'merge'* command in Arc/Info (Grid):

Grid> *bigtile* = **merge** *tile1 tile 2*

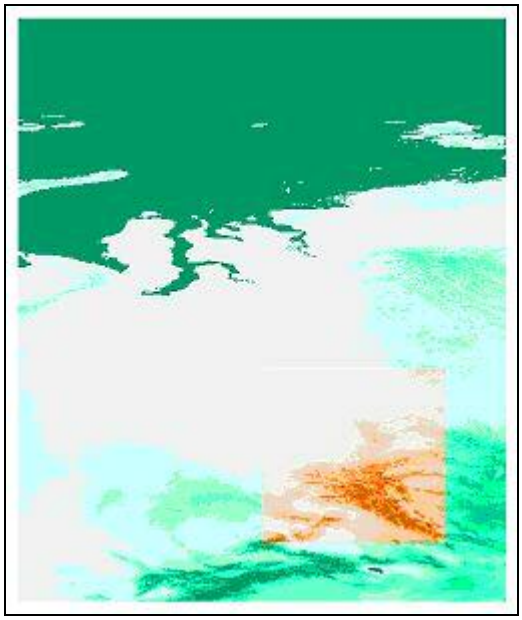


Figure 4-5. Merged GTOPO30 tiles showing area of interest (in orange)

To reduce the amount of elevation data that must be manipulated in substantial steps, the area of interest (Ob/Tom watershed) is cropped out from the merged tiles. Cropping out the area of interest (between longitudes 80^{o} to 95^{o} East, and latitudes 45^{o} to 60^{o} North) is facilitated by another Arc/Info Grid command called the *'setwindow'* command:

Grid > **setwindow** LLx LLy URx URy

Grid > *croppedgrid = bigtile*

After using these two statements in succession in the 'grid' session of ARC/INFO, the grid "*croppedgrid*" contains the raster DEM information for only the area that lies within the boundaries specified in the '*setwindow'* command.

The "*croppedgrid*" data are reprojected to EASE-grid projection (Section 3.4), Arc/Info requires a *projection file* that states the current and desired projection and other parameters such as units, etc. A sample projection file is shown in Fig 4-6.

```
input
  projection geographic
  units dd
  parameters
output
  projection lambert_azimuth
  units meters
  parameters
    6371228.0    ' radius of the sphere of reference
    0 0 0.00     ' longitude of center of projection
    90 0 0.00    ' latitude of center of projection
    0.0          ' false easting
    0.0          ' false northing
end
```

Figure 4-6. Projection file *'geog2ease'*

The important terms and contents in Figure 4-6 are explained below:

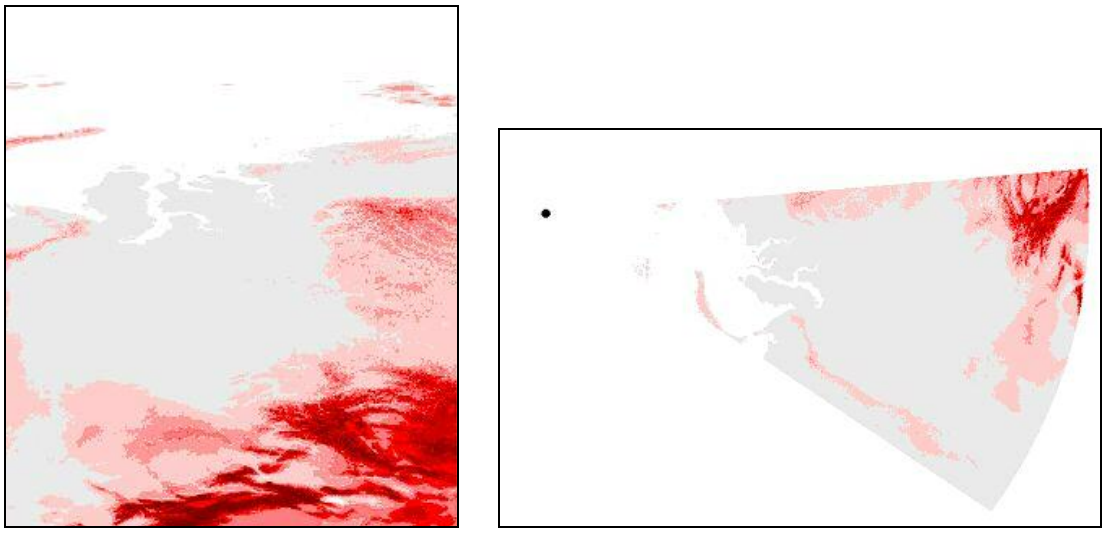
- Input projection: Geographic (latitude-longitude)
- Input units: Decimal Degrees (**dd**)
- Output (desired) projection: Equal Area Lambert Azimuth (north)
- Output units: meters
- Longitude of center of projection is specified as 0 and latitude is 90 N indicating the North pole as the center of projection (EASE grid North)

The projection file *'geog2ease'* of Figure 4-6 is a text file created in UNIX as required by Arc/Info and used successfully to re-project the cropped tile to EASE grid projection from the original geographic projection. A resampling technique (bilinear / nearest neighbor / cubic convolution) is also chosen and specified.

Grid> *easetom* = **project** (*croppedgrid, geog2ease, bilinear, 1000, 0.0, 0.0*)

where 1000 represents the desired resolution of 1000m or 1 kilometer, *geog2ease* is the projection file, *bilinear* is the resampling method chosen, and *easetom* is the output grid.

Figures 4-7a and 4-7b show how the same Ob watershed region appears in geographic and EASE grid projections. The North Pole has been plotted for reference in Figure 4-7b. As displayed in Figures 4-7a and 4-7b, on reprojecting the northmost GTOPO30 tiles to EASE grid, the North Pole, which is represented by the entire top edge of the Lat-Long image, is reduced to a single point in the EASE image. Similarly square (or rectangular) shapes in Lat-Long projection are transformed to wedge shapes in the EASE-grid projection. The DEMs are ready for hydrologic analysis at this stage.



(a) Geographic projection (North to the top of image)

(b) EASE Grid projection (The north pole is the dot towards the left of the image)

Figure 4-7. The Ob River watershed region in different map projections.

4.3 DELINEATION OF WATERSHEDS AND HRUs

This section provides details on the delineation and processing of the Tom watershed, and the subsequent generation of the 5 subwatersheds and 20 non-contiguous HRUs. It also explains how the watershed obtained from the GTOPO30 tiles was validated using the watershed provided by the NASA collaborators.

4.3.1 Digital Ob watershed outline supplied by the NASA HSB Team

As discussed in Section 3.2, the research collaborators at NASA had already completed their own mapping and subdivision of the watershed using a combination of paper maps and image processing software. The objective of the steps described in this section is to create the same subdivisions within the GIS raster framework.

The team at NASA provided the project with a digital version of the Ob River watershed. This digital map was stored as 8 smaller maps, each of which was a bulky

9MB text file that contained cells with binary values i.e., 1 and 0 for boundary and no-boundary respectively, and the X-Y (latitude, longitude) location parameters of each cell.

Microsoft EXCEL was used to convert the column text data to grid form. The given location parameters of each cell were used to place each cell value accurately with respect to each other and form a raster grid of binary values: boundary and no-boundary. This required the breaking down of each 9 MB *.txt* file into smaller files to make the process faster and comply with data limitations of MS EXCEL.

X	Y	value
74.01562	56.99219	0
74.04688	56.99219	0
74.07812	56.99219	0
74.10938	56.99219	0
74.14062	56.99219	1
74.17188	56.99219	1
74.20312	56.99219	1
74.23438	56.99219	0
74.26562	56.99219	0
74.29688	56.99219	0
74.32812	56.99219	0
74.35938	56.99219	0
74.39062	56.99219	0
74.42188	56.99219	0

(a) Sample initial text file with X, Y location and binary value

0	0	0	0	0	0	0
1	0	0	0	0	0	0
0	**1**	0	0	0	0	**1**
0	0	**1**	0	0	**1**	0
0	0	0	**1**	**1**	0	0
0	**1**	**1**	0	0	0	0
1	0	0	0	0	0	0

(b) Sample binary grid output with **1** as subarea boundary.

Figure 4-8. Example of creation of binary value watershed boundary grid from initial table of X, Y coordinates and value for each cell.

The matrix text files created in EXCEL were ingested into Arcview using the features/commands for reading ASCII grid data. A header file was created for each of the eight maps by using the pixel (cell size) dimensions and geolocation information for each map. These eight maps were then opened in ArcView. They appeared discontinuous and clumsy when all eight were opened simultaneously: the watershed boundaries did not match well along the edges of the eight maps. This was because the eight maps had different pixel dimensions (along X and Y axes) and Arcview system assumes square pixels i.e., $\Delta x = \Delta y$.

It was therefore necessary to convert all the files to a uniform square cell size. The uniform square cell-size was chosen to be the one that was most common among the eight maps: 0.015625 decimal degrees. The other maps with dissimilar cell-sizes were converted to this value by multiplying their cell-sizes by some factor and resampling the data to leave out any voids in the new matrix. For example, an earlier map with twice the Y-dimension (0.03125) and the chosen X-dimension (0.015625) would be converted into a new map with twice the number of rows (along Y-axis) and a new Y-dimension of 0.015625. Figure 4-9 demonstrates this example.

<u>Original</u>
X-dim = 0.015625
Y-dim = 0.031250

1	0	0
0	1	0
0	1	0

<u>Converted</u>
X-dim = 0.015625
Y-dim = 0.015625

1	0	0
1	0	0
0	1	0
0	1	0
0	1	0
0	1	0

Figure 4-9. Example of re-sampling of uneven maps to the chosen dimensions

MATLAB was used to re-sample these relatively big matrices (of the order 512 x 512). The final, corrected Ob watershed map appears as shown in Figure 4-10. This simple binary map did not distinguish between river channels and watershed boundaries. In addition, there were numerous discontinuities in the boundaries, due to the sequence of steps used in creating the map. Due to changes in personnel, software and equipment, the NASA collaborators were unable to reproduce or correct their digital data.

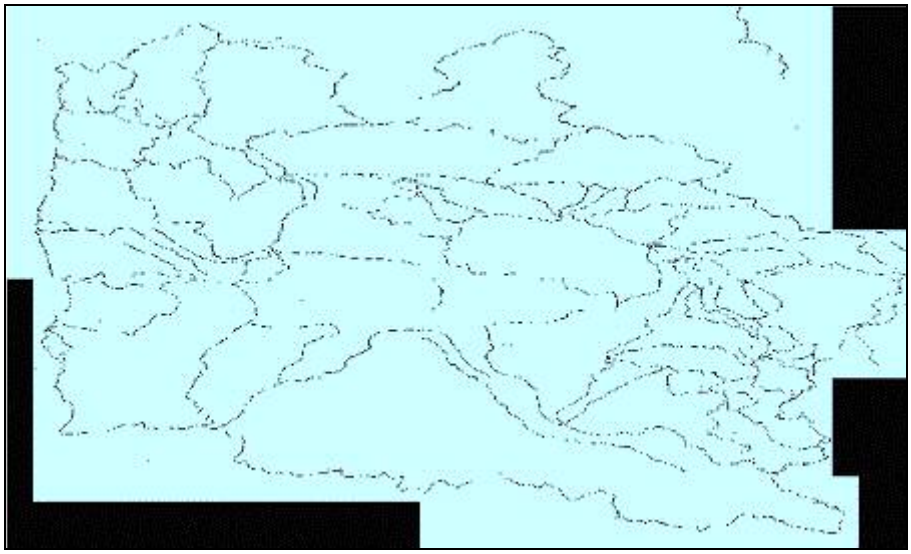


Figure 4-10. Mosaic of eight maps given by NASA as they appear in Arcview

4.3.2 Using the GTOPO30 tiles to delineate the Tom watershed

Although the binary boundary map did not prove very useful for digital analysis (boundaries were not continuous at all places, creating open polygons), it did allow a visual comparison of the NASA team's choice of subareas with those delineated automatically as described in Section 3.7.1.

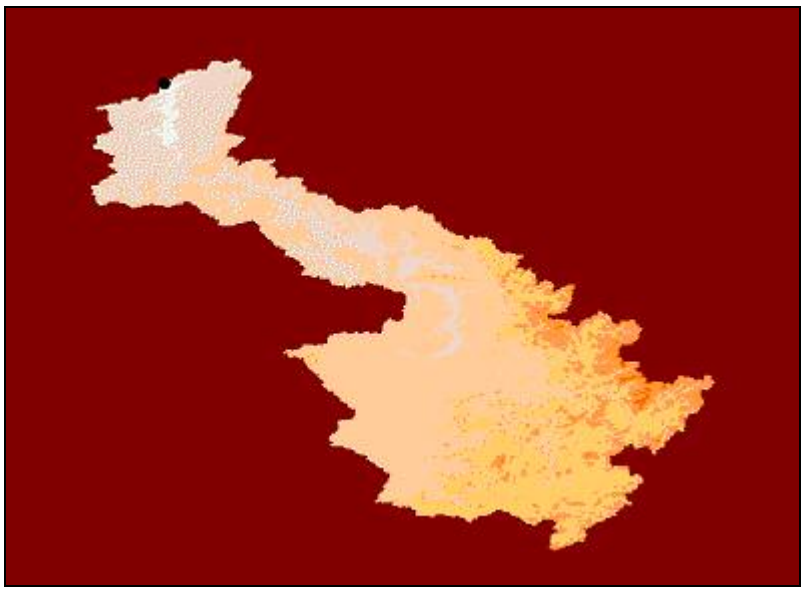
ArcView 3.0b has an extension called "CE465Hydro" (Dr. Glenn Moglen, personal communication, 2003) that has some useful tools to delineate a watershed from

a DEM provided the exact basin outlet pixel is identified. The NASA collaborators provided the project with the geographic coordinates of the outlet of the Tom River watershed. This extension was used successfully to get an accurate raster grid of the Tom watershed. The CE465Hydro performs the following operations on the DEM grids imported in ArcView:

- FILL : For hydrologic modeling purposes, DEMs must be devoid of pits except in areas with actual inland catchments. The FILL process in ArcView / Arc/Info eliminates artificial pits.
- FLOW DIRECTION : This is a measure of the direction in which a given cell is likely to discharge any incident water. Eight flow directions – leading to one of any of 8 cells surrounding any given cell in a grid – are possible (Figure 3-7b).
- FLOW LENGTH : This is a measure of the distance along the flow path (determined by the flow direction grid) from a given cell to its drainage basin outlet. A detailed stream network is generated on running this operation as shown in Figure 3-7c.

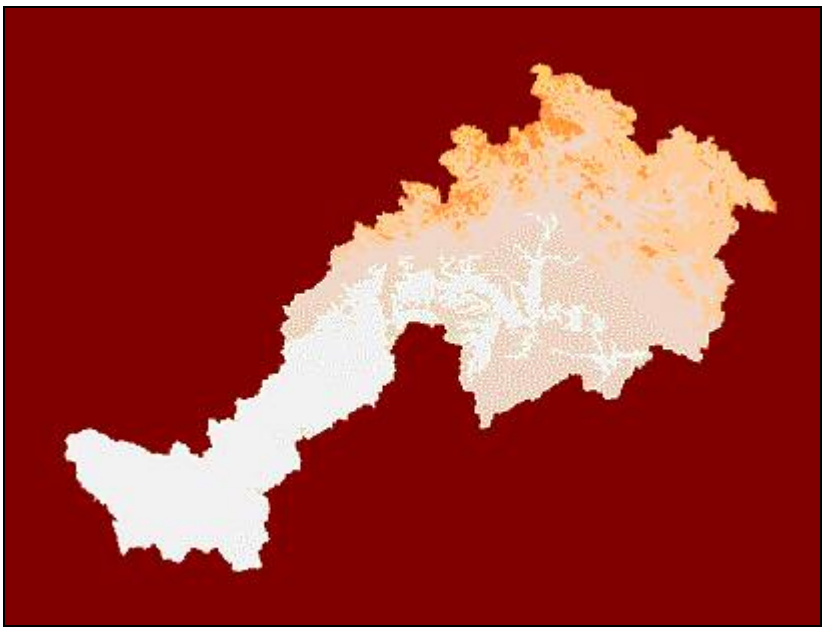
This extension generated the Tom River watershed using the DEM file and the watershed output co-ordinates. The resultant Tom River watershed is shown in Figure 4-11a (Geographic projection) and in Figures 4-11b and 3-7d (EASE grid projection).

The five subareas (A, B, C, D, E) were also created similarly by using the geographic coordinates of the outlet points for each of them as defined by the NASA HSB team in their previous analysis. Figure 4-12 displays the Tom R. watershed and its 5 subareas as viewed in ArcView in EASE grid projection.



(a) Geographic (Lat-Long) projection

(The black dot marks the outlet of the watershed, North is towards the top of the image)



(b) EASE-grid as viewed in Arcview

(The north pole is to the left of the image)

Figure 4-11. Tom R. watershed in different projections for visual comparison of shape.

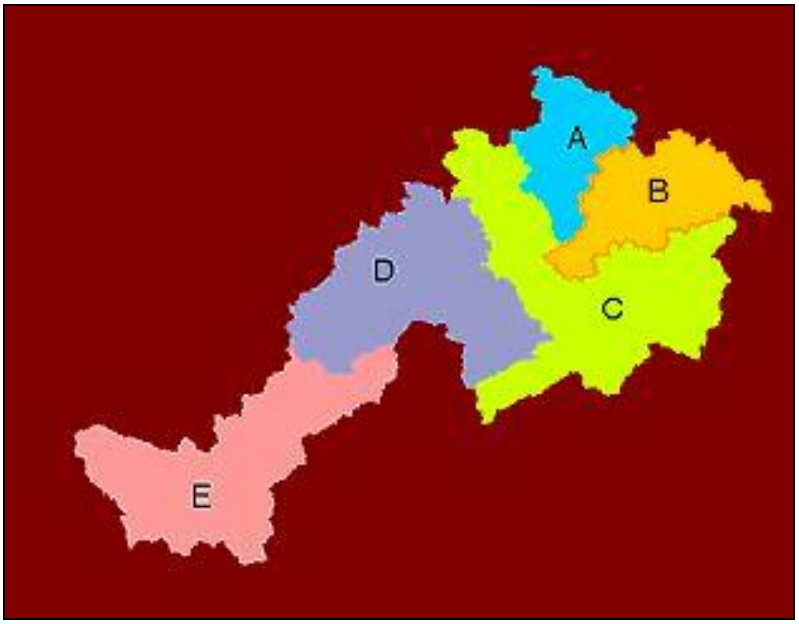


Figure 4-12. The five subareas in EASE grid projection

These subareas and the main Tom river watershed generally match well with the paper maps used by the NASA team. The GIS analysis indicates a larger overall watershed area than the paper map analysis, probably due to the more precise capture of contributing areas near the watershed divide by the GIS compared to visual analysis. The increase in size is contributed by subareas A, C, D, and E, and only partially balanced by a decrease in subarea B. The smallest subarea, A, is 78% larger in the GIS analysis, probably due to a combination of additional area at the overall watershed divide and a different placement of the divide between subareas A and B. These subareas may be more accurate than the paper maps (which were created by hand) because they were generated by ArcView (assuming the elevation data from USGS are accurate). Table 4-2 displays the subarea area comparison between the NASA HSB analysis and GIS-generated maps.

Table 4-2. Subarea area comparison

	Area in km^2			
Subarea	NASA HSB team	GIS analysis	Difference	% Difference
A	3320	5929	2609	78
B	10592	8874	-1718	-16
C	15888	18251	2363	15
D	13600	15281	1681	12
E	13600	15672	2072	15
Total	57000	64007	7007	12

4.3.3 Overlaying landcover data to determine Hydrological Response Units (HRUs) and comparison with NASA's HRUs

The landcover information for the region obtained from GLCF (and explained in detail in Section 3.6) are cropped and reprojected to EASE-grid projection using the same commands in Arc/Info as used for the GTOPO30 DEM tiles i.e., *'setwindow'* and *'project'* except that "nearest" is used for the resampling method rather than "bilinear", because the data are categorical.

The HRUs are determined by using similar land cover types over each of the five subwatersheds. There are four major landcover types for each watershed:

- Deciduous Needleleaf
- Mixed Forest
- Woody Savanna
- Croplands

These are the four landcover types used by the NASA HSB Team to obtain their HRUs for the same watershed. The landcover data obtained from GLCF initially had 13 different categories of land cover. These were grouped together based on similar

attributes and reduced to the four landcover types used by the NASA HSB team. Twenty HRUs are thus created: FIVE areas x FOUR landcover types on each.

Since the landcover over the region is not contiguous, the HRUs created using them are also non-contiguous. Figure 4-13 displays the 20 HRUs as viewed in ArcView in EASE-grid projection.

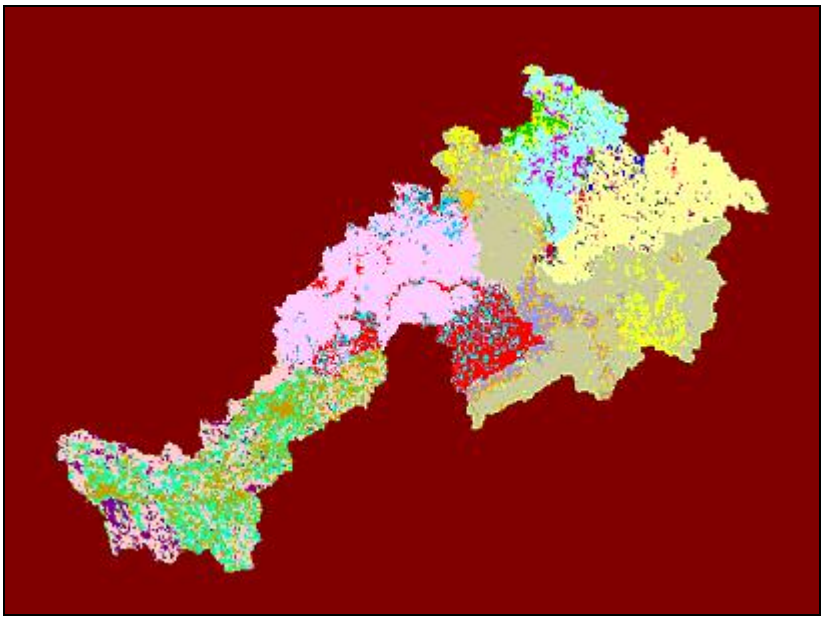


Figure 4-13. All 20 non-contiguous HRUs in EASE grid projection

These 20 HRUs match satisfactorily with the NASA HRUs, but not perfectly. The areas do not match exactly, as is to be expected with the different analysis techniques. Also, the NASA HSB Team has 19 HRUs: the “croplands” HRU in subarea B was ignored because of its negligible area, whereas the ArcView analysis determines that HRU to be nearly 3% of the total area in subarea B, making it significant enough to be included as an individual HRU in the Tom watershed. A comparison of the area of each HRU from the

NASA dataset and the project's dataset is displayed in Table 4-3. Table 4-4 gives more statistics on the HRUs and the subareas.

Table 4-3. Comparison of area of each HRU estimated by the NASA HSB Team and determined by GIS analysis.

Sub	HRU	HRU #	Area in km^2		HRU as % of Subarea, by area	
area	Land Cover type		NASA	GIS	NASA	GIS
					(%)	(%)
A	Deciduous Needleleaf	1	564	595	17	10
A	Mixed Forest	2	2058	4024	62	68
A	Woody Savannas	3	332	687	10	12
A	Croplands	4	365	623	11	10
	TOTAL AREA:		3320	5929	100	100
B	Deciduous Needleleaf	5	1059	308	10	3
B	Mixed Forest	6	9003	8050	85	91
B	Woody Savannas	7	530	255	5	3
B	Croplands	20	N/A	261	N/A	3
	TOTAL AREA:		10592	8874	100	100
C	Deciduous Needleleaf	8	1588	1484	10	8
C	Mixed Forest	9	10645	13328	67	73
C	Woody Savannas	10	477	1577	3	9
C	Croplands	11	3178	1862	20	10
	TOTAL AREA:		15888	18251	100	100
D	Deciduous Needleleaf	12	1360	214	10	1
D	Mixed Forest	13	9520	10164	70	67
D	Woody Savannas	14	1360	1566	10	10
D	Croplands	15	1360	3337	10	22
	TOTAL AREA:		13600	15281	100	100
E	Deciduous Needleleaf	16	680	1147	5	7
E	Mixed Forest	17	4080	6342	30	40
E	Woody Savannas	18	2040	4118	15	26
E	Croplands	19	6800	4065	50	26
	TOTAL AREA:		13600	15672	100	100

Because a GIS map of the NASA team's HRUs is not available, a direct comparison of the exact locations of these non-contiguous regions is impossible. In fact, because the HRUs are treated as lumped entities in the PRMS model, the exact location of each pixel included in a HRU is irrelevant to the model. Therefore, it is appropriate to compare model results for the NASA team's HRUs to observations for the University of Maryland HRUs.

4.3.4 Effect of order of operations: Watershed delineation and reprojection

In the procedure described above, the watershed and its subareas were delineated in the geographical projection DEM, and then these binary masks were reprojected into the EASE-grid. These two processes are not necessarily invertible, because of the difference in resampling techniques used in reprojecting the watershed and subarea masks (nearest) as opposed to the DEM (bilinear).A simple experiment was performed at this stage of the project to get an estimate of the change in spatial data information if the watershed and subareas were delineated after the DEM was reprojected to EASE-grid.

The cropped out DEM for the Tom River region (inset orange portion of Figure 4-5) was entirely reprojected to EASE-grid using the *'project'* command and the projection file *geog2ease* (Figure 4-6) in ArcInfo. Thereafter the outlet point pixel was sought in this reprojected DEM and the CE465Hydro extension in ArcView was used to delineate a fresh Tom River watershed in EASE-grid. This new Tom R. watershed was compared with the one delineated earlier in Lat-Long projection and then converted to EASE-grid. It was observed that there was no perceivable change in the number of pixels. They were equal to 64007 (1km x 1km) pixels in either case.

The conclusion of this experiment was that no change was observed because of the relatively small size of the Tom R. watershed compared to the entire northern region of the Northern hemisphere. Also, the study region lies between latitudes 45^{o}N and 60^{o}N, where the amount of distortion in cells upon reprojection is probably not significant enough to cause a significant change in an areal spread of 64,000 km^2. Distortion would increase as one approaches the North Pole. Another issue was the relatively small cell size. The cell size of the DEM tiles is 30-Arc-second (or 0.5 minute) of a degree. A bigger cell size (say, 2 degrees) would show higher distortion of cells and probably some visual and areal change in the Tom R. watershed upon reprojection.

4.4 NSIDC SNOW DATA ANALYSIS

This section of this thesis discusses the step-by-step algorithm of developing the final product of the GIS analysis: The SCA time series. At this stage of the project, the 20 HRU masks are ready for *map calculation* (intersection) with the NSIDC snowcover data layers (Surface Type Masks). Acquisition of the snowcover data from the National Snow and Ice Data Center (NSIDC) is explained in detail in Section 3.3. This section discusses unanticipated difficulties faced in creating the appropriate snow/no snow data layer (theme) from the NSIDC dataset with regard to proper georeferencing of the snowcover data and resolution issues.

4.4.1 Georeferencing NSIDC data to correctly align with other datasets

Regular orders of Surface Type Mask datasets from the NSIDC catalog revealed that each new dataset, accompanied with its independent header (*.hdr*) file, aligned in a

different position along the Y-axis. The X-axis referencing appeared to be constant and correct in all when compared to other spatial datasets such as elevation (DEMs from the USGS GTOPO30 website) and land cover (from the GLCF website). On closer inspection of the header file of each separate dataset, it was discovered that the parameter that governed the placement of the snowcover grid along the Y-axis (‘*ulymap’*) had irregular values that almost appeared to be random. This flaw had to be corrected.

A sample snowcover dataset (in EASE-grid as provided by NSIDC) was selected over the Tom R. area for an arbitrary date. This dataset was imported in ArcView along with the EASE-grid projection of the entire two GTOPO30 tiles shown in Figure 4-7b. The first few NSIDC snow data sets obtained did not appear properly positioned; they did not overlap the DEM data for the region. In order to test the georeferencing information created by NSIDC’s automated data server, a NSIDC snowcover dataset was specified to have the North Pole as its upper left corner. The North Pole is a natural choice of reference point, because it has (X,Y) coordinates (0,0) in the EASE-grid’s polar azimuthal coordinate system. The NSIDC snowcover dataset aligned correctly in the X-direction but was significantly shifted from its proper position along the Y-axis (Figure 4-14). The entire snowcover grid would have to be translated along the Y-direction by a certain number of rows to be correctly positioned in the EASE-grid coordinate system..

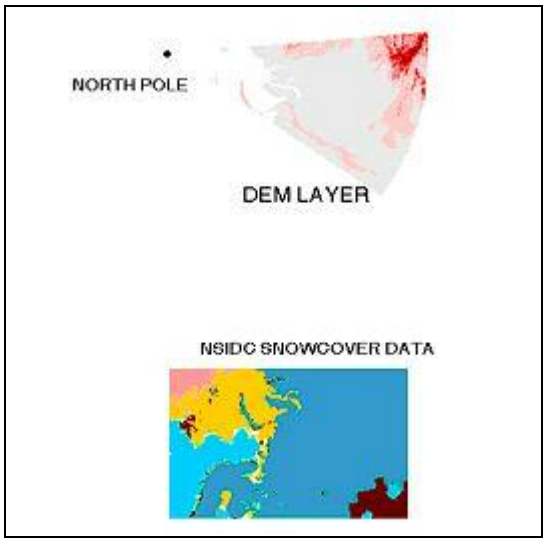


Figure 4-14. Typical location of a sample NSIDC data layer with respect to a georeferenced DEM layer as viewed in ArcView. The North Pole (shown as a dot) has (X,Y) coordinates (0,0) in the EASE-grid projection. The NSIDC data layer was specified to have its upper left corner at the North Pole. This image demonstrates an error in the NSIDC's automated creation of the georeferencing header file.

The header file for the snowcover dataset had to be corrected so that the grid properly aligns with the DEM grid, and thereby gets correctly georeferenced. Figure 4-15 shows a typical header file of the NSIDC snowcover dataset.

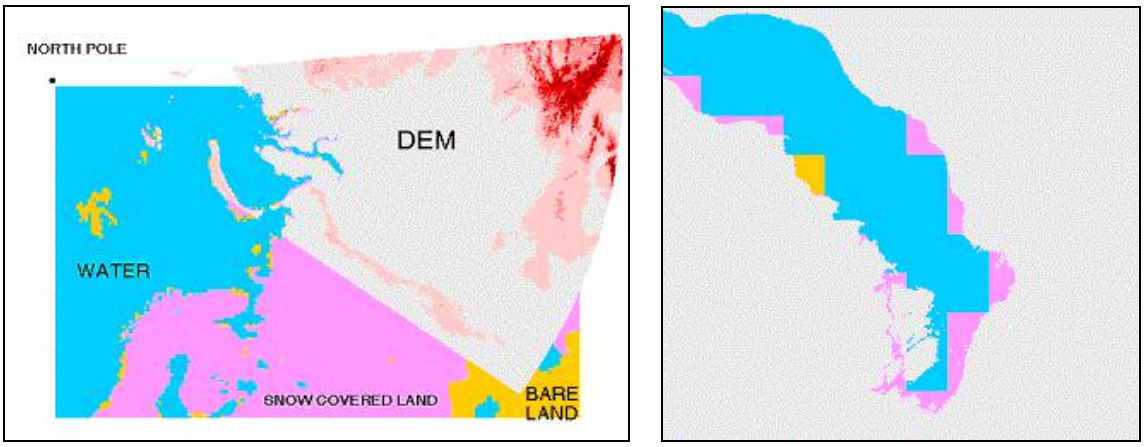
```
Sample GIS header file for subsetted data from GISMO request #2003925183753

units meters
nrows 226                          from GISMO Subset Grid Height
ncols 308                          from GISMO Subset Grid Width
nbits 16                bytes per pixel * 8 *** Varies by file, may need to be changed ***
byteorder M                        (big-endian)
layout bsq                         band sequential - one band.
ulxmap 2978021.97                  in meters = (ulx - 0.5 - centerx)*(mapscale(km)/grid
columns per map unit)*1000
ulymap -7605487.085   in meters = -(uly + 0.5 + centery - gridheight)*(mapscale(km)/grid
rows per map unit)*1000
xdim 5013.505          in meters/pixel = (mapscale(km)/grid columns per map unit)*1000
ydim 5013.505          in meters/pixel = (mapscale(km)/grid rows per map unit)*1000
```

Figure 4-15. Sample header file for a NSIDC snow dataset

Most of the terms in this sample NSIDC data header file are the same as in the GTOPO30 tiles header file shown in Figure 4-4. The remaining terms are explained as follows:

- nrow and ncols : represent the number of rows and columns in the dataset ordered.
- nbits : depicts the number of bytes per pixel.
- byteorder ; is M for *Motorola Byte Order* (Big-endian, most significant byte first).
- bsq : Band Sequential layout
- ulxmap : the X-coordinate of the Upper Left corner of the downloaded dataset (in meters).
- ulymap : the Y-coordinate of the Upper Left corner of the downloaded dataset (in meters).

The snowcover theme was translated in the Y-direction by changing the *ulymap* parameter by multiples of the cellsize, thus effectively moving the entire grid by a fixed number of rows. This task was a trial and error approach, and was performed repeatedly with new values of *ulymap* until both images (snowcover and DEM) matched perfectly with each other. Fortunately there is a long, jagged and curved coastline along northern Europe that facilitated the visual alignment of the two data layers. To ensure accurate registration, the images were zoomed to the corner of specific critical pixels leaving room for no doubt that the snowcover dataset matched the DEM dataset. Figure 4-16a shows the aligned coastlines and Figure 4-16b shows the two images zoomed in to the extent of individual pixels.



(a) Northern Europe and Russia (EASE-grid) (b) Zoomed to pixels

Figure 4-16. Aligned coastlines of the USGS DEM and NSIDC snowcover dataset indicating the georeferencing of the snowcover theme.

At this stage the value of *ulymap* that correctly georeferences the NSIDC snowcover dataset is determined. Thereafter the final dataset required for the generation of the desired SCA time series was ordered for all days between 24 July 1981 to 31 December 1985. A snowcover image from this new dataset for a randomly chosen date from among the available days was aligned with the DEM layer to double check and make any possible corrections to the new header file that accompanied this new dataset. The snowcover dataset was now correctly georeferenced and ready for *map calculation* with the 20 HRU masks for all 1511 days with available snowcover data. It should be noted that this georeferencing step was necessary only because of an apparent processing error in the NSIDC GISMO. The purpose of georeferencing in header files is to guarantee that digital data are correctly and precisely aligned in their coordinate systems. NSIDC will be alerted to this problem.

4.4.2 Final map calculations and generation of SCA time series

There are a total of ***2,548*** 5km x 5km pixels in the entire Tom R. watershed. Table 4-4 shows the number of 5km x 5km pixels for each subarea and HRU, and comparisons in area computed from 5km x 5km pixels and 1km x 1km pixels. It may be observed from these tables that the computed areas do not match perfectly for both resolutions. In fact, subarea B has wide ranging disparities. This is due to the imperfect matching of the watershed boundary pixels, which are at a resolution of 1km, and the NSIDC data, which has a resolution of 5km. The 5km pixels, when overlaid on the 1km resolution image, cover many extra 1km pixels than they should be covering in certain areas, and are unable to cover some 1km pixels that they should be covering. This causes a change in the computed area for different resolutions. Also refer to Table 4-3 for comparison of 1km resolution HRU areas with the corresponding areas computed by the NASA team.

Figure 4-17. Sample of the watershed layer overlaid on a typical snowcover dataset. This snowcoverage is for 15 May 1985 (selected at random).

Table 4-4. Comparison of HRU areas computed from datasets with different resolutions.

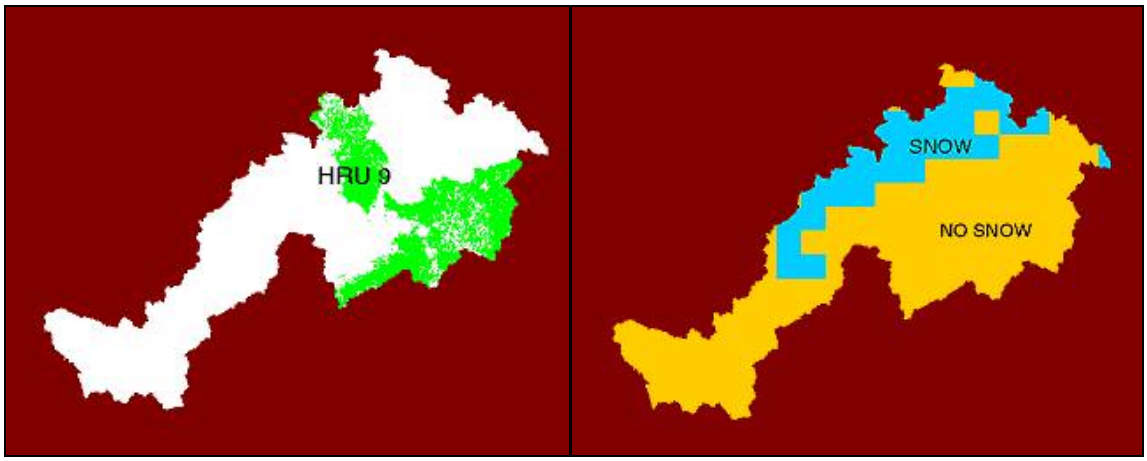
Subarea	HRU	Number of 5km x 5km pixels	Area (km^2) from 5km pixels	Area (km^2) From 1km pixels	Error (% with respect to 1-km resolution)
A	1	25	625	595	5.04
A	2	165	4125	4024	2.51
A	3	22	550	687	-19.94
A	4	24	600	623	-3.69
A	**Total**	**236**	**5900**	**5929**	**-0.49**
B	5	7	175	308	-43.18
B	6	318	7950	8050	-1.24
B	7	11	275	255	7.84
B	20	18	450	261	72.41
B	**Total**	**354**	**8850**	**8874**	**-0.27**
C	8	67	1675	1484	12.87
C	9	524	13100	13328	-1.71
C	10	59	1475	1577	-6.47
C	11	74	1850	1862	-0.64
C	**Total**	**724**	**18100**	**18251**	**-0.83**
D	12	8	200	214	-6.54
D	13	407	10175	10164	0.11
D	14	63	1575	1566	0.57
D	15	136	3400	3337	1.89
D	**Total**	**614**	**15350**	**15281**	**0.45**
E	16	41	1025	1147	-10.64
E	17	266	6650	6342	4.86
E	18	155	3875	4118	-5.90
E	19	158	3950	4065	-2.83
E	**Total**	**620**	**15500**	**15672**	**-1.10**
TOTAL		**2548**	**63700**	**64007**	**-0.48**

The 5km NSIDC snowcover dataset is resampled from an initial 25km resolution satellite image series. Each original snow pixel appears to take up twenty-five 5km x 5km snow cover area (SCA) pixels. Figure 4-17 displays a sample snowcover theme overlaid

on the Tom R. watershed as viewed in ArcView. The blue areas in Figure 4-18 are the snowcovered pixels and the area covered by the yellow pixels is that which is devoid of snow. This snowcover data theme is for 15 May 1985. Figure 4-17 gives a sense of how one edge of a blue snow pixel, which actually consists of five 5km snow pixels along the edge, is equivalent to twenty-five 1km watershed pixels (when visually compared to the boundaries of the watershed).

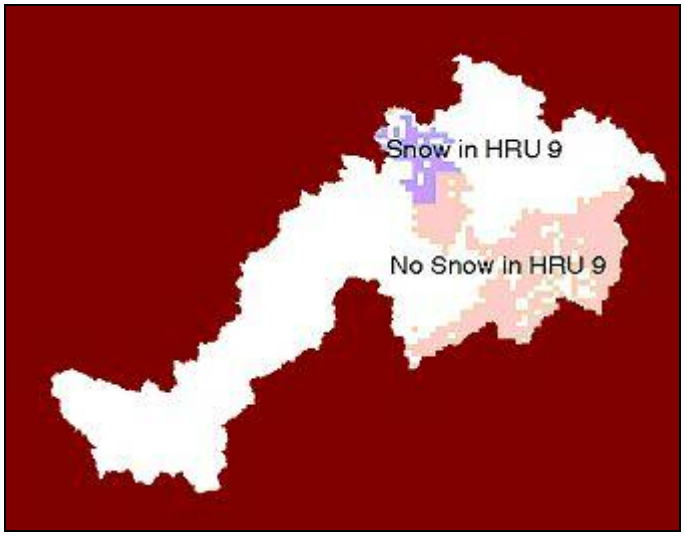
Each of the HRU masks are now *map multiplied* with the snowcover theme, and the resultant themes portray the number of snowcovered pixels in each respective HRU. For example, Figure 4-18 shows a sample *map multiplication*. The mask of HRU 9 is multiplied with the snowcover data layer for 15 May 1985. In the resultant grid (Figure 4-18c), the purple pixels are the snowcovered pixels in HRU 9 for that day, and the gray ones are not snowcovered in HRU 9 *only* on that day. It should be noted here that all pixels outside HRU 9 remain untouched by this map calculation on the HRU 9 mask. Figure 4-18 also gives a sense of the difference in resolutions of 1km, 5km and 25km themes. Figures 4-18a and 4-18c show the same HRU 9 in resolutions of 1km and 5km respectively. It is evident that HRU 9 has lost some area upon being converted to 5km x 5km pixel size (1.71% to be precise, Table 4-4).

This procedure is repeated for each HRU and each day with available snowcover data between 24 July 1981 and 31 December 1985. No map calculations were required for the times when the entire watershed was snow covered i.e., the intense winter months. The snowcovered area (SCA) for such days would simply be 100% or the total number of pixels in each HRU.



(a) HRU 9 in the Tom R. watershed
1 km resolution

(b) Snowcover data for 5/15/85
5 km resolution



(c) The output theme after map multiplication. The pink pixels indicate the snow-covered portion of HRU 9. 5 km resolution

Figure 4-18. An example of a typical map multiplication involving a HRU and the snowcover on a randomly chosen date. Note that pixels outside HRU 9 remain untouched.

After all the map calculations, the SCA time series generated for each of the twenty HRUs are discussed and analyzed in detail in Chapter 5. Figure 4-19 gives a sample of a portion of a generated time series. This is for HRU 9 for days with available snowcover data between dates 1/1/1982 and 12/31/1983, i.e., two full years.

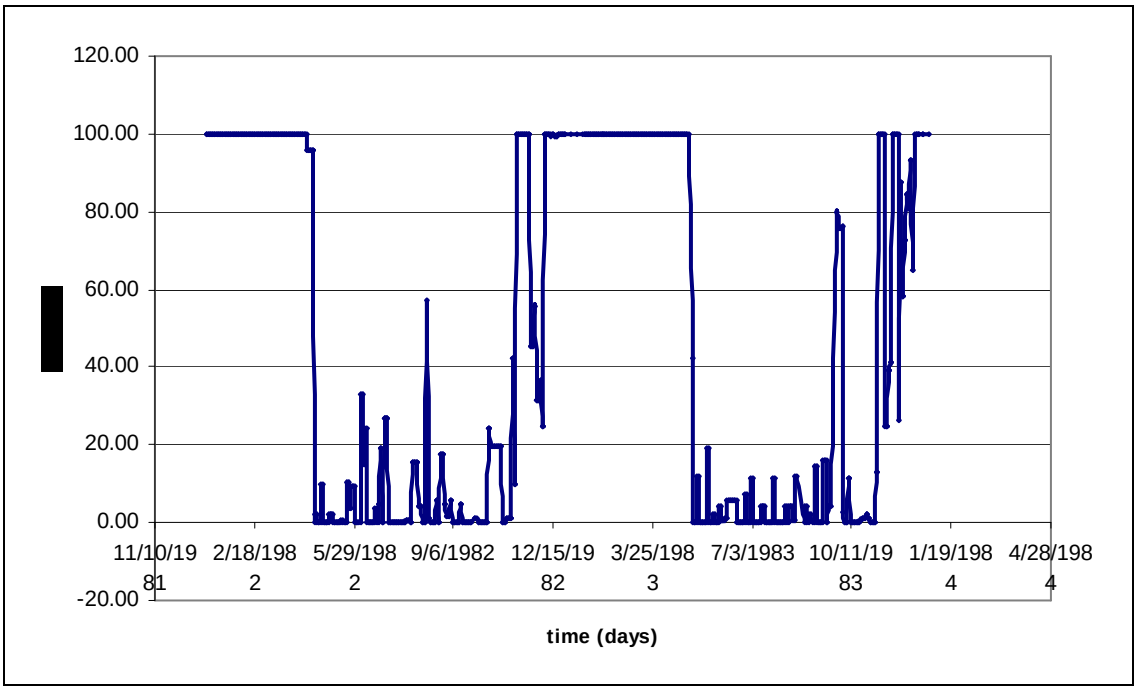


Figure 4-19. Sample time series generated from map calculations of HRUs with snowcover data. Figure shows SCA time series for HRU 9 for days between 1/1/1982 and 12/31/1983.

As observed in Figure 4-19, there are numerous spikes in the time series and the snowmelt period appears to be relatively short. The SCA drops from 100% to 0% in virtually 7 to 9 days. The possible causes for such results will be discussed in Chapter 5, where the project results are also compared with the PRMS time series generated by the NASA team.

CHAPTER FIVE

RESULTS

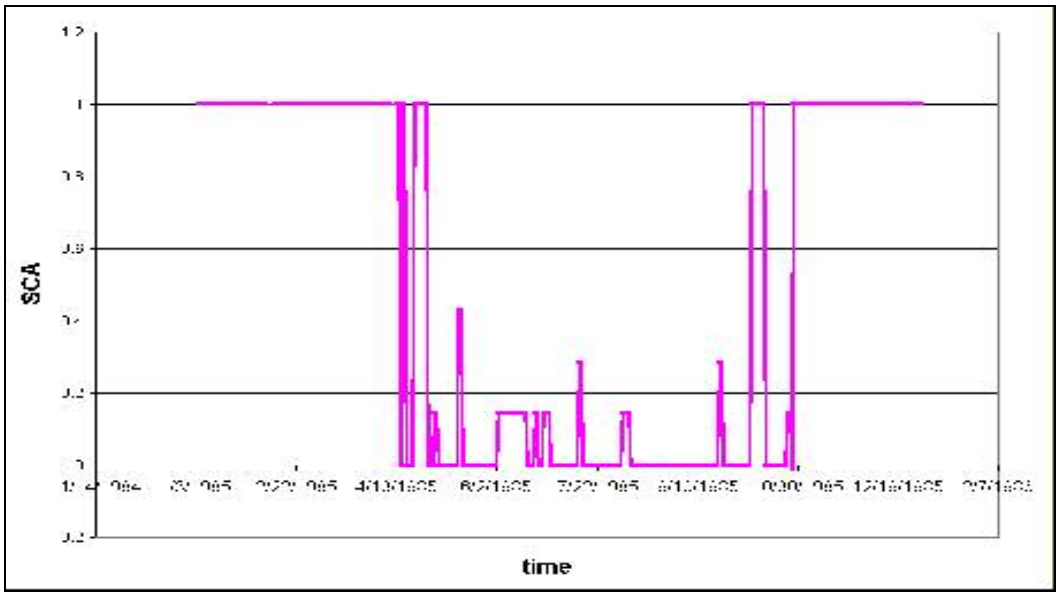
5.1 OVERVIEW

After executing the steps detailed in the algorithms in Chapter 4 and making optimal use of the datasets described in Chapter 3, the SCA time series from satellite data were generated. This chapter is devoted to study the characteristics of the created time series, compare them with the corresponding preliminary results of the model (PRMS), and also perform analyses on the resultant series to evaluate their usefulness and further improve their quality and accuracy.

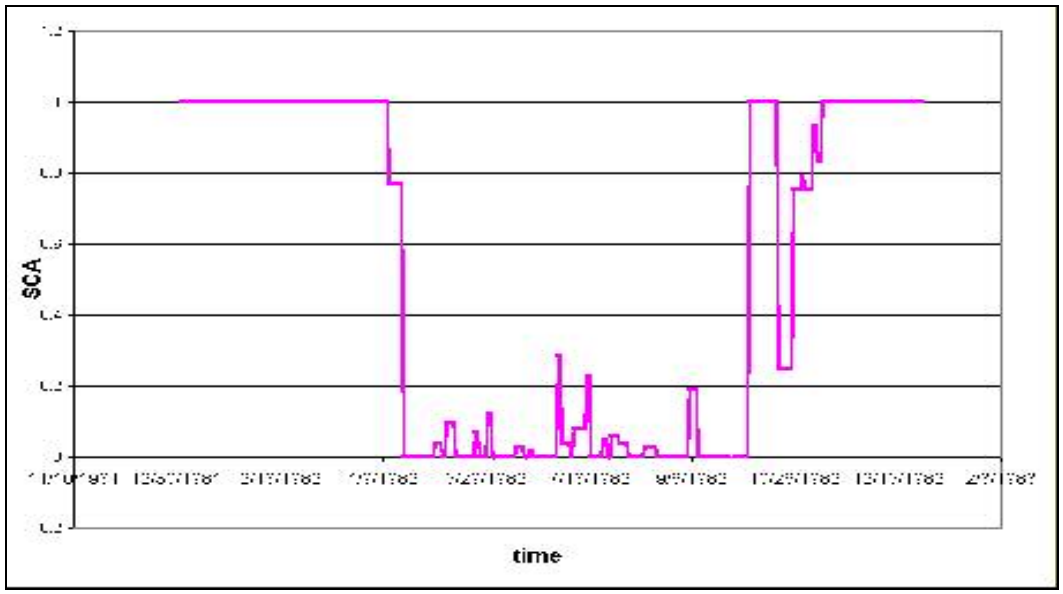
5.2 THE SATELLITE GENERATED SCA TIME SERIES

At this stage of the project one hundred SCA time series were ready for assessment (100 = 20 HRUs in watershed x 5 calendar years). Approximately 20,000 map calculations were performed to generate the time series, excluding days when the entire Tom R. watershed was snow covered or when it was fully devoid of any snow. Such special days had a SCA of 1.00 or 0.00, respectively. Technically, each individual data element in the generated time series is the direct result of an individual map calculation.

All 100 time series and the comparison of 95 of them with available corresponding time series from the NASA HSB team's PRMS model are displayed in Appendix A. Figure 5-1 shows two selected time.



a) HRU 5 (Area = 175 km^2), 1985



b) HRU 19 (Area = 3950 km^2), 1982

Figure 5-1. Samples of SCA time series generated for 20 HRUs for each of 5 years.

The satellite-generated SCA time series are shown as discontinuous points in Appendix A because data on some days were missing in the NSIDC snowcover data catalog for the region. These days are listed in Appendix D. Due to the presence of these

gaps in the satellite data the SCA time series are represented as discontinuous points when being compared to the continuous output from PRMS, which gives SCA values for every three hours. In Figures 5-1 a and b, the discontinuous points have been joined by curves to impart a better understanding of the output. The properties of these time series will be assessed in the following sections.

5.2.1 Characteristics of the SCA Time Series

As displayed in the examples in Figure 5-1, the snow depletion portions of the time series are not as gradual as one would normally expect them to be. As discussed in Section 2.5, snow-covered area is generally observed to decline over a period of weeks or even months, due to spatial variability in snow mass and melt energy. The sudden drops in the SCA curves during the melt season from completely covered to completely uncovered in a few days do not represent a very hydrologically realistic picture of the actual happenings on the ground. The possible reasons for this occurrence will be explored in Section 5.3.

The remaining portions of the time series, with unexpected spikes in the SCA in the middle of summer, and then the instant rise from zero SCA to 100% SCA in just 2-3 days, is also not in line with our knowledge of the snow accumulation process. Possible reasons for these findings will also be discussed in the following sections.

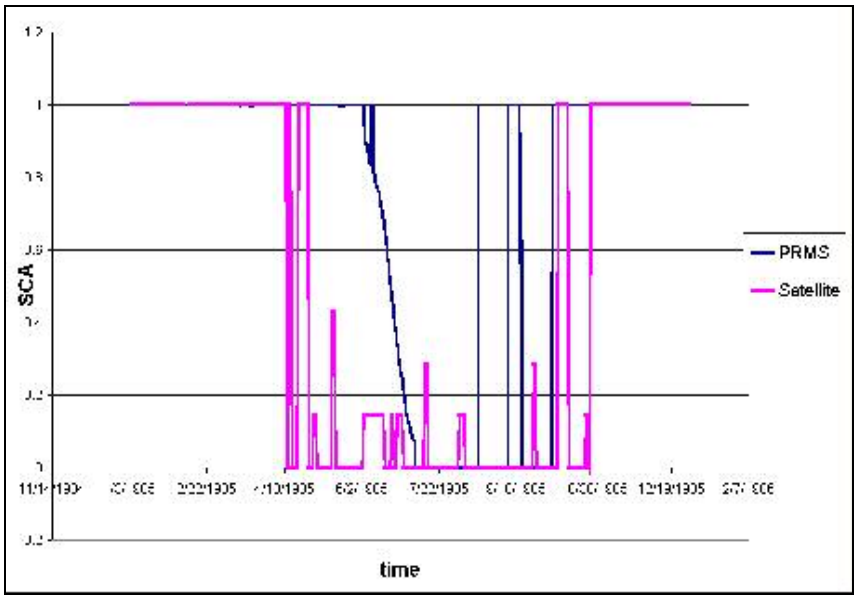
5.3 COMPARISON WITH PRMS OUTPUT TIME SERIES

As mentioned earlier, Appendix A contains the figures of 95 SCA time series, where the results from satellite data are compared with corresponding outputs from the

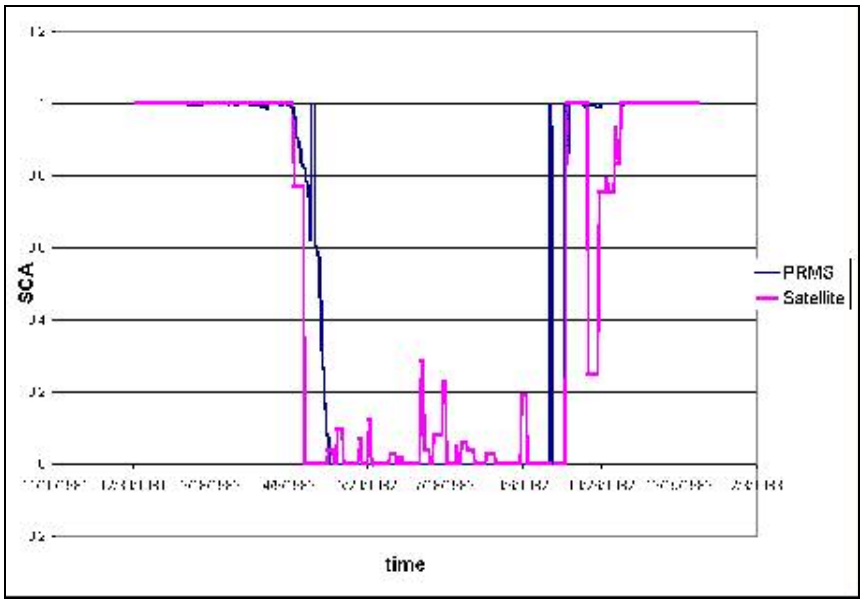
PRMS model as used by the NASA HSB team. It must be remembered again that the PRMS results are still uncalibrated and very preliminary. Even as this thesis is being written, the NASA HSB team continues to improve the output of the model by improving the Snow Water Equivalent (SWE) inputs to the model. The last five SCA time series in Appendix A are for HRU 20 in subarea B. This HRU was not included in the NASA HSB team's model as it was deemed insignificant, and no PRMS SCA outputs were provided for it. Therefore, results for HRU 20 from only the satellite data are shown from Figure A-96 to Figure A-100. Figure 5-2 shows the comparison of satellite generated SCA time series with the corresponding series generated by PRMS for the same two examples displayed in Figure 5-1. The example in Figure 5-2a is a poor match, while the example in Figure 5-2b is a relatively good match compared to the other results.

The comparative SCA time series figures in Appendix A reveal that many of the PRMS generated time series also show a sudden drop in the SCA (Figures A-10, A-30, A-58, etc.). Almost all PRMS output series show a sudden rise in SCA from 0 to 100%. In general, the model HRUs are snow-free throughout the summer and do not display the periods of temporary partial snowcover exhibited by the satellite analysis. A visible phase difference in the snow depletion and snow accumulation timing between the model and satellite SCA (e.g., Figures A-2, A-9, A-29, etc.) is generally evident. It is observed that on an average the satellite SCA drops from 100% to zero earlier than the PRMS output during the melt season and rises back to 100% SCA later than the PRMS output during the snow accumulation season. However, some of the series show a better match either during snow depletion (Figures A-18, A-33, etc.), or during the accumulation (Figures A-16, A-70, etc.). These differences may be attributed to the fact that the PRMS results

supplied by NASA are still uncalibrated – and therefore unrealistic -- and that the SCA time series from satellite data are misleading for reasons to be discussed below. Some special cases include ones where the predicted SCA from PRMS never reaches zero for the entire calendar year (Figures A-44, A-55, etc.).



a) HRU 5 (Area = 175 km^2), 1985



b) HRU 19 (Area = 3950 km^2), 1982

Figure 5-2. Sample comparative figure showing SCA time series from both sources

5.3.1 Statistical Comparison of the Two Time Series

It is clear by visually inspecting Figure 5-2 and the Figures in Appendix A that the two methods of estimating SCA do not agree well, except during extreme seasons of winter and summer. During the snow depletion and accumulation seasons, when remote sensing data would be most valuable to understand hydrologic processes and calibrate/validate models, the satellite data do not appear very useful (this observation is discussed further in Section 5.3.1). Nevertheless, this thesis proceeds to apply quantitative analysis to these time series in anticipation of a future time when improved remote sensing products are available. Although neither the satellite analysis nor the model output are perfected, this section presents statistical tools that can be used to compare satellite and model SCA analyses in future work, and discusses the meaning of these statistical measures based on the preliminary analysis developed in this project.

The ultimate goal of the research, of which this study is a part, is to use satellite imagery as observational truth against which to test the performance of spatially distributed models. The PRMS model computes both spatially-distributed snow water equivalent (SWE) and HRU snow-covered area (SCA) as output variables. Currently, as discussed in Chapter 2, SCA is more readily available than SWE from space-based remote sensing. Although problems have been identified with the specific satellite product analyzed here, the study proceeds to explore possible statistical tools that could be used to evaluate model SCA output by comparing it to observed (satellite) SCA.

Ninety-Five SCA time series from the PRMS model were compared to the corresponding satellite-derived SCA time series, using a variety of statistical measures to determine their agreement. The first of these measures is the linear correlation

coefficient, R. Computing the coefficient of correlation (**R**) for the two time series for any full calendar year may be a misleading measure because of the perfect match in the two series during extreme winter months (SCA = 1.00) and during most of summer (SCA = 0.00). Hence it was decided that **R** be computed only for the snowmelt season to quantitatively assess the mismatch in the two series. Snow accumulation months were omitted in this analysis because of the high phase difference.

A FORTRAN program (McCuen, 2003) was used to analyze the time series. This program was applied to the snow depletion period of the two time series for each HRU for the years 1982-1985. Snow depletion data for 1981 are not available because NSIDC has remotely sensed data for the *Surface Type Mask* category for this region only for days after 24 July 1981 (Appendix D). The program was applied to SCA time series from both sources, for each HRU, during the time period of May 1 to July 31 for the years with available SCA data for these days. Because the research team eventually intends to use satellite imagery as truth against which to test model results. The satellite-derived SCA time series are treated as "observations" in this analysis.

In addition to the correlation coefficient, R, and the coefficient of determination R^2, the program computes the relative Standard Error ratio (Se/Sy) in the two series. Se/Sy is an indicator of the random variation or relative difference in values of the two time series. The FORTRAN program also provides the *correlogram*, which displays how the coefficient of cross-correlation (**R**) changes by changing the lag duration between the two series. Figure 5-3 displays a sample correlogram output from the program for the snow depletion period of the example time series shown in Figures 5-1b and 5-2b i.e., HRU 19 for the calendar year 1982. The curve represents values of the coefficient of

cross-correlation (**R**) for one-day time lags between the PRMS and satellite data time series. The dashed horizontal line at the center denotes the zero-lag instance for the two SCA time series i.e., in such a case the plot of the two time series would be as shown in Figure 5-2b. The values on the left are the **R** values shown by the curve for that day, and the values on the right indicate the lag duration in days (41 in this case, equal to ± 20 days and a zero-lag day).

The PRMS time series (say, x1) leads the satellite data time series (say, x2) in the cross-correlation analysis described by the correlograms. Therefore, a sample value of R for a lag of (say) *m* days on day *i* would be computed by the following equation:

$$R(\text{lag } m) = \frac{E[\, x1(i) - x2(i-m) \,]}{s_{x1}\ s_{x2}}$$

where s_{x1} and s_{x2} are the standard deviations of the variables, and E[] is the expectation operator. A positive lag *m* means that model output are being compared to satellite data reported *m* days earlier, and a negative lag *m* means that model output are being compared to satellite data reported *m* days later. For the example shown in Figures 5-2b and 5-3, the R values are highest (about 0.9575) for a lag of m = 9 days (value 11 on the right column of the correlogram. This indicates that the model time series agrees best with the satellite time series reported 9 days earlier, and that the two series may match better statistically if the PRMS simulated snowmelt values shift backward in time by 9 days.

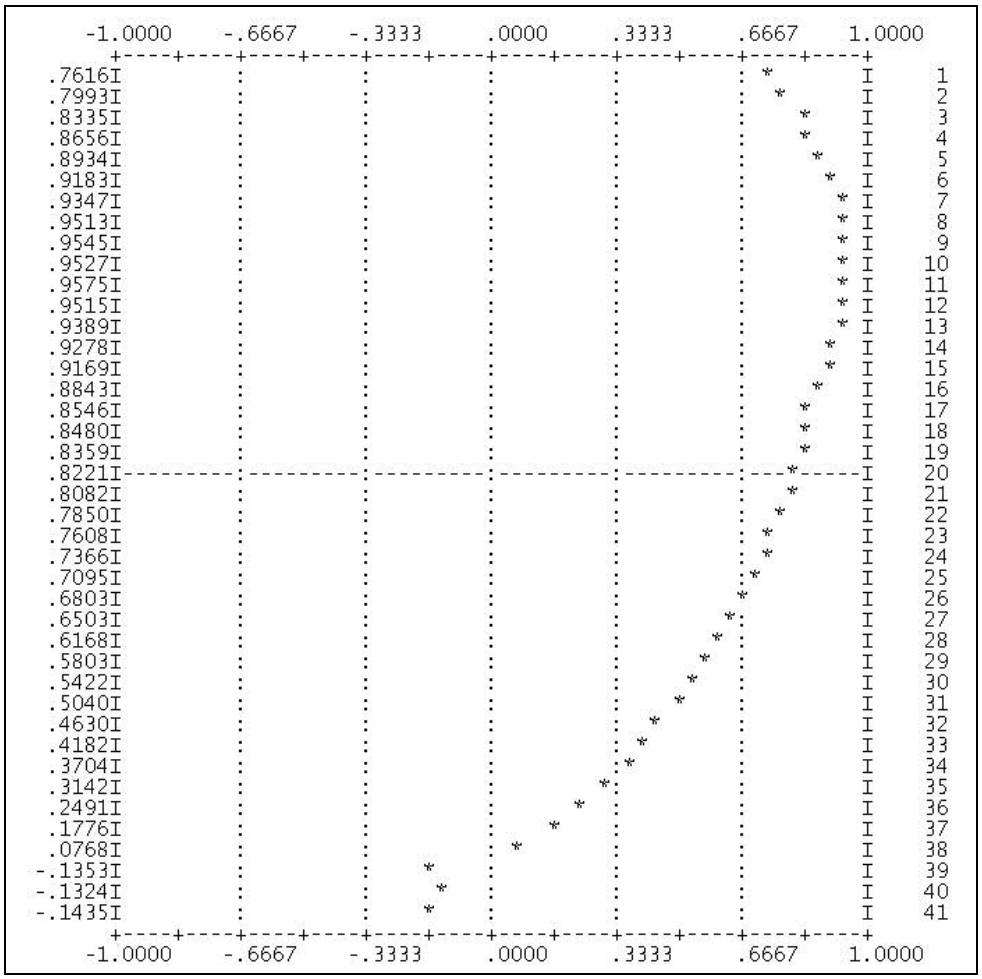


Figure 5-3. Sample Correlogram with 20 days of lag for HRU 19, calendar year 1982.

Correlograms thus show promise as tools to fine tune the modeled time series to match the satellite observations , with respect to errors in timing of snow disappearance. All 76 correlograms, for the snowmelt period of 19 HRUs for four years are displayed in Appendix B. They show results to a maximum lag of ±20 days. Most figures in Appendix B indicate that lag does not play a very significant role in determining the accuracy of these two snow depletion time series (Figures B-24, B-34, etc). On the other hand, some figures (B-57, B-69, B-73, etc) indicate a dramatic change in cross-correlation values on changing the lag-duration. Most of the latter kind are in HRUs in the lower elevations and

the relatively flat regions. Many of these correlograms have negative values for R at various lag-days, which indicates a negative correlation between the snow depletion values of the two time series. This largely occurs in instances where one time series indicates complete snowmelt and an SCA of 0% while the other still maintains an SCA of 100%. However, negative R values are not rational because SCA is not expected to decrease in one time series while increasing in the other simultaneously. Small negative values of R can be attributed as random variation.

Table 5-1 in Section 5.3.2 displays the values of R and R^2 for the snow depletion season for each of the 19 HRUs for the years 1982 to 1985. These R values are not the ones given by the correlogram analysis at zero-lag. Instead, they are computed using MS EXCEL with an even longer SCA depletion season, starting at Spring Equinox (March 21) and ending on July 31. As can be observed from this table, the R values do not appear to follow any increasing or decreasing trend with time, which is expected because each R value is derived from two SCA time series samples whose accuracy is not known to be changing with time. Figure 5-4 shows that R values for this extended snowmelt season are positive and have a very wide range for the HRUs. Many HRUs show high R values (0.7 to 0.9) for the years 1982 and 1983, and somewhat lower (<0.7) for the year 1985. Another observation is that while some HRUs show a distinct “low-high-low-high” pattern over the four years, some others remain constantly high or constantly low. HRU 19, which represents croplands in the most downstream watershed has a distinctly high R value (≈ 0.9) between the two SCA depletion time series for the years 1982, 1983, and 1984, and a sharp drop to 0.55 for the year 1985.

Since satellite SCA data was missing for some days, these "gaps" in the time series were filled up by interpolating values near the voids in the time series. This step was taken only for the statistical analysis of the snow depletion time series by the FORTRAN program and by MS EXCEL. These voids were not filled when generating figures displayed in Appendix A. In all, 8 days (2.17%) of satellite SCA data are missing from a total of 368 days for the snowmelt season time series selected for the analysis by FORTRAN, and 25 days (4.69%) of satellite SCA data are missing out of 532 days for the analysis by MS EXCEL. Overall, 110 days (6.82%) of data are missing for the entire dataset out of a total of 1611 days (Appendix D).

Figure 5-5 shows the values of R^2 in the same format as Figure 5-4 showed R values. In this case, R^2, represents the percentage of variance in the observed (satellite) time series that is explained or captured by the model time series. Figure 5-5 shows that R^2 values for the various HRUs have a very wide range (0.1 to 0.8) for the years 1982, 1983 and 1984. They are lower year 1984, and very low (<0.5) for the year 1985, indicating that in those years the PRMS model is not capturing the variance in the satellite SCA time series very well.

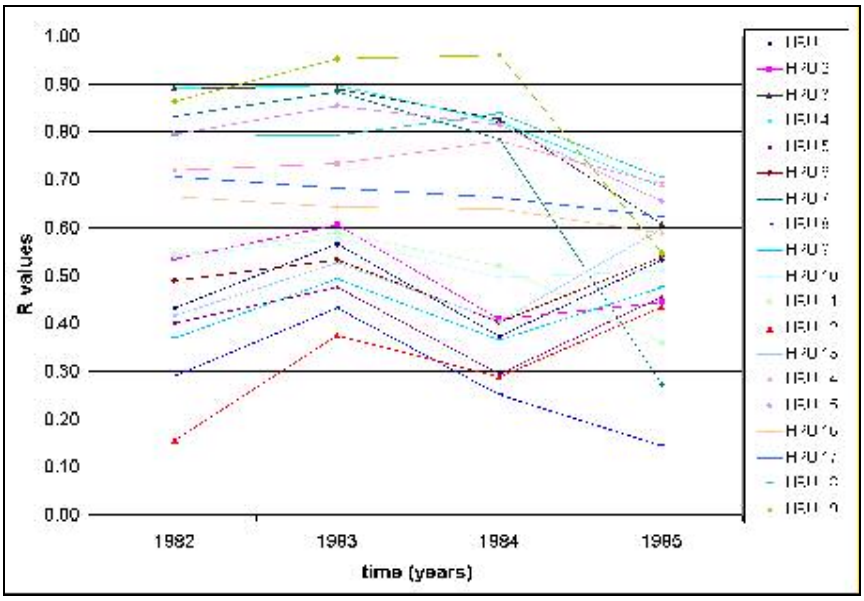


Figure 5-4. Cross-correlation coefficient (R) values of the two SCA time series for snow depletion period of March 21 to July 31, for each HRU.

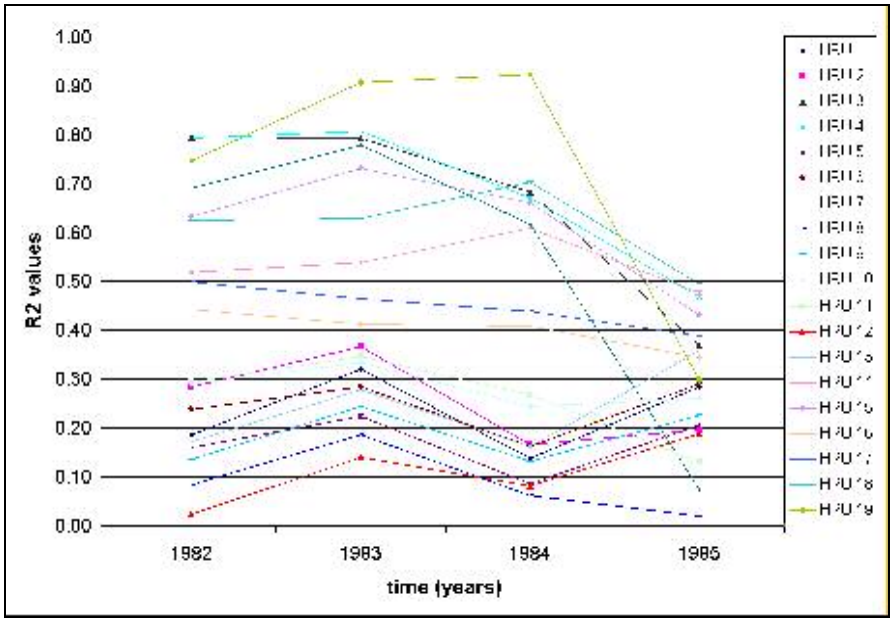


Figure 5-5. R^2 values of the two SCA time series for the snow depletion period of March 21 to July 31, for each HRU.

Lastly, Figure 5-6 shows the distribution of Se/Sy values obtained as a result of the analysis on the two SCA time series using the FORTRAN program (McCuen, 2003).

This is based on the snowmelt period selected for the correlogram analysis, i.e., May 1 to July 31. Se/Sy is often computed as $(1 - R^2)^{1/2}$. As observed in Figure 5-6, Se/Sy is greater than 0.9 in most cases indicating a high relative mismatch and thus poor similarity between the two time series. The lowest Se/Sy values are for HRU 19 (≈0.6, indicating 60% standard error ratio).

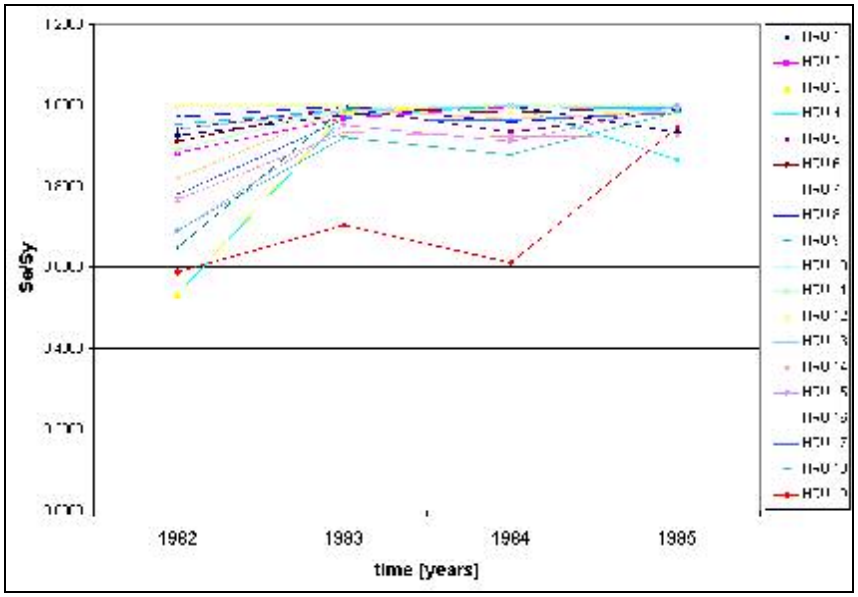


Figure 5-6. Se/Sy values for the two SCA time series for the snow depletion period May 1 to July 31, for each HRU.

5.3.2 The Nash-Sutcliffe Coefficient

The two SCA time series involved in this research (from satellite data and PRMS output) have strong serial correlation among the data values because they represent a serial depletion in SCA. Menoes (2003) found that the serial correlation in SCA time series was so great that the entire time series represented only a single statistical degree of freedom. In such time series, which have a serial correlation, it is not strictly correct to

apply the correlation coefficient, R, and the coefficient of determination, R^2 (Ward, 1992). A useful goodness of fit criterion for such continuous time series is the Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970), computed as follows:

$$R^2 = 1 - \frac{\Sigma (Q_i - Q_i')^2}{\Sigma (Q_i - Q)^2}$$

where Q_i = the measured or observed time series i.e., the satellite data SCA time series in this case, Q_i' = the computed time series or the PRMS output in this case, and Q = the average value of the observed (satellite SCA time series. Nash-Sutcliffe coefficients are computed for the extended SCA depletion period of March 21 to July 31 using MS EXCEL, and are displayed with corresponding R (regression coefficient) and R^2 (coefficient of determination) values in Table 5-1.

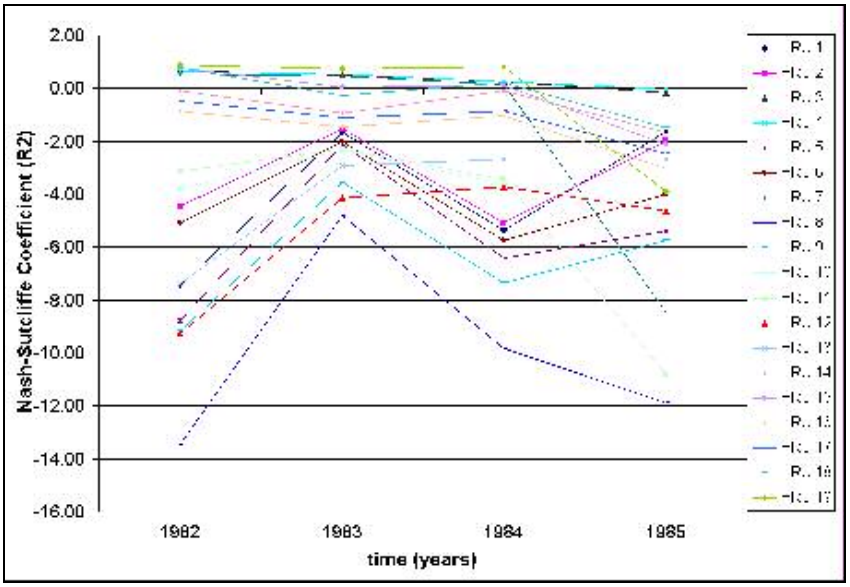


Figure 5-7. Nash-Sutcliffe coefficients of the two SCA time series for snow depletion period of March 21 to July 31, for each HRU.

The Nash-Sutcliffe coefficient values (also denoted by R^2) should usually vary from 0 to 1, with 1 indicating a perfect fit. Computationally, these coefficients could be negative

but that becomes rather meaningless as far as interpretation or results are concerned. Figure 5-7 shows the Nash-Sutcliffe coefficients for the SCA depletion period from March 21 to July 31. It is observed that only six of the nineteen HRUs have a positive R^2 value for the years 1982, 1983 and 1984, and all have a negative value for the year 1985. This indicates a reasonably imperfect match of the two SCA depletion time series, which was also observed by earlier criteria of R, R^2 and Se/Sy. HRU 19 stands out again, with high Nash-Sutcliffe values (> 0.75) for the years 1982 to 1984. Table 5-1 indicates that only few other values exceed 0.5.

Table 5-1. Values of Coefficient of Correlation (R), R^2 and Nash-Sutcliffe Coefficient for the two SCA depletion time series for the period March 21 to July 31.

	Coefficient of Correlation (R)				**Coefficient of Determination (R^2)**				**Nash-Sutcliffe Coefficient (R^2)**			
	1982	**1983**	**1984**	**1985**	**1982**	**1983**	**1984**	**1985**	**1982**	**1983**	**1984**	**1985**
HRU 1	0.43	0.57	0.37	0.53	0.19	0.32	0.14	0.28	-7.46	-1.65	-5.37	-1.62
HRU 2	0.53	0.61	0.41	0.44	0.28	0.37	0.17	0.20	-4.44	-1.51	-5.07	-1.99
HRU 3	0.89	0.89	0.83	0.61	0.80	0.79	0.68	0.37	0.66	0.50	0.27	-0.15
HRU 4	0.89	0.90	0.82	0.68	0.79	0.81	0.67	0.47	0.65	0.53	0.25	-0.01
HRU 5	0.40	0.47	0.29	0.46	0.16	0.23	0.09	0.21	-8.78	-2.13	-6.41	-5.38
HRU 6	0.49	0.53	0.40	0.54	0.24	0.28	0.16	0.29	-5.10	-2.03	-5.73	-3.99
HRU 7	0.83	0.88	0.79	0.27	0.69	0.78	0.62	0.07	0.55	0.48	0.13	-8.47
HRU 8	0.29	0.43	0.25	0.14	0.08	0.19	0.06	0.02	-13.44	-4.81	-9.81	-11.87
HRU 9	0.37	0.49	0.36	0.48	0.14	0.24	0.13	0.23	-9.15	-3.56	-7.35	-5.74
HRU 10	0.54	0.58	0.50	0.51	0.29	0.33	0.25	0.26	-3.81	-2.33	-3.62	-4.61
HRU 11	0.54	0.59	0.52	0.36	0.29	0.35	0.27	0.13	-3.12	-2.28	-3.43	-10.80
HRU 12	0.16	0.37	0.29	0.43	0.02	0.14	0.08	0.19	-9.25	-4.14	-3.72	-4.62
HRU 13	0.42	0.53	0.40	0.60	0.17	0.28	0.16	0.36	-7.38	-2.91	-2.66	-2.67
HRU 14	0.72	0.73	0.78	0.69	0.52	0.54	0.61	0.48	-0.10	-0.92	-0.09	-1.66
HRU 15	0.80	0.85	0.81	0.66	0.63	0.73	0.66	0.43	0.77	0.05	0.10	-2.12
HRU 16	0.67	0.64	0.64	0.59	0.44	0.41	0.41	0.34	-0.86	-1.44	-1.03	-3.01
HRU 17	0.71	0.68	0.66	0.62	0.50	0.46	0.44	0.39	-0.48	-1.11	-0.86	-2.48
HRU 18	0.79	0.79	0.84	0.70	0.63	0.63	0.70	0.50	0.78	-0.28	0.23	-1.49
HRU 19	0.86	0.95	0.96	0.55	0.75	0.91	0.92	0.30	0.88	0.75	0.83	-3.89

The results displayed in Table 5-1 and Figures 5-4 to 5-7 clearly indicate that the time series generated by the NASA teams PRMS model do not match closely with the remotely sensed satellite data SCA time series . This calls for an investigation of the major causes that contributed to this mismatch. Some of the suggested crucial reasons are discussed in Section 5.4.

5.4 THEORIES FOR THE INTERPRETATION OF RESULTS

With reliable and believable satellite observations, any of the quantitative measures applied in Section 5.3 could be used as calibration targets to improve the model results. Model parameters could be adjusted objectively or subjectively in order to obtain a better match to the remote sensing observations. An example of the use of satellite SCA time series as calibration/validation targets is Menoes' (2003) work with National Operational Hydrologic Remote Sensing Center (NOHRSC) snow cover data with a process based snow accumulation and ablation model.

As observed in Sections 5.1 and 5.2, however, the satellite SCA time series does not appear to be a good and reliable representation of the actual snow depletion or accumulation in the Tom R. basin, based on physical reasoning. Due to spatial variability in precipitation, stored snow mass, and melt energy, snow cover would not be expected to change from 0 to 100%, or vice versa, in a matter of days (Section 2.5). There may be multiple reasons for this occurrence. This section will focus on the supposedly most dominant reason, and also briefly explain the possible contribution of other causes for the inaccuracy. Three possible explanations for the sudden disappearance of snowcover in the SSM/I Surface Type Mask satellite time series are explored:

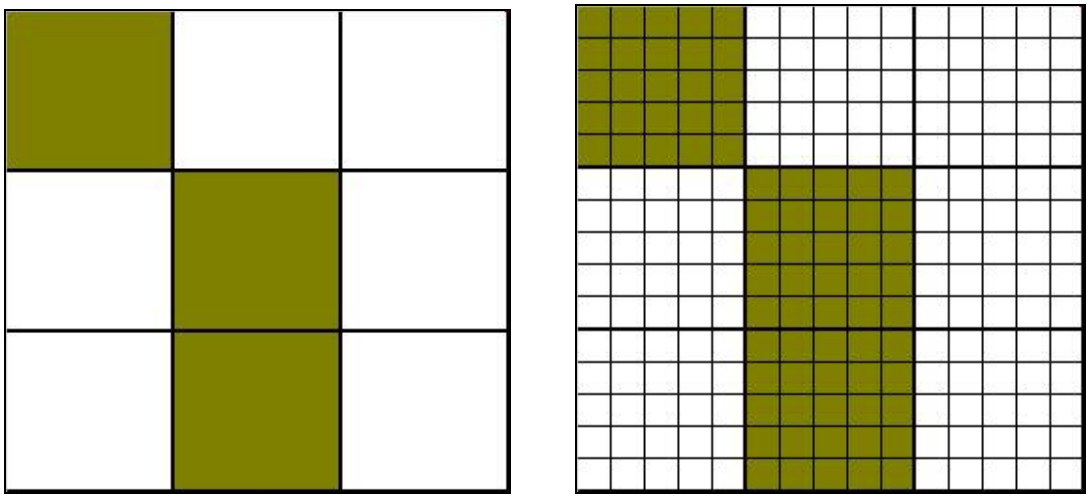
1. The effects of coarse resolution
2. Sensor properties and Snow Identification Algorithm
3. Possible physical explanations

5.4.1 NSIDC Snow Data Resolution

The *Surface Type Mask* data ordered and downloaded from NSIDC for analysis in this project are labeled as "AVHRR Polar Pathfinder Twice-Daily 5 km EASE-Grid Composites" data at NSIDC. This dataset was selected for the project because the 5-km resolution appeared to be compatible with the 1-km land use and DEM. However, after downloading and displaying the product, we learned that this 5-km resolution Surface Type Mask data has been resampled from data initially recorded by SSM/I sensors at a resolution of 25-km. NSIDC has made corrections in the resampled 5-km data at the coastlines to reduce the jaggedness caused by the 25-km resolution. However, they have not altered the resampled 5-km data at regions farther away from the coast. This downscaling of pixel size from 25 km to 5 km cannot provide information finer than the 25-km scale. Figure 5-8 demonstrates an example of resampling a 25-km grid to a new 5-km grid, similar to the process experienced in this project. Despite breaking a single 25-km x 25-km pixel into twenty-five 5-km x 5-km pixels, the values allotted to the new 25 pixels taking up the place of the earlier single large pixel remain the same. This does not improve the accuracy of the image compared to the true condition on the ground.

1-km resolution DEM and GLCF Landcover data were used in the creation of the HRU masks. It is true that *map multiplying* 5-km snow data layers with the 1-km HRU masks in Section 4.4.2 gives more finely-resolved results than would be obtained if 25-

km snow data were map multiplied with the 1-km HRU masks. The latter case would result in 25-km resolution pixels for the final SCA output, and the entire Tom R. watershed would be covered by much fewer (approx. 25 times fewer) pixels than the 5-km pixels and roughly 125 times fewer pixels than 1-km pixels. (The conversion factor may not be exactly 25 or 125 respectively because of the irregular shape of the Tom R. watershed.)



a) Sample 25-km grid b) Resampled 5-km grid

Figure 5-8. Example of resampling a coarser resolution grid into a finer resolution grid.

The resampled 5-km resolution snow data provided in the "AVHRR Polar Pathfinder Twice-Daily 5 km EASE Grid Composites" dataset did not prove more beneficial or accurate than the 25-km dataset in this region. An original *satellite sensor resolution* of 5-km would provide more accurate results than *image resolution* of 5 km (resampled from original 25-km resolution data).

5.4.2 Simulation to Demonstrate Effect of Resolution on SCA

The sudden drop in satellite generated SCA values as observed in Figure 5-1 appears physically unreasonable. Hydrologically, a one-day or several-day disappearance of snow cover from a large region implies that all of the snow depth or snow water equivalent (SWE) throughout that region disappears at exactly, or very nearly, the same time. Whereas, we know that both snow depth and melt rate are highly variable even over short distances. A snow pack would be expected to melt away and disappear in a patch or zones according to elevation, orientation, and meteorological factors. It is true that we do not have the benefit of direct observation of snow in the headwaters of the Ob; however, it is difficult to imagine that these regions would behave differently than seasonally snow covered regions elsewhere. This section aims at demonstrating the effect of grid resolution in the interpretation of actual ground data (SCA in the case of this research).

The idea is to simulate a sample snow cover depletion process and analyze how that depletion process is captured by images of the same *state of truth* with different resolutions. A totally random distribution of 'snow' (value = 1) and 'no snow' (value = 0) pixels over the 1-km grid is not very realistic because in real life snow falls and melts over an area with some spatial correlation. That is to say, if a pixel is snow covered, then the probability of pixels around it being snow covered is high, compared to pixels that are not snow covered.

A truly accurate model for the distribution of snow as described above would have to consider various parameters that affect the spatially variable accumulation and depletion of snow. Such parameters include elevation, slope, wind direction, solar

radiation, etc., for every point in the study region. This would require the creation of a spatially distributed model in itself, with each individual pixel as an independent unit. A properly specified and parameterized model driven with spatially distributed meteorological variables would be able to produce a spatially variable model snowpack with an appropriate spatial correlation structure. Due to the fact that creation of such a complex model is not the primary objective of this research, and due to time and data constraints, simulation of snow depletion in this section is performed by a simple model that, however, makes an effort to capture some logic of spatial correlation in snow area.

The simulation experiment is set up over an assumed area of 125 x 125 pixels of (say) 1-km x 1-km resolution. This square shaped 1-km "simulated satellite image" will then be resampled to a 5-km x 5-km grid, thereby reducing the number of pixels from 15,625 to just 625 (25 times reduction in number of pixels). The 1-km x 1-km grid is also resampled into 25-km x 25-km pixels. This results in twenty-five 25-km resolution pixels in the third grid. All three grids represent the same area over the same simulated SCA on that square region.

The proposed model uses a Markov Chain to simulate snowcover over the 1-km resolution image. Table 5-2 summarizes the application of the Markov Chain in the model. The Markov Chain model is typically used in the time domain to describe transitions from one system state to the next in discrete time steps and long-term (steady-state) probabilities of the system being in a given state at a point in time. Here the Markov Chain model is applied in space to describe transitions from one location to its neighbor and the overall spatial fraction of pixels expected to be found in each state.

Let state 0 = ‘no snow’ and state 1 = ‘snow’. Therefore, if some pixel A is snow covered (state = 1) then the probability that the pixel following it (in a pre-determined direction) is also snow covered (state = 1) is equal to P_{11}, which is known as the “Transition Probability” in Markov Chain theory. Similarly for P_{00}, P_{01} and P_{10}. The one-dimensional Markov Chain model is applied in a spiral pattern in an attempt to improve spatial structure on the simulated snow cover. The starting point is the center pixel of the 9 x 9 pixel square in the upper left of the 126 row x 126 column domain. It is initialized randomly based on a Bernoulli trial with P(success) equal to the analytical probability of snow or predicted snow covered area for the domain. The control then moves out in a spiral motion from the center of the first cell covering all 81 pixels in the first 9 x 9 square. Figure 5-9 shows this spiral flow of control. The snow state of each pixel is dictated by the transition probabilities and a Bernoulli probability random number generation in Microsoft EXCEL. Control proceeds from the final pixel in the first square to the center of the adjacent 9 x 9 pixel square and continuing in a single spatial chain to assign values to all pixels in the domain. The values in the 126^{th} row and 126^{th} column are ignored, using all values in a 125 x 125 pixel grid.

For a given day, a snow cover field or image is produced by specifying the theoretical percent snow cover for that day, P_1, and a transition probability P_{11} (specifying the likelihood that the neighbor of a snow-covered pixel will also be snow-covered). With these two parameters fixed, the other transition probabilities are dictated by the mathematics of the Markov Chain theory. By the axioms of probability,

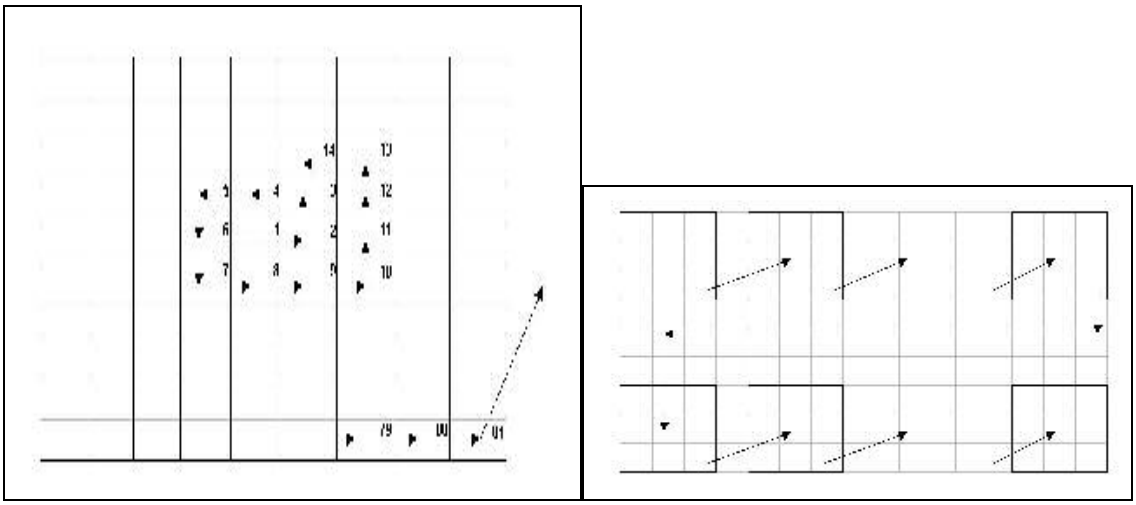
$$P_{00} + P_{0\,1} = 1,$$

$$P_{10} + P_{11} = 1\,,$$

and by the definition of a steady state (or, in this case, overall spatial state),

$$P_1 = P_0 \times P_{01} + P_1 \times P_{11},$$

$$P_0 = P_0 \times P_{00} + P_1 \times P_{10}.$$



a) Control flow within a 9x9 square b) control flow from the last pixel of one to the center of the next 9x9 square

Figure 5-9. Spiral control flow from the center of a 9 x 9 square.

Table 5-2. Markov Chains and model logic.

Pixel B ->	**0**	**1**		**Pixel B ->**	**0**	**1**
Pixel A				**Pixel A**		
0	$P_{0\,0}$	$P_{0\,1}$		**0**	s	1-s
1	$P_{1\,0}$	$P_{1\,1}$		**1**	n	1-n

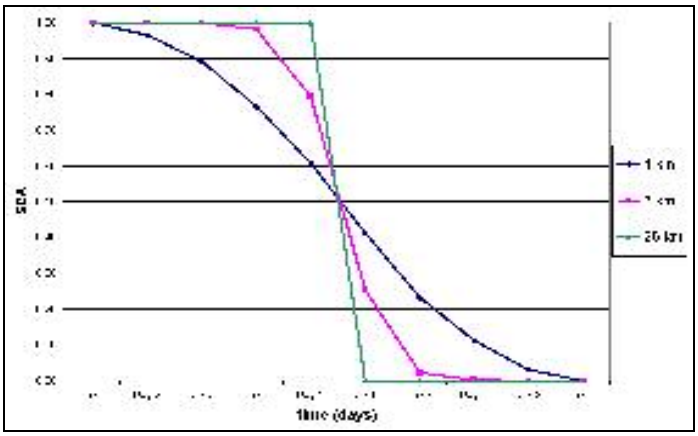
A single realization of an experiment consists of creating a sequence of ten snow cover fields with decreasing fraction of snow cover, representing sequential dates in the snow depletion season. Unlike physical reality, the patterns of snow cover are independent of each other i.e., there is no mathematical relationship between a particular pixel's snow state on one date and the next. The fractional snow cover is serially

correlated by design.. Eight separate experiments were conducted with this mechanism. Each experiment consists of ten realizations using the same parameter set (four transition probabilities), which are averaged to reduce sampling variation. A higher number of realizations with more random samples would reduce the sampling variation even further. However, the initial variation in predicted SCA values for each day in multiple realizations is not very high (< 5%); therefore, it is believed that ten simulations for each day for each experiment serve to demonstrate the effect. The results of each simulation for these eight experiments are displayed in Appendix C.

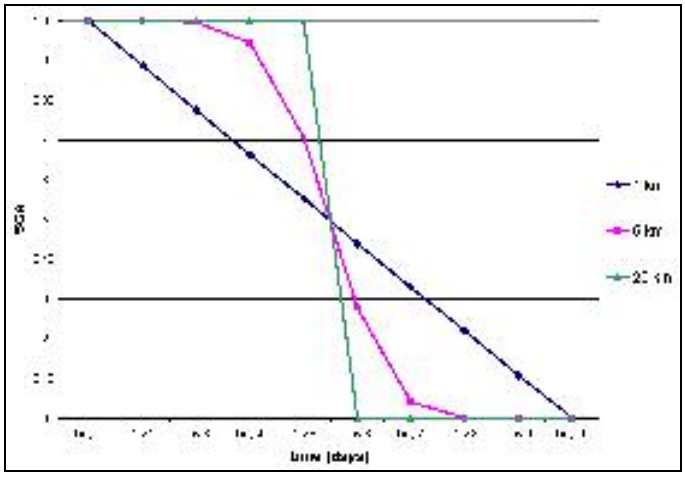
Figure 5-10 displays a simulated snow depletion time series over ten days as captured by grids with resolutions of 1-km, 5-km, and 25-km. A random, but clustered, snow cover is generated in the 1-km grid as explained earlier. The value in each pixel of the 5-km grid is determined by the twenty-five 1-km pixels it directly replaces. If more than 50% of the 1-km pixels have a state = 1 then the resulting 5-km pixel will also have state = 1, otherwise the larger pixel will have state = 0. Similar technique is used to determine values of each of the 25-km pixels. Therefore, all three images at any date represent the same simulated “truth”. Figures 5-10a and 5-10b show the snow depletion process when the simulated snow cover follows a sigmoid (S-shaped) and a linear decrease, respectively. Figure 5-11 illustrates a single realization of simulated images of the sigmoidally depleting SCA (Figure 5-10a), as shown by the three resolutions. The individual probabilities and SCA values for the ten realizations represented by the curves in Figure 5-10a are displayed in Experiment 1 of Appendix C.

Figures 5-10 and 5-11, and the tables and figures in Appendix C show the effect of increasing pixel size. They demonstrate how a gradual snow depletion curve at a

spatial resolution of 1-km becomes somewhat inaccurate at the 5-km resolution and very unrealistic at 25-km resolution, where the SCA drops off from 100% to 0% in a single time unit. This is the most likely reason for the presumed inaccuracy of the satellite snowcover data. Conclusions on the effect of this resolution-caused inaccuracies are discussed in Chapter 6. An interesting feature of Fig. 5-10(a, b) is that the curves for all three resolutions cross at SCA = 0.5; in the following section, this observation is used to suggest a new measure of agreement between satellite and modeled SCA time series.



a) SCA decreases sigmoidally over ten days.



b) SCA decreases linearly over ten days.

Figure 5-10. Snow depletion curves as they appear in grids of different resolutions for the same snowmelt process.

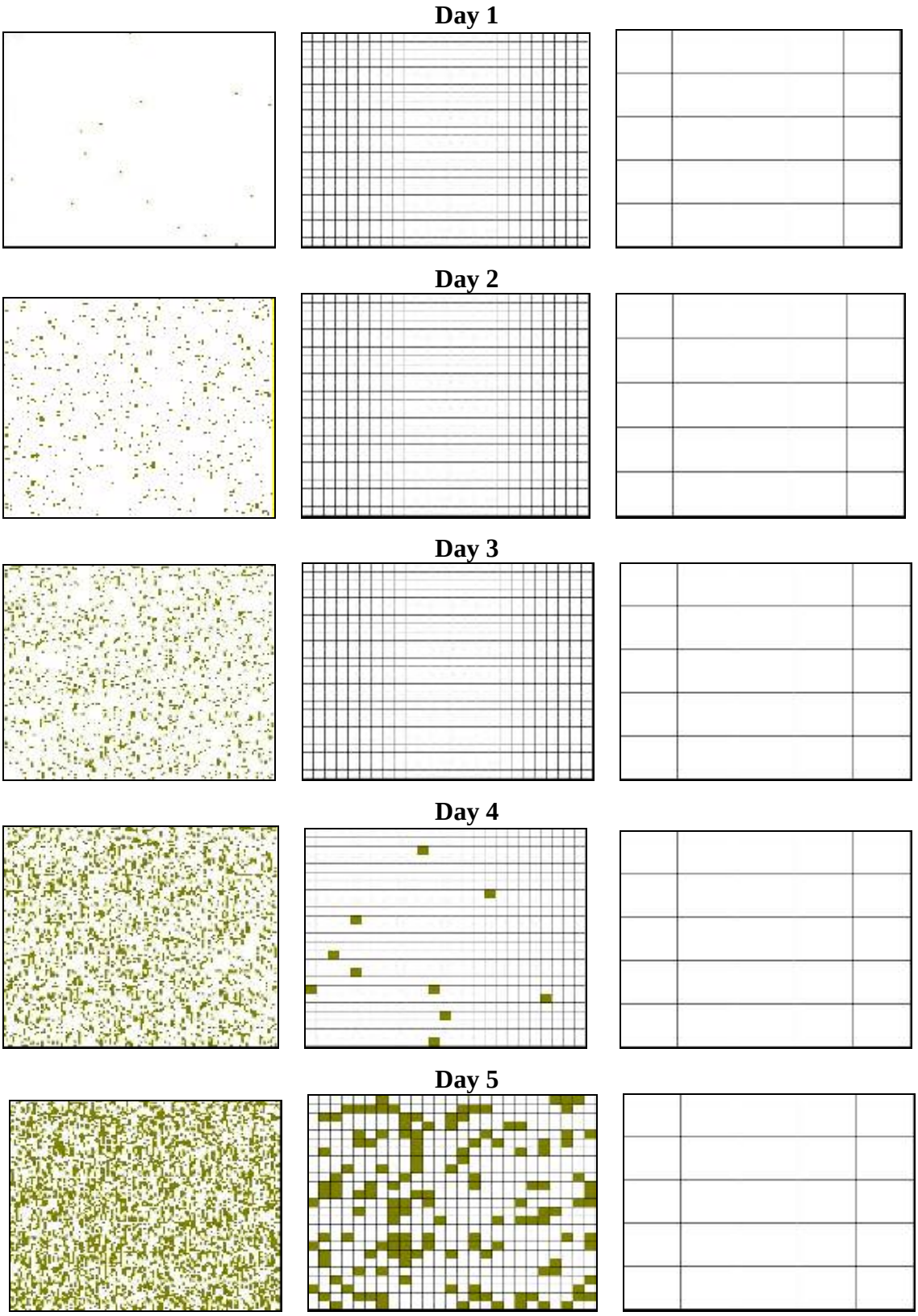


Figure 5-11. Sigmoidally depleting SCA (continued on next page)

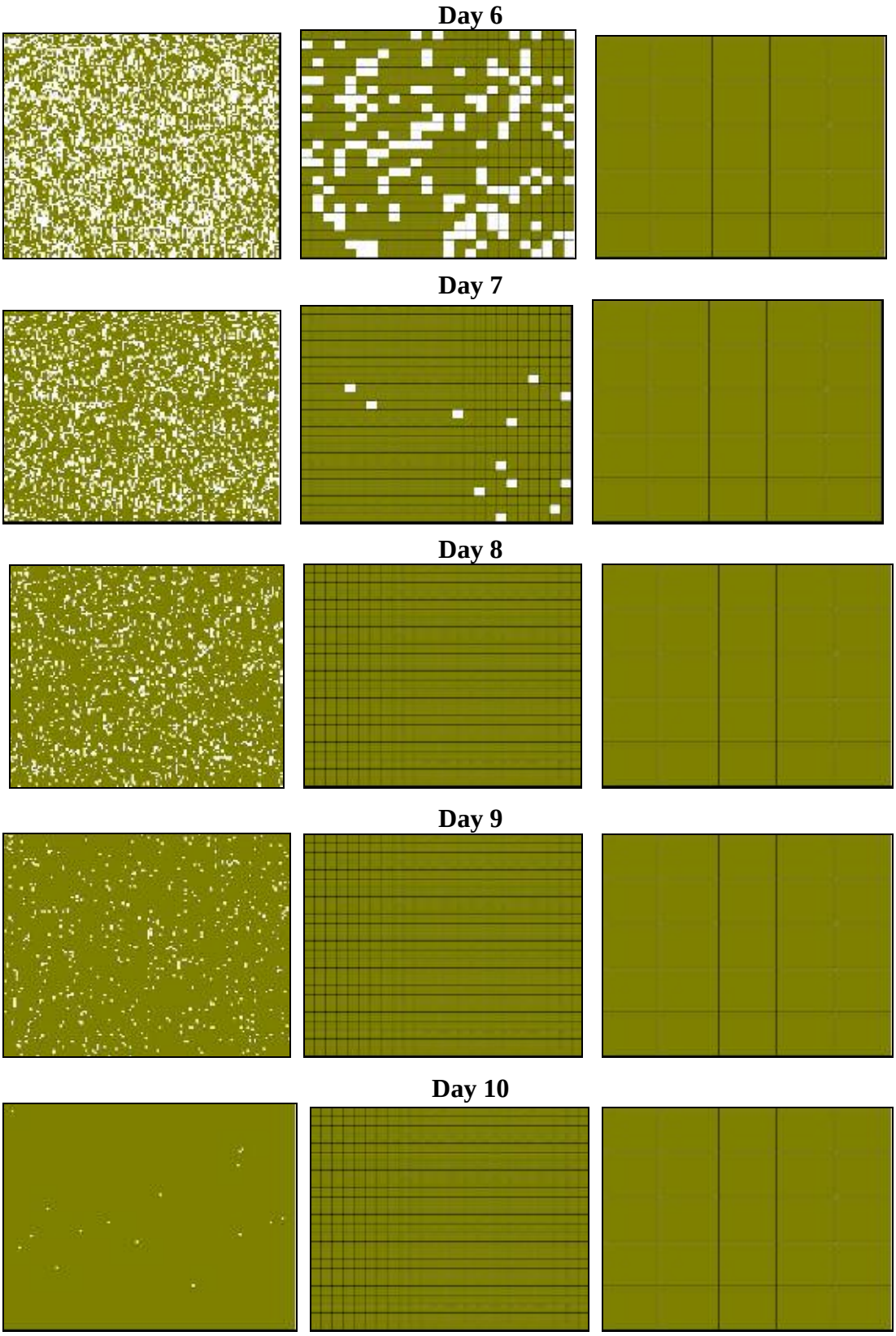


Figure 5-11 (continued). Sigmoidally depleting SCA

5.4.3 T_{50} Analysis

As discussed in Section 5.3.2, the SCA time series from the satellite data as well as the model have strong serial correlation since they represent a continuous depletion of SCA from 100% to 0%. Since they may be so strongly correlated that the entire time series is described by one statistical degree of freedom (Menoes, 2003), a simple measure may be a better indicator of how well the two time series match than the coefficients of correlation and determination.

It is observed in Section 5.4.2 that despite different resolutions of the satellite imagery capturing the same SCA depletion process, the SCA depletion curves meet at the same point for SCA = 0.5 (Figure 5.10a, b). A new variable, $T_{50,}$ is defined to be the day of the year when SCA drops to 50%, in the respective SCA depletion time series. If grid resolution is the main factor contributing to the mismatch of the two time series, T50 may allow even these coarsely resolved satellite images to be useful measure in evaluating model performance.

If T_{50}(model) - T_{50}(satellite data) = 0 then indeed coarse resolution is the main contributor to the imperfect match. However, if T_{50}(model) - T_{50}(satellite data) $\neq$ 0 then resolution is perhaps not the only issue and there are certainly other factors involved. Figure 5-12 shows a scatter diagram of T_{50}(model) versus T_{50}(satellite data). It is observed that all of the values lie more towards the PRMS model side from the 1:1 line, indicating that T_{50}(model) > T_{50}(satellite data). (This finding is consistent with the earlier observation from cross-correlation analysis that the model SCA depletion lags the satellite depletion in time.) It can, therefore, be inferred that other factors also contribute reasonably to the imperfect match of the two time series. The fact that PRMS model

being used by the NASA HSB team is still not fully calibrated and the output is in its preliminary stages should also be kept in mind.

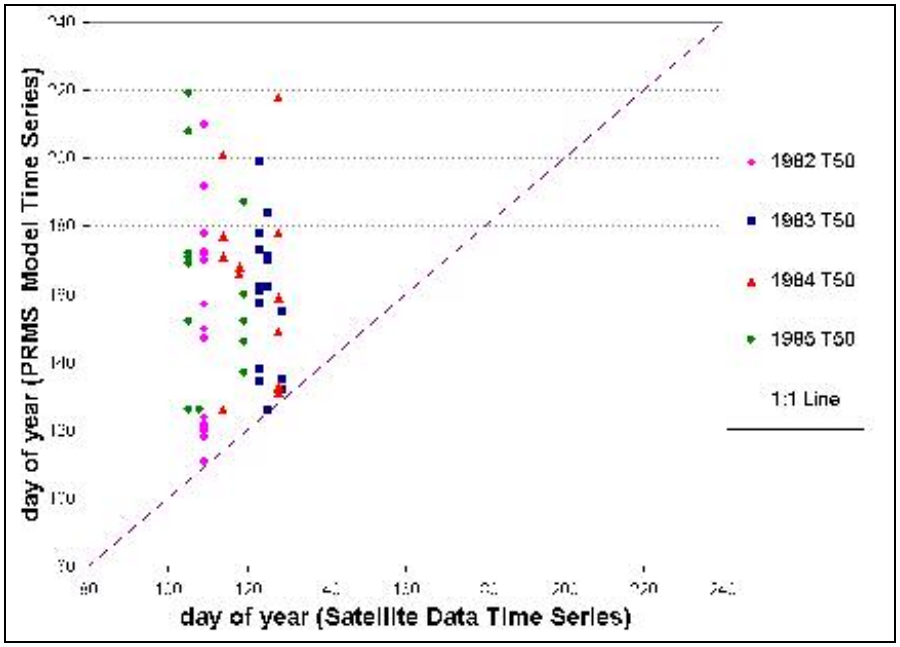


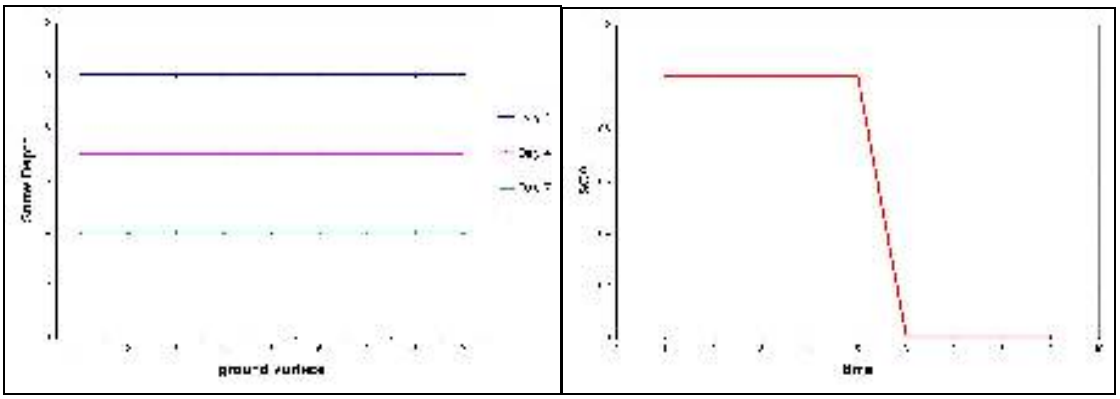
Figure 5-12. T_{50} analysis diagram.

5.4.4 Satellite Snow Algorithm

The NSIDC Surface Type Mask data used in this research is interpreted by snow data detected by SSM/I sensors that record brightness temperatures in different frequency bands. The brightness of snow changes as it gets older, particularly during the melt season when the liquid water fraction of the snowpack is increasing, due to the different reflective properties of the phases of water across the electromagnetic spectrum. Therefore, it is quite likely that SSM/I sensor wrongfully detected a much lower signal for a relatively higher amount of snowcover in the event of snow lying undisturbed for some duration of time.

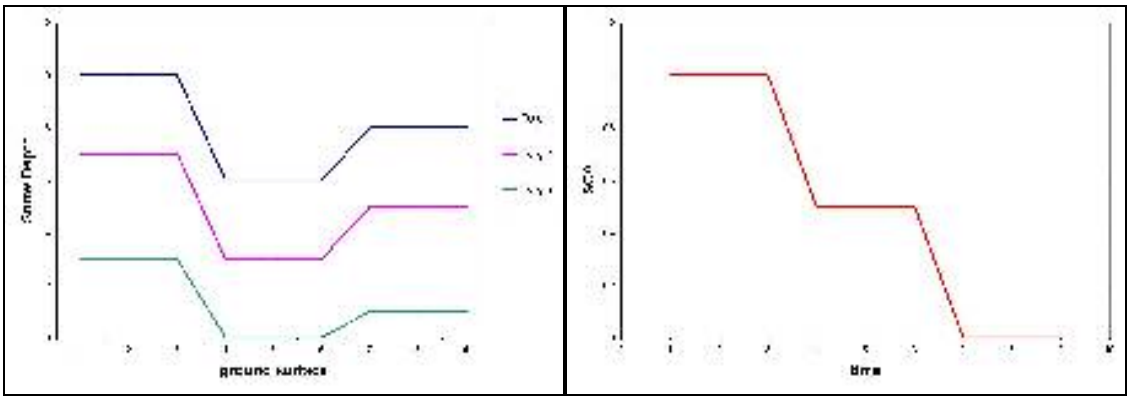
5.4.5 Possible Physical Explanation

Land is considered “snow covered” if it has even the tiniest bit of snow depth on it. For example, a snow depth of (say) 0.1 cm to 3 m would count as equally “snow covered.” Consider a region covered with a uniform snow depth and SWE. It starts to melt and melts away slowly at a uniform rate all across the region. Runoff and streamflow result, because water mass is being removed from the snow pack; but the region is 100% “snow covered” from the time melt begins until the time the snow is completely melted away. This idea is displayed in Figure 5-13. In such a case, the snow cover could drop from 100% to zero in one or two days. However, this situation is physically unlikely because snow depth, snow density, SWE and melt rate all vary spatially. Snow depth, SWE, melt rate and snow density are defined at points in space; however, snow covered area is a spatial fraction. SCA does not necessarily decline at the same rate as areally integrated SWE; there is a complex mathematical relationship between the two (Liston, 1999). The snow cover depletion curve is a result of spatial variability in SWE and melt rate across a study area or HRU. The more homogenous a study area, the more abrupt the SCA drop from 100 to 0 percent will be. Figures 5-14 and 5-15 demonstrate this idea and supplement Figure 5-13.



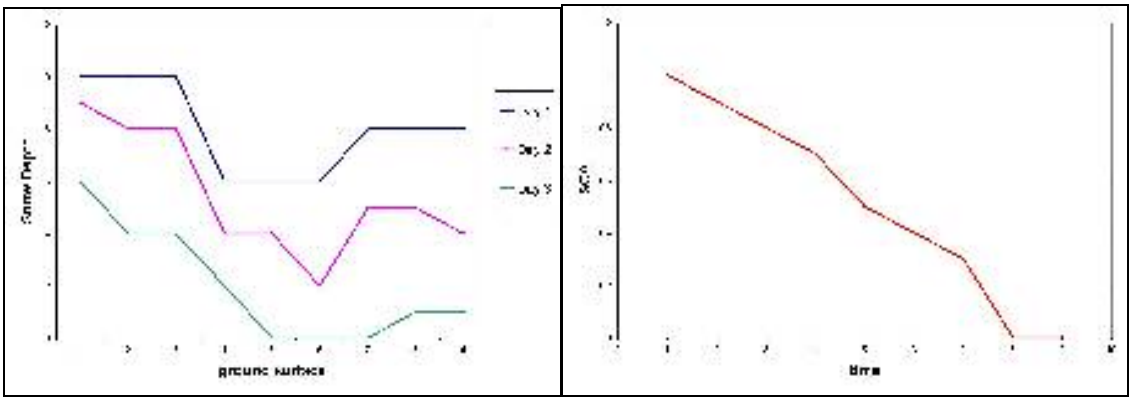
a) Snow depth decreasing with time b) Corresponding SCA depletion curve

Figure 5-13. Uniform snow depth and uniform melting. SCA drops abruptly.



a) Snow depth decreasing with time b) Corresponding SCA depletion curve

Figure 5-14. Non-uniform snow depth but uniform melting. Somewhat gradual SCA depletion curve expected.



a) Snow depth decreasing with time b) Corresponding SCA depletion curve

Figure 5-15. Non-uniform snow depth and non-uniform melting. Gradual SCA depletion curve expected.

5.4.6 Other Factors contributing to Mismatch of Results

A number of other factors could be responsible for the mismatch of the two time series, probably less than the coarse resolution. One evident reason is that the PRMS data supplied by the NASA HSB team is still in its preliminary stage and uncalibrated. Improved PRMS results may help eliminate some of the phase difference in the two series during snowmelt and snow accumulation. The NASA HSB team's model also shows sudden depletion of snow in certain cases (HRU 6, 1985, etc., Appendix A). Other reasons include possible human errors by the NASA HSB team in the delineation of the watershed and subareas using hand-drawn paper maps. Yet another possible source are the technical errors encountered during the recording, storage, and retrieval of snowcover data from the satellite sensors.

Overall, all of these possible factors have apparently contributed to the imperfect match of the two SCA time series. Conclusions derived from the results discussed in this chapter, and from the algorithms described in Chapter 4 will be detailed in Chapter 6.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 OVERVIEW

Chapter 5 discussed and analyzed in detail the SCA time series for the Tom river basin generated by using SSM/I "Surface Type Mask" data, and its comparison with preliminary PRMS model results. The aim of Chapter 6 is to derive conclusions from those results and analysis and suggest ideas for future research that may prove beneficial in applying Remote Sensing technology to hydrological applications, especially of the kind addressed in this research project: distributed modeling.

6.2 CONCLUSIONS ON SCA TIME SERIES AND IMAGE RESOLUTION

This research project achieved its objectives of:

a) Deriving SCA time series for the part of the Ob River basin using remotely sensed satellite data. Using a combination of digital elevation model data, and snow/no snow images, in a Geographic Information System, time series of percent snow cover were successfully generated for HRUs of the Tom River basin. These methods are general and could be applied to any region and any snow/no-snow remote sensing product.

b) Comparing the SCA time series from (a) with the results from a spatially distributed hydrological model. The time series of SCA from remote sensing and GIS analysis can be compared using statistical tools, specifically the correlation coefficient, the coefficient of determination, lagged cross-correlation, and the Nash-Sutcliffe coefficient. These statistics are most meaningful when applied to selected portions of

the time series (melt season and accumulation), rather than full-year analysis, in cases such as those studied here, where long periods of total or zero snow cover would distort the statistics. The time at which SCA drops to 50% was identified as a potentially useful comparison measure, which is not sensitive to the resolution of the snow/no-snow image.

The objective of using the satellite data time series for model calibration and validation was not achieved in this study. In fact, it is quite questionable whether this particular satellite product is appropriate for model calibration/validation at the 1-km resolution.

As discussed in Section 5.3.1 the major contributor to the imperfect match of the two SCA time series was the coarse resolution (5-km resolution resampled from 25-km) of the "Surface Type Mask" dataset at NSIDC. It may be concluded here that maybe this kind of research was not ideally suitable for the 1980s because of the kind of snow cover imagery available for those times. Satellite remote sensing techniques in the early 1980's were still in the nascent stages and image resolution for most civilian satellites was coarse (≈25km). This did not permit great accuracy of results derived from such satellite images and ground truth or simulated models, as demonstrated by the results of this study.

Without doubt, the resultant SCA time series would have been more accurate if the resolution was finer (say, even 5km x 5km). Finer resolution imagery would capture the ground truth more accurately and smoothen the snow depletion curve, to something that appears hydrologically more realistic, as demonstrated by examples and simulation in Section 5.3.2. The current state of the art in satellite remote sensing of snow is much

superior to the 1980's data used in this project. A prominent example of the current quality of satellite remote sensing imagery (especially of snow), would be the impressive snow data collected by the MODIS sensor. MODIS provides snow data images at a resolution of as fine as 500m, 50 times better than the 25-km data used in this study. The MODIS image shown in Fig.6-1 gives an idea of present technology in satellite remote sensing of snow. Unfortunately, the MODIS products are only available since 2000, while the in-situ hydrological records needed for modeling the Ob basin were discontinued when the former Soviet Union was dissolved.

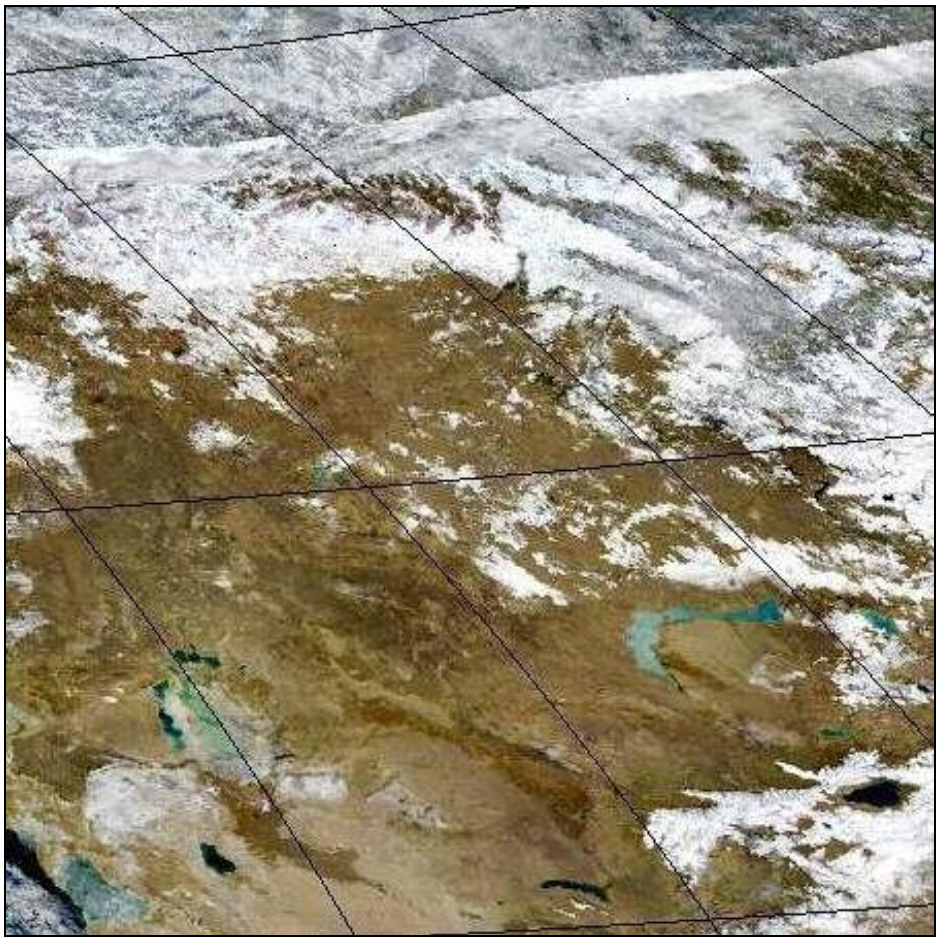


Figure 6-1. Sample snow image from MODIS for southern Siberia
For 11/9/2003, Goode Homolosine Projection (MODIS, 2003)

A major lesson learned from this study is the importance of a careful and thorough exploration of the data sources and algorithms used in creating a publicly-available dataset. The "Surface Type Mask" included in the "AVHRR Polar Pathfinder Twice-Daily 5 km EASE-Grid Composites" was derived from an SSM/I product with a 25-km resolution. The resolution of the satellite sensor initially recording the images from the earth's surface is much more important than the resolution of the images themselves. The re-sampling of a 25-km resolution satellite image into a 5-km image may not enhance any qualitative accuracy of the image if not re-sampled using some special disaggregation technique.

6.3 CONCLUSIONS ON THE DATASETS AND TOOLS USED

The utility of EASE-grid in research studies involving large areas in the polar regions of the earth is ably demonstrated by this project. This project demanded an "Equal Area" projection because of the need for computing areas of individual HRUs using the number of pixels they take up in a remotely sensed image of the area. Besides EASE-grid, this project involves the use of data from various sources. Data collected from different satellite sensors at different times, and stored in databases independent of each other, were assimilated and analyzed as part of this research project. Integrating all these data to get fruitful results and making them compatible with each other has been a valuable learning process.

Some dead ends were reached while taking different approaches to many problems in this research. One of the most notable was the intention to use the paper maps used by the NASA HSB team to delineate the Tom R. watershed and subareas. The

eight separate paper maps were successfully merged together in a digital format, but they failed to give an appropriate and reliable outline of the watershed. The alternative approach of using the DEM files from USGS proved to be more informative and accurate in determining the Tom R. watershed and the units of the spatially distributed model – the HRUs. Another important revelation was the inconsistencies of the header files of snow data images downloaded from NSIDC. Without any assumption, every time data was requested from NSIDC, the downloaded images were checked for georeferencing. The header file was changed accordingly to correct the georeferencing.

6.4 FUTURE RESEARCH

Future work to expand on the results of this study should include attempts to use much finer satellite imagery that capture the actual hydrological dynamics on the earth's surface much more accurately than the data used in this research. Current satellite sensors like MODIS would be a good choice for high resolution snow data. Even though one of the secondary objectives of this research was to use the SCA time series derived from satellite data to calibrate a spatially distributed hydrological model, the primary objective was to see how good a match the two datasets were in the first place. In future, the spatial model results calibrated on some other parameters, such as point measurements of SWE, may yield a much better match with fine resolution satellite imagery, than absolutely uncalibrated model results. Lastly, the need to develop a superior technique to simulate snow (accumulation and depletion) was identified when analyzing the results of this research. In the future, a very complex snow simulation model may be developed that simulates the spatial distribution of snow. Such a model would be beneficial to studies on

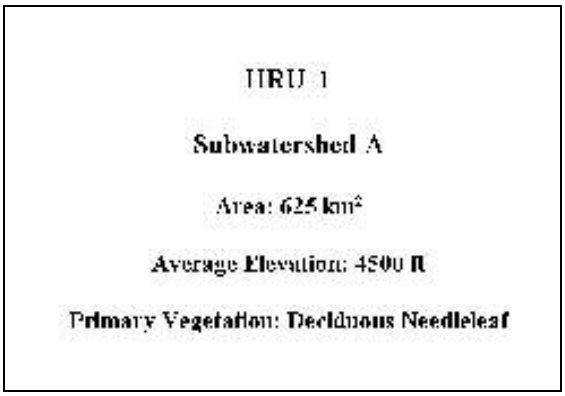
snow hydrology, including studies on the effect of image resolution on hydrologic modeling.

Hydrologists seeking to use satellite data products to calibrate and validate distributed models are advised to develop a thorough understanding of the different sensors, the spatial resolution of the products, and the algorithms used to convert remotely sensed information (such as brightness temperatures) to hydrologically meaningful quantities. This knowledge will allow hydrologists to understand the strengths and weaknesses of available datasets, and select appropriate products to support their investigation.

APPENDIX A

Comparative plots of the SCA time series from both data sources: satellite data from NSIDC and uncalibrated results from the NASA HSB team's PRMS model.

- Each of the following pages contains five plots, one each for the five calendar years that spanned the target time period for this research (1981-1985).
- The Area covered by each HRU as mentioned in this appendix is computed by ArcView from the 5-km resolution HRU masks (Table 4-4).
- The average elevation data was provided by the NASA HSB team.
- The Primary Vegetation information is derived from grouping land cover categories from the GLCF database.
- Figures A-96 to A-100 are for HRU 20 for which PRMS results were not available.



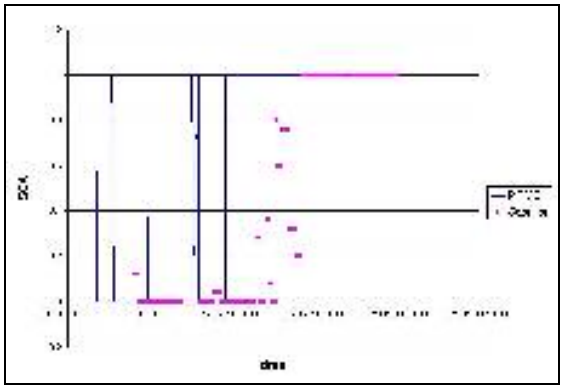


Figure A-1. HRU 1, 1981

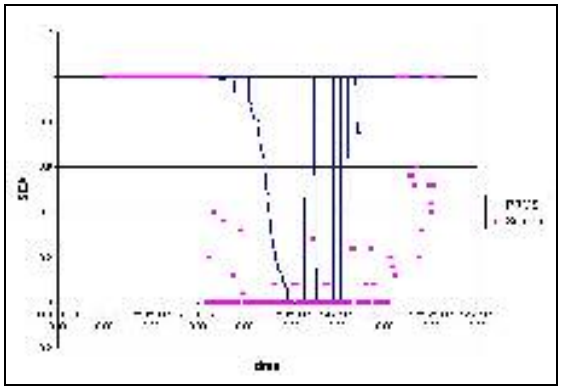


Figure A-2. HRU 1, 1982

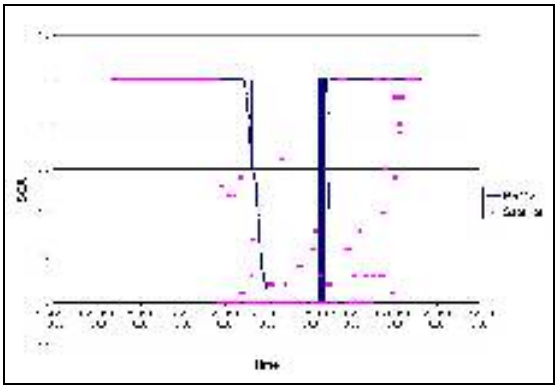


Figure A-3. HRU 1, 1983

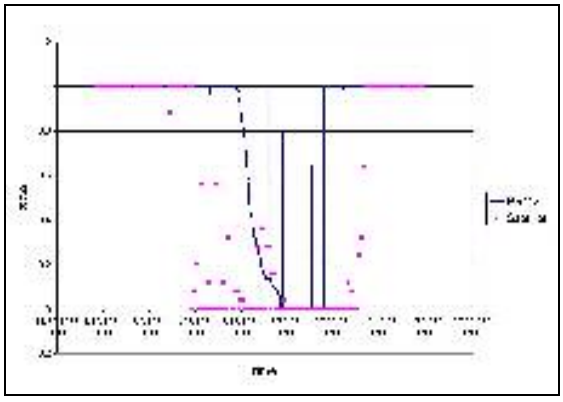


Figure A-4. HRU 1, 1984

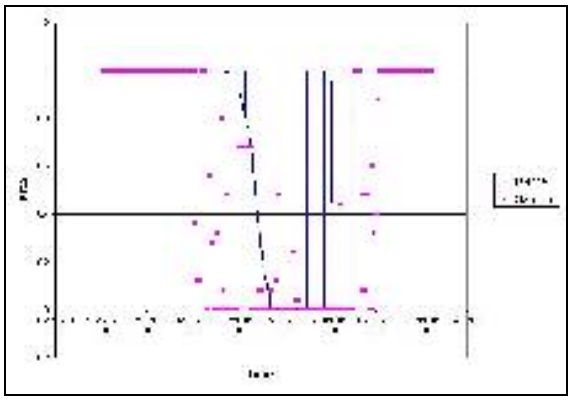


Figure A-5. HRU 1, 1985

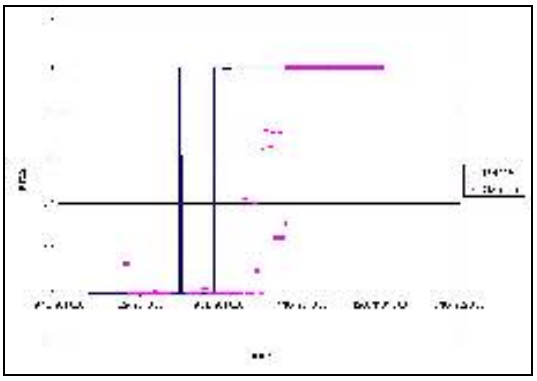


Figure A-6. HRU2, 1981

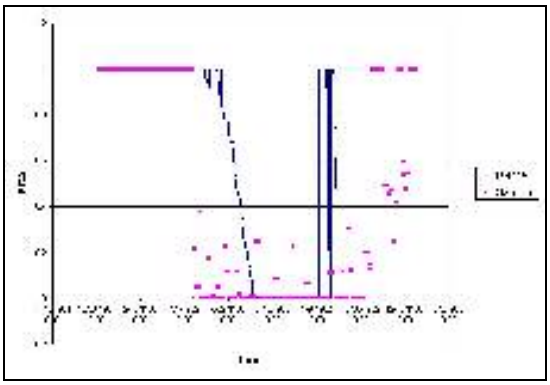


Figure A-7. HRU 2, 1982

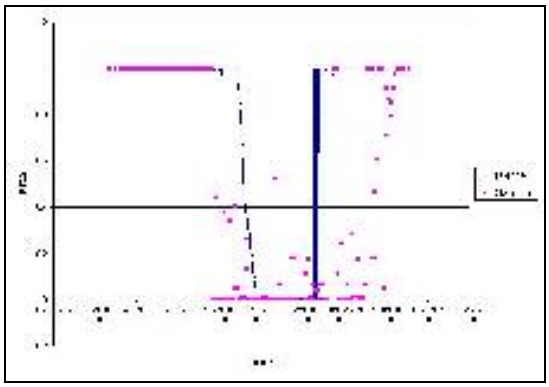


Figure A-8. HRU 2, 1983

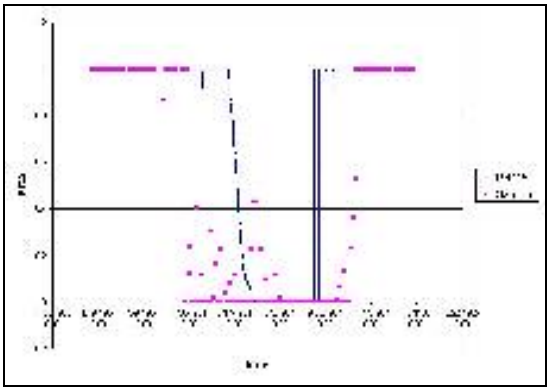


Figure A-9, HRU 2, 1984

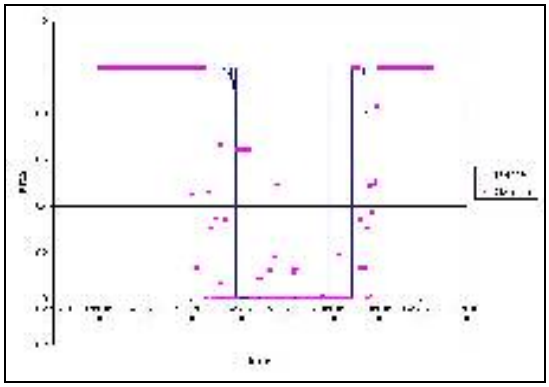


Figure A-10. HRU 2, 1985

HRU 3

Subwatershed A

Area: 550 km²

Average Elevation: 1500 ft

Primary Vegetation: Woody Savannas

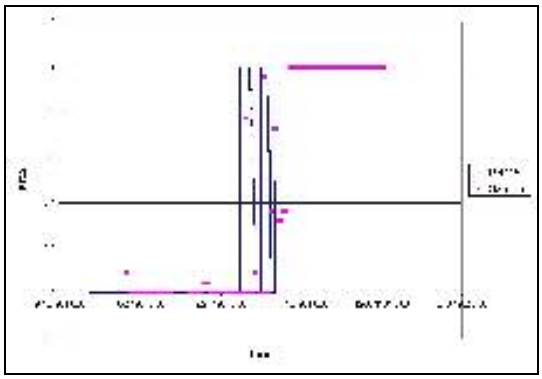


Figure A-11. HRU 3, 1981

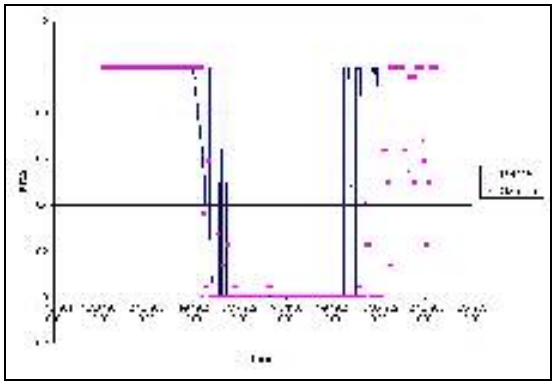


Figure A-12. HRU 3, 1982

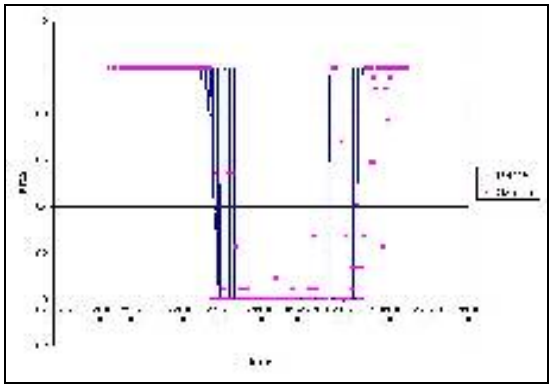


Figure A-13. HRU 3, 1983

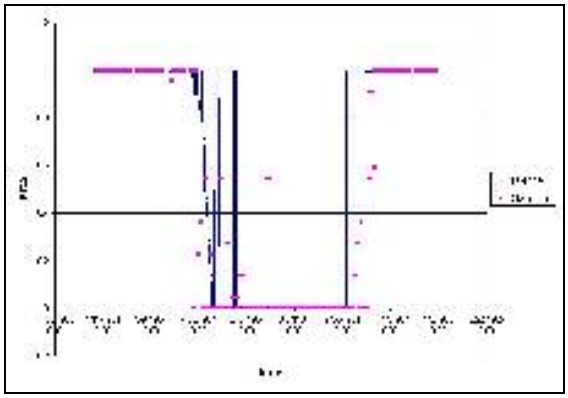


Figure A-14. HRU 3, 1984

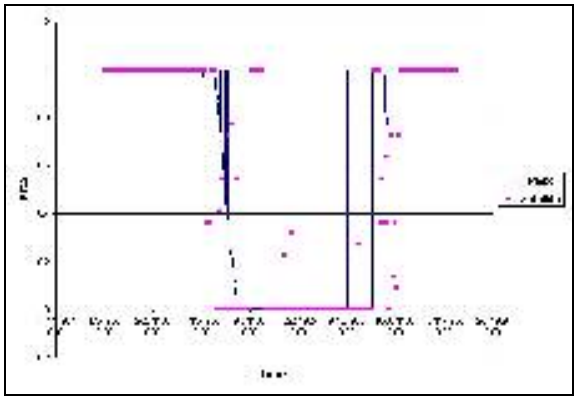


Figure A-15. HRU 3, 1985

HRU 4

Subwatershed A

Area: 600 km²

Average Elevation: 1500 ft

Primary Vegetation: Croplands

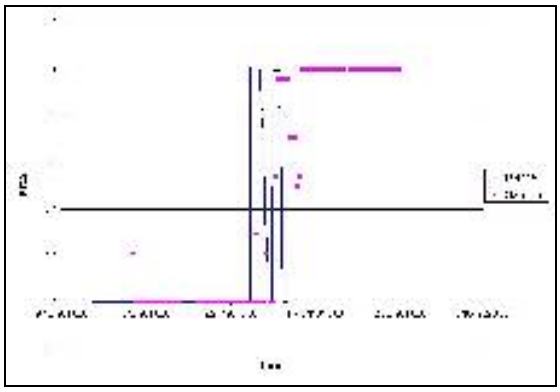


Figure A-16. HRU 4, 1981

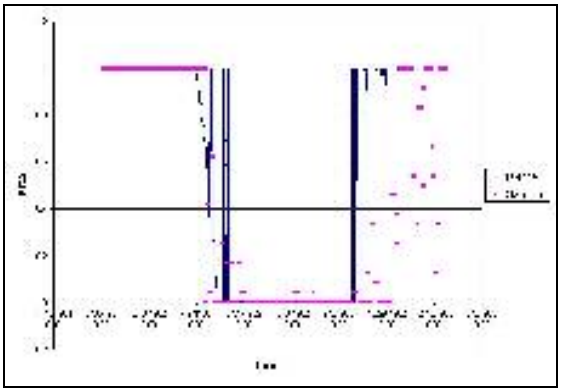


Figure A-17. HRU 4, 1982

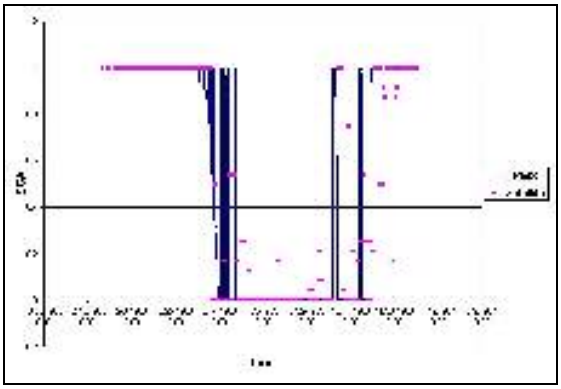


Figure A-18. HRU 4, 1983

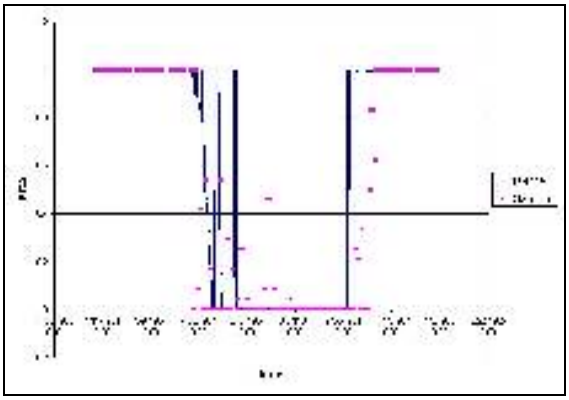


Figure A-19. HRU 4, 1984

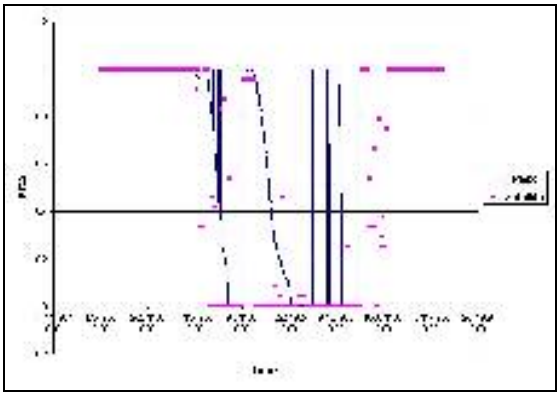


Figure A-20. HRU 4, 1985

HRU 5

Subwatershed B

Area: 175 km²

Average Elevation: 4500 ft.

Primary Vegetation: Deciduous Needleleaf

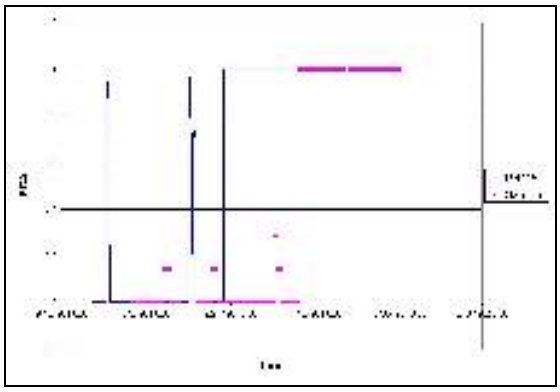


Figure A-21. HRU 5, 1981

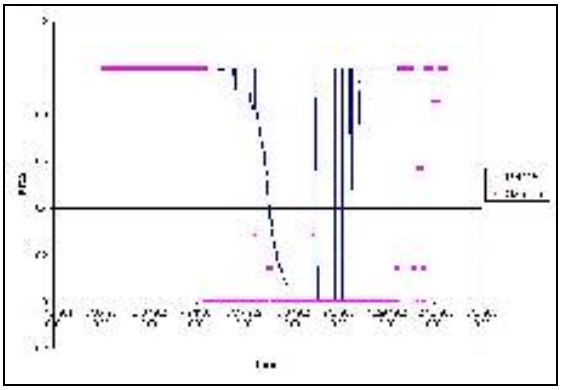


Figure A-22. HRU 5, 1982

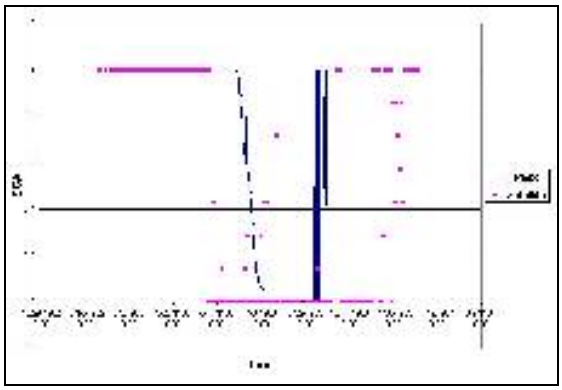


Figure A-23. HRU 5, 1983

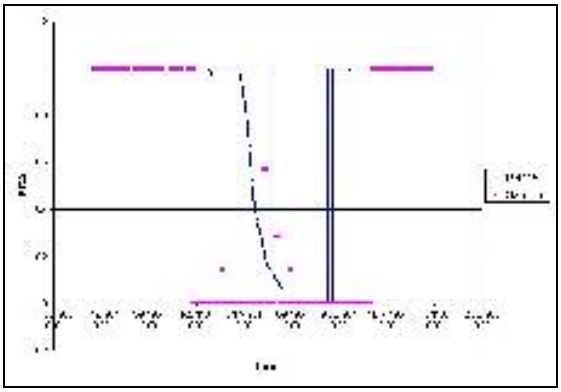


Figure A-24. HRU 5, 1984

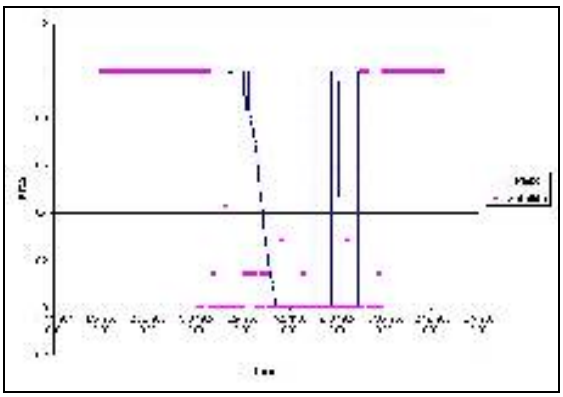


Figure A-25. HRU 5, 1985

HRU 6

Subwatershed B

Area: 7950 km²

Average Elevation: 3500 ft.

Primary Vegetation: Mixed Forest

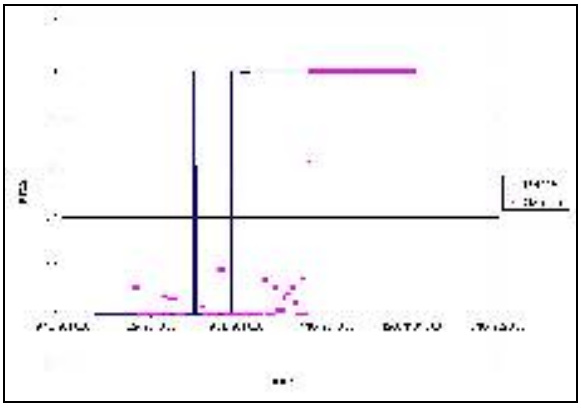


Figure A-26. HRU 6, 1981

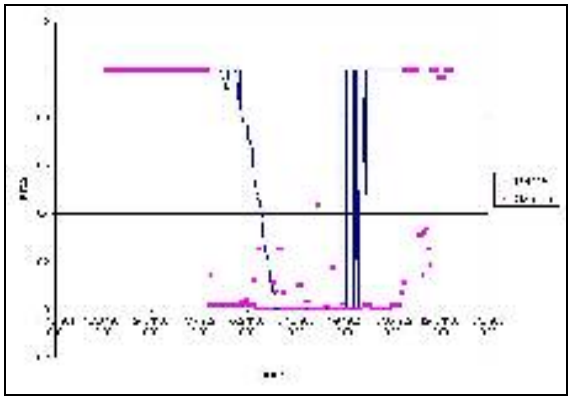


Figure A-27. HRU 6, 1982

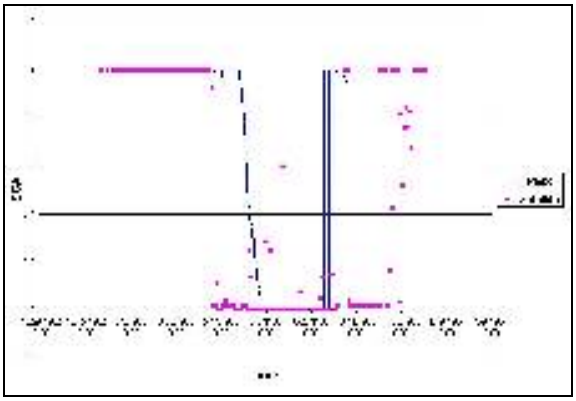


Figure A-28. HRU 6, 1983

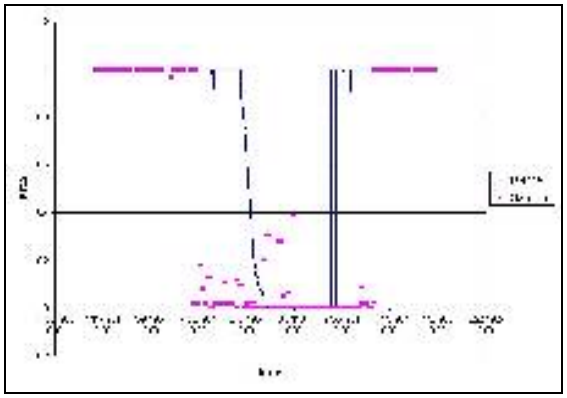


Figure A-29. HRU 6, 1984

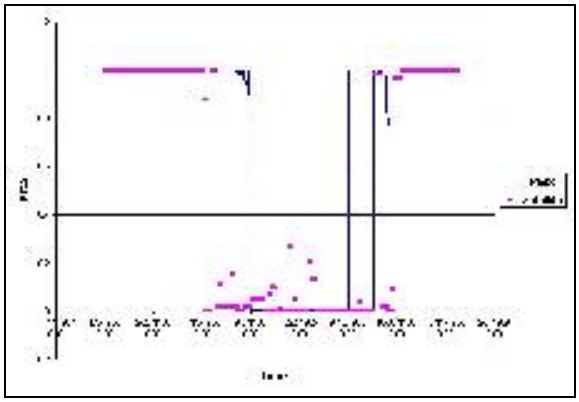


Figure A-30. HRU 6, 1985

HRU 7

Subwatershed B

Area: 275 km²

Average Elevation: 1500 ft.

Primary Vegetation: Woody Savannas

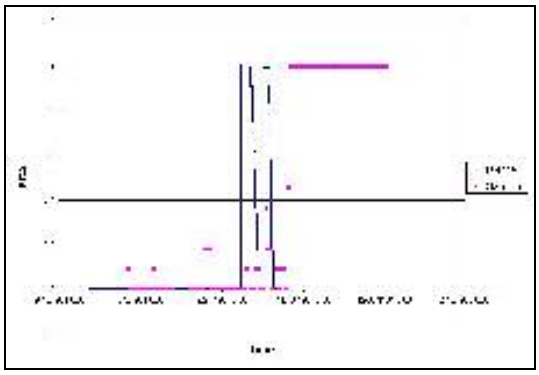


Figure A-31. HRU 7, 1981

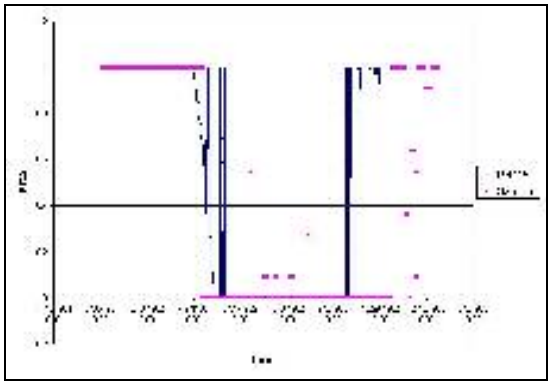


Figure A-32. HRU 7, 1982

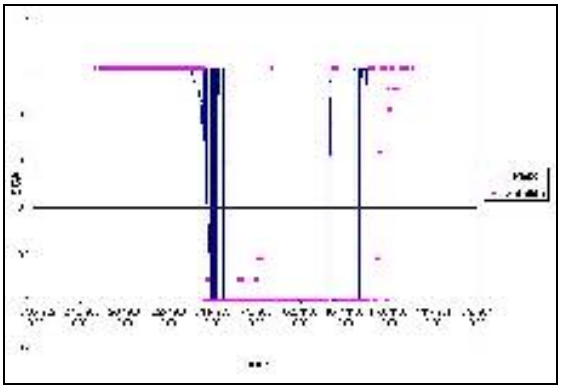


Figure A-33. HRU 7, 1983

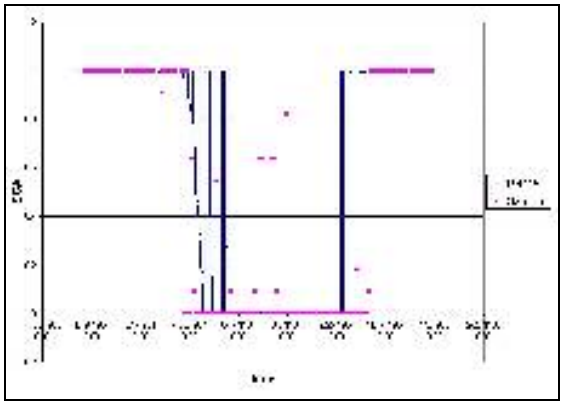


Figure A-34. HRU 7, 1984

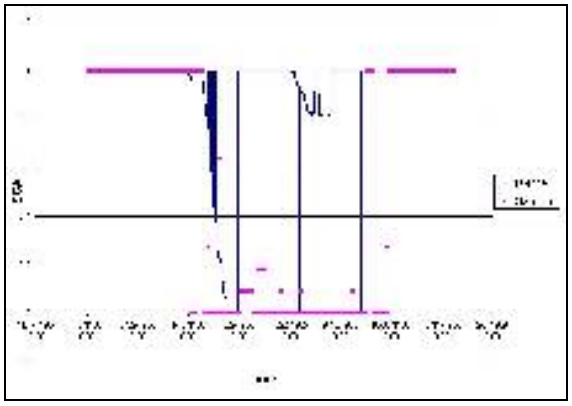


Figure A-35. HRU 7, 1985

HRU 8

Subwatershed C

Area: 1675 km²

Average Elevation: 4500 ft.

Primary Vegetation: Deciduous Needleleaf

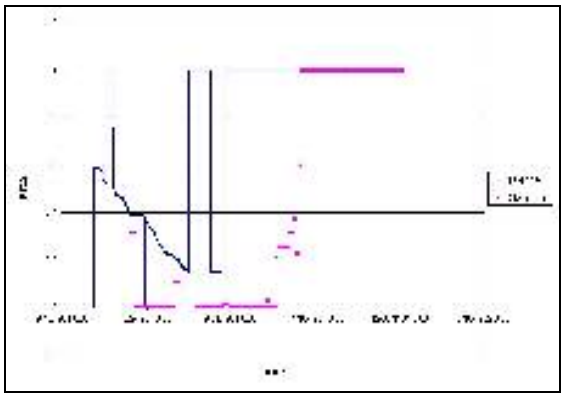


Figure A-36. HRU 8, 1981

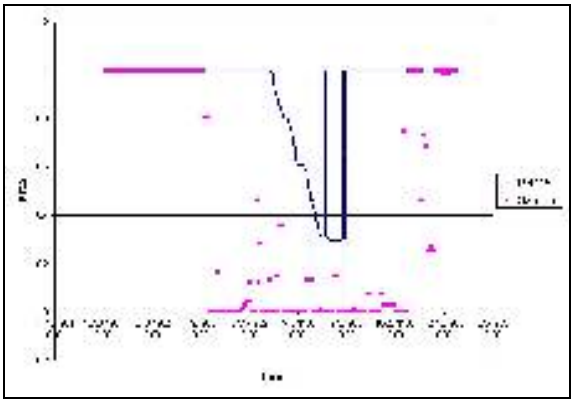


Figure A-37. HRU 8, 1982

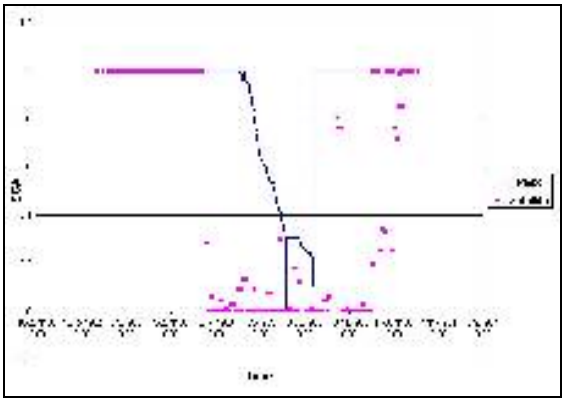


Figure A-38. HRU 8, 1983

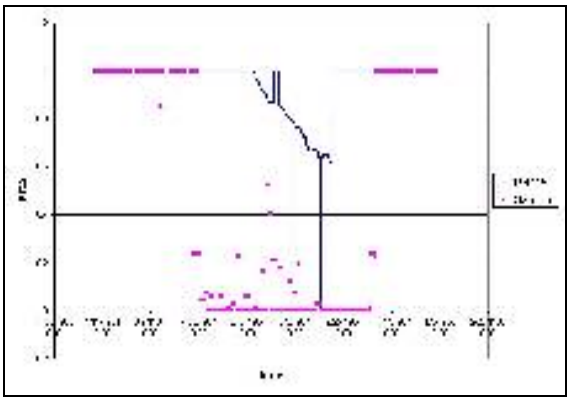


Figure A-39. HRU 8, 1984

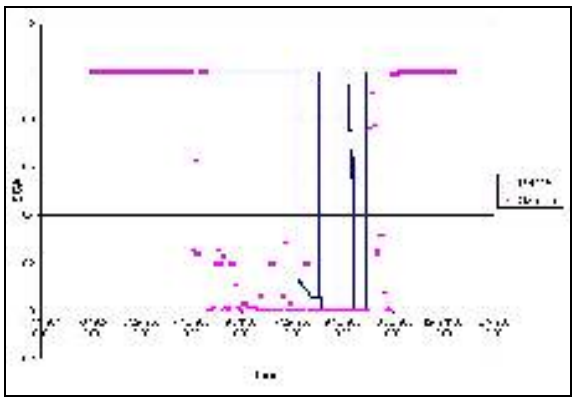


Figure A-40. HRU 8, 1985

HRU 9

Subwatershed C

Area: 13100 km²

Average Elevation: 3500 ft

Primary Vegetation: Mixed Forest

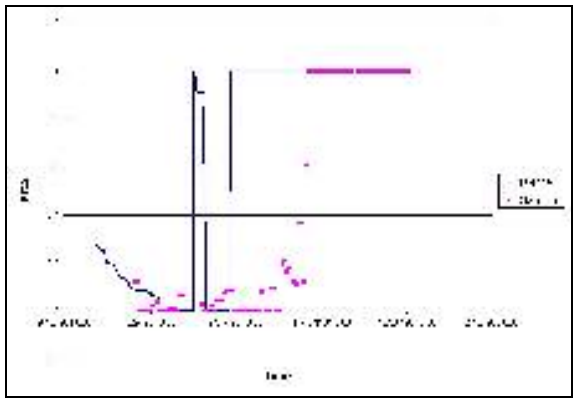


Figure A-41. HRU 9, 1981

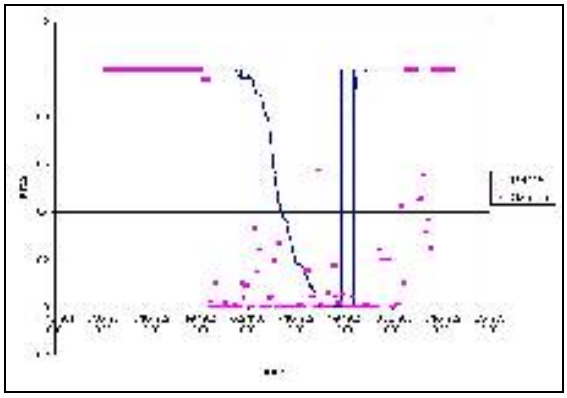


Figure A-42. HRU 9, 1982

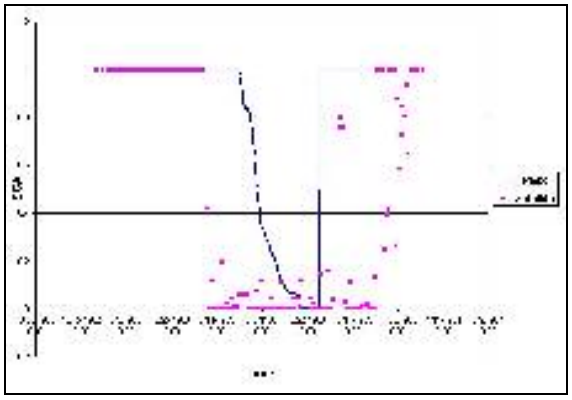


Figure A-43. HRU 9, 1983

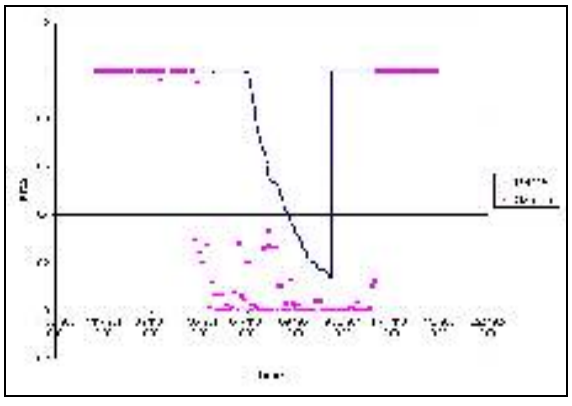


Figure A-44. HRU 9, 1984

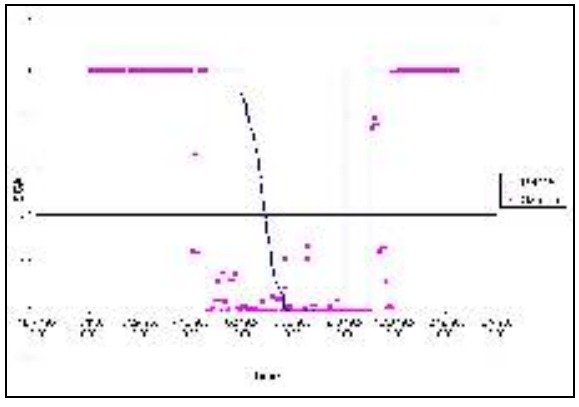
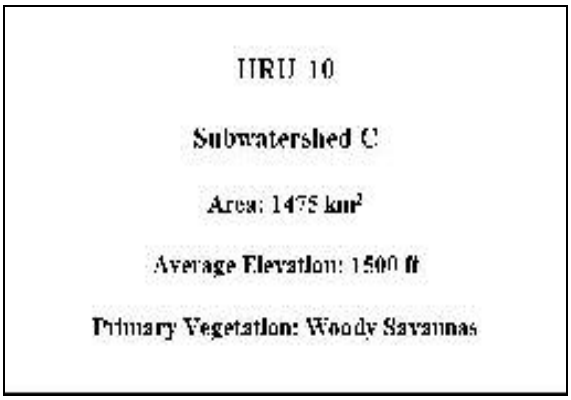


Figure A-45. HRU 9, 1985



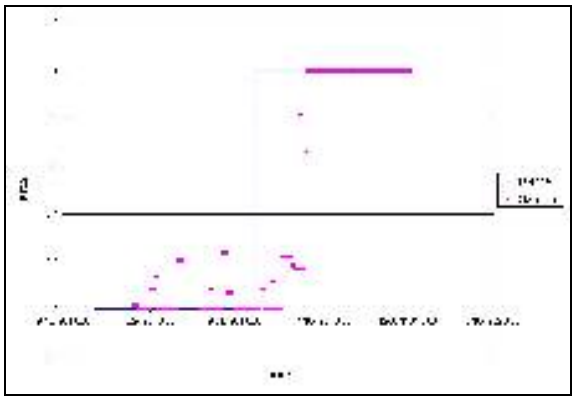


Figure A-46. HRU 10, 1981

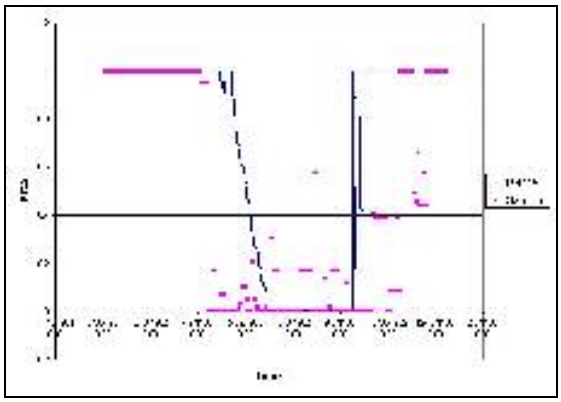


Figure A-47. HRU 10, 1982

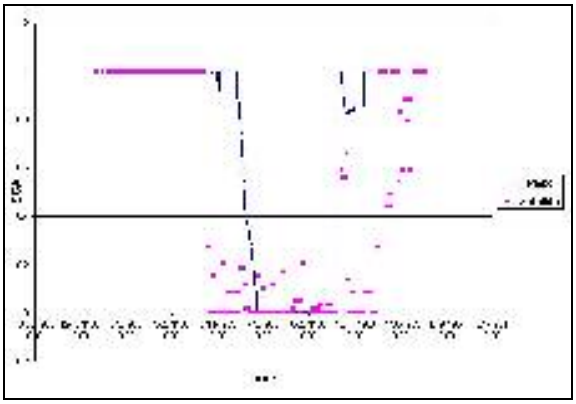


Figure A-48. HRU 10, 1983

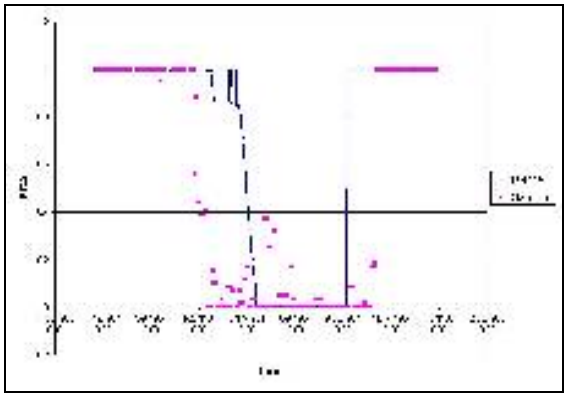


Figure A-49. HRU 10, 1984

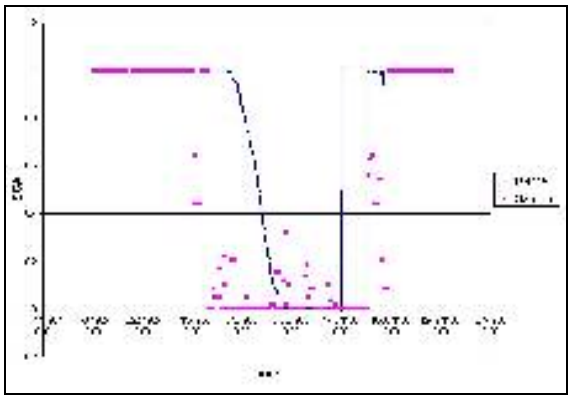


Figure A-50. HRU 10, 1985

HRU 11

Subwatershed C

Area: 1850 km²

Average Elevation: 1500 ft

Primary Vegetation: Croplands

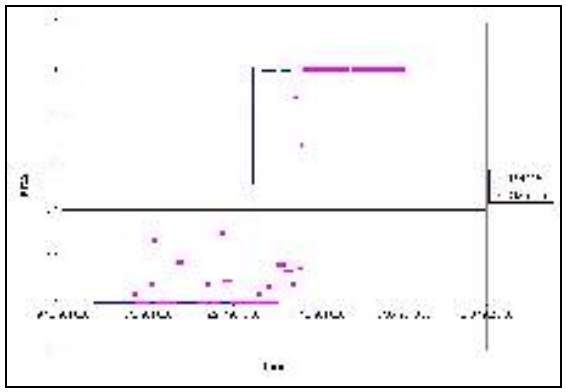


Figure A-51. HRU 11, 1981

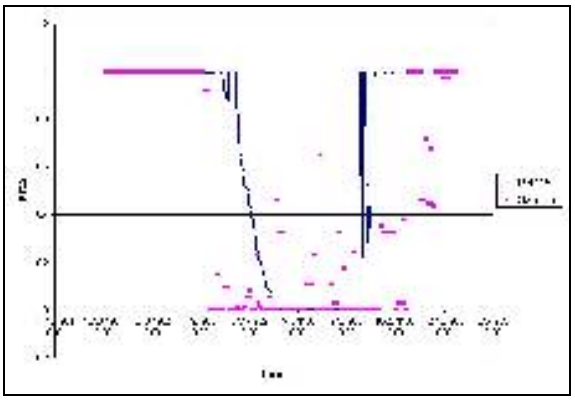


Figure A-52. HRU 11, 1982

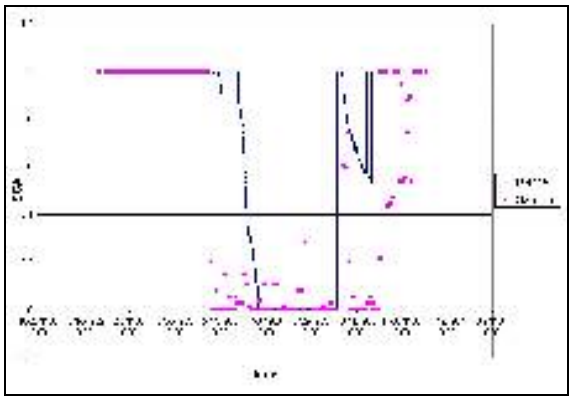


Figure A-53. HRU 11, 1983

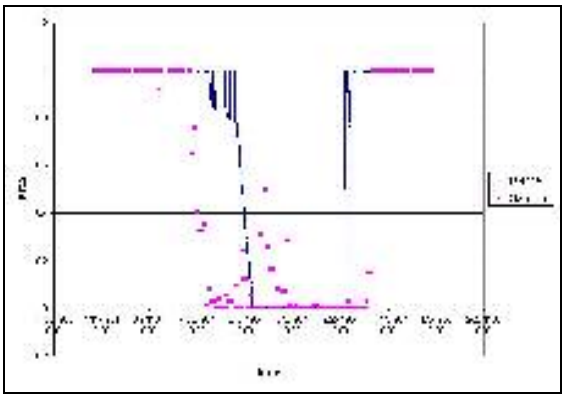


Figure A-54. HRU 11, 1984

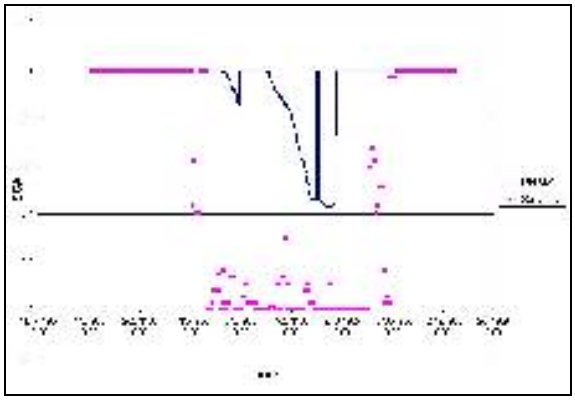


Figure A-55. HRU 11, 1985

HRU 12

Subwatershed D

Area: 200 km²

Average Elevation: 4000 ft.

Primary Vegetation: Deciduous Needleleaf

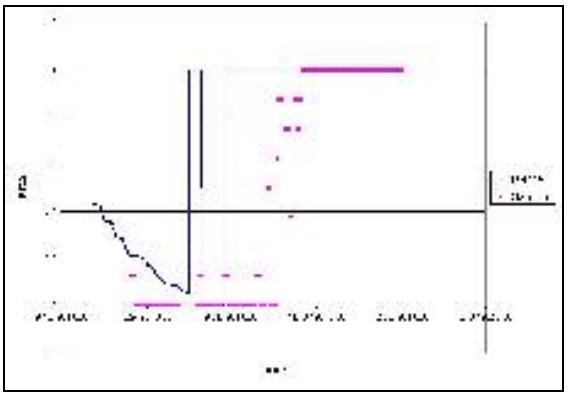


Figure A-56. HRU 12, 1981

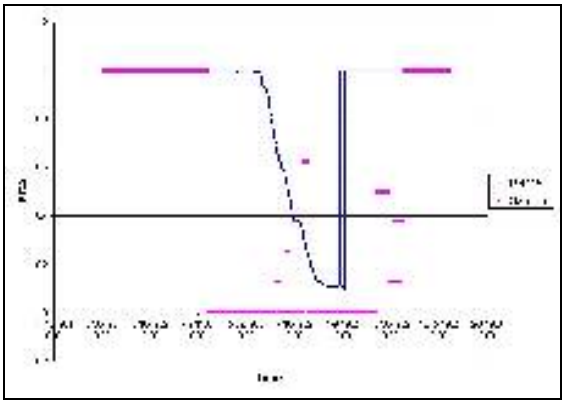


Figure A-57. HRU 12, 1982

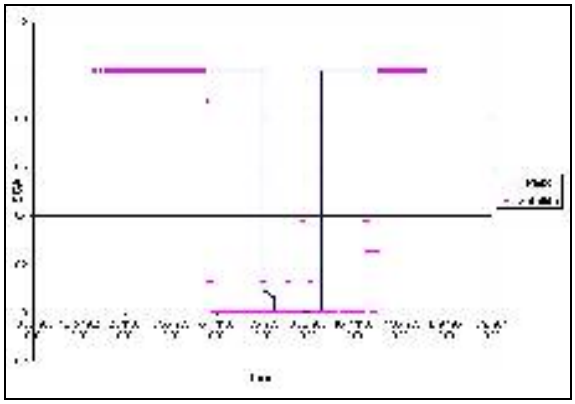


Figure A-58. HRU 12, 1983

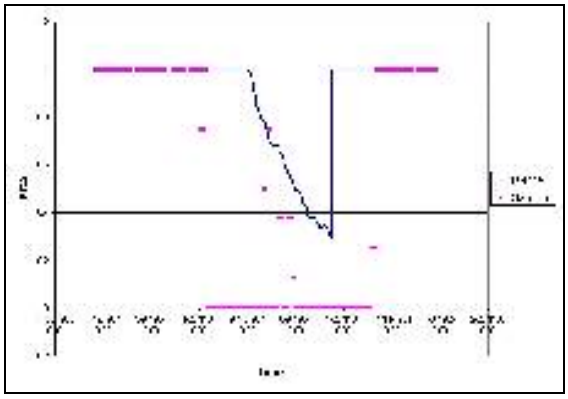


Figure A-59. HRU 12, 1984

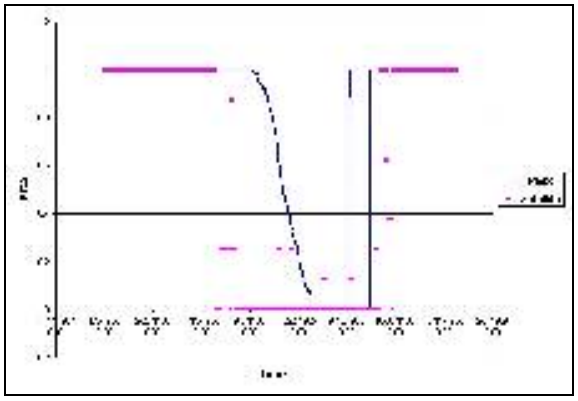
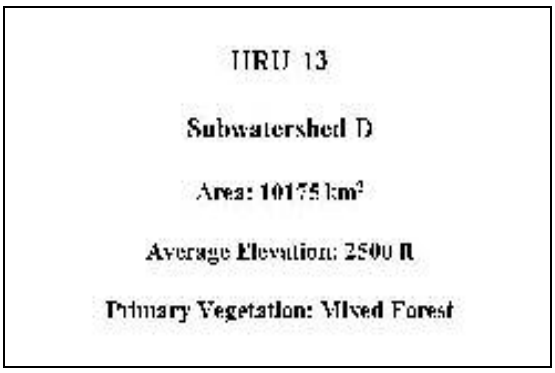


Figure A-60. HRU 12, 1985



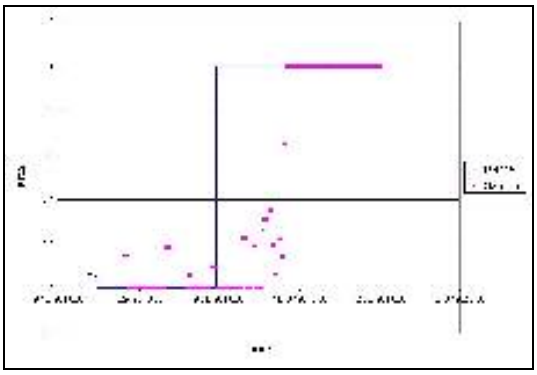


Figure A-61. HRU 13, 1981

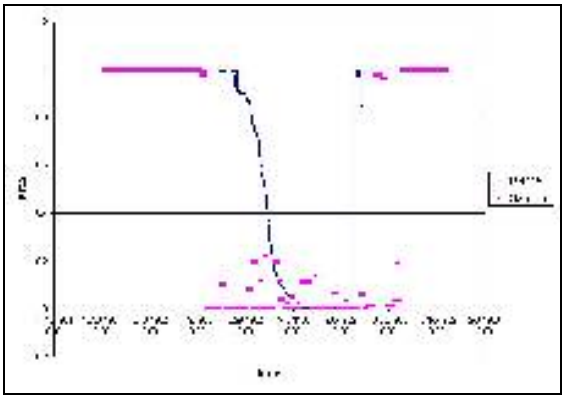


Figure A-62. HRU 13, 1982

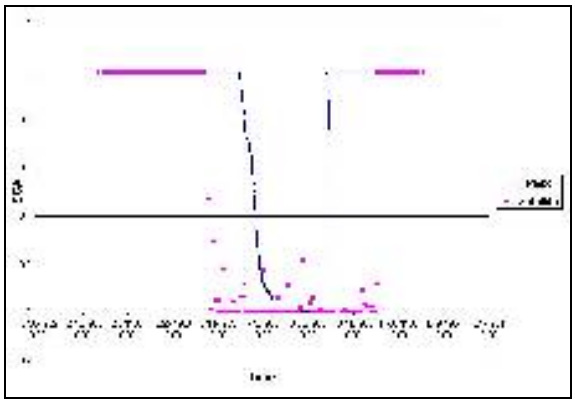


Figure A-63. HRU 13, 1983

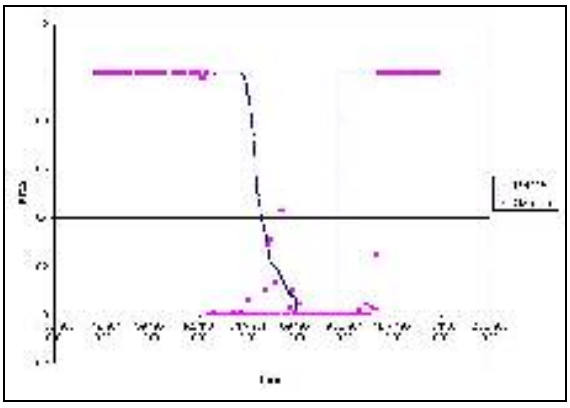


Figure A-64. HRU 13, 1984

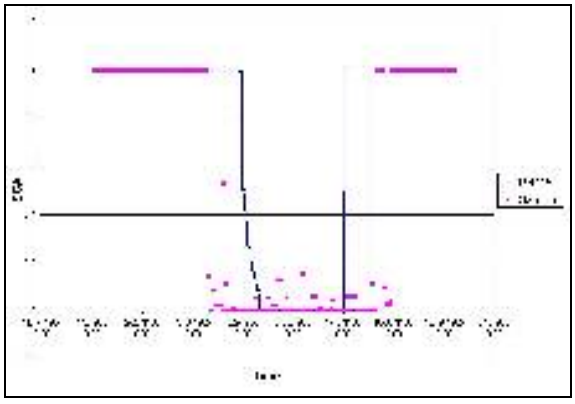


Figure A-65. HRU 13, 1985

HRU 14

Subwatershed D

Area: 1575 km^2

Average Elevation: 1100 ft

Primary Vegetation: Woody Savannas

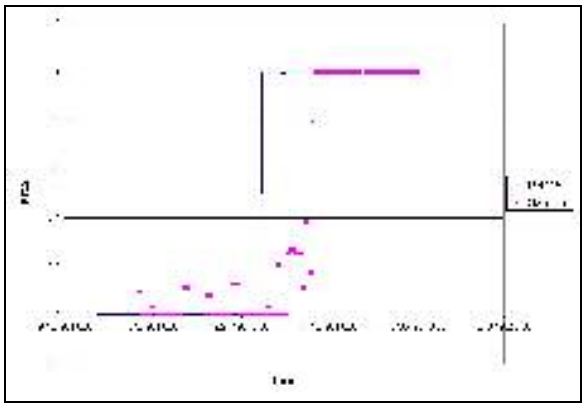


Figure A-66. HRU 14, 1981

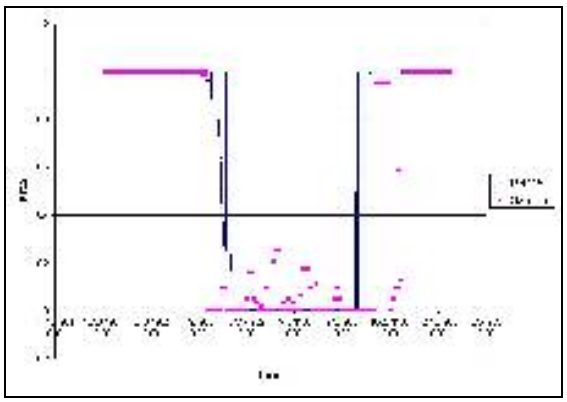


Figure A-67. HRU 14, 1982

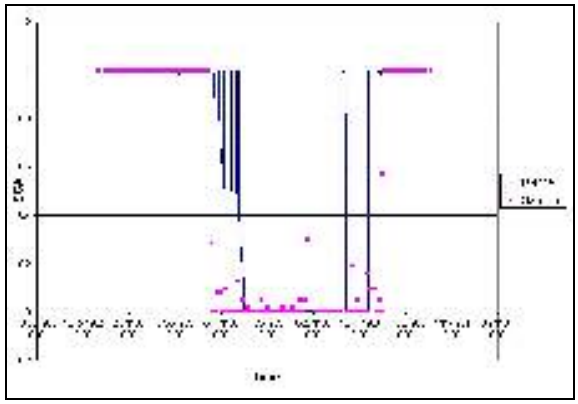


Figure A-68. HRU 14, 1983

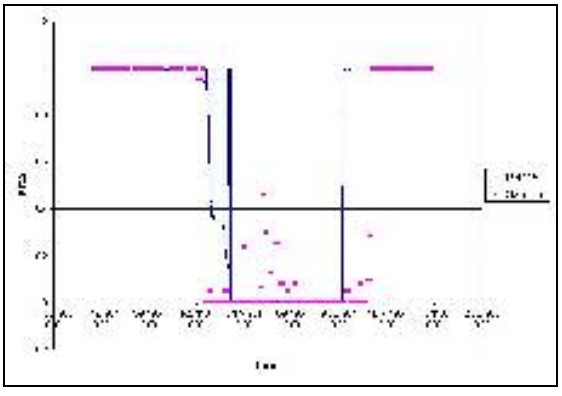


Figure A-69. HRU 14, 1984

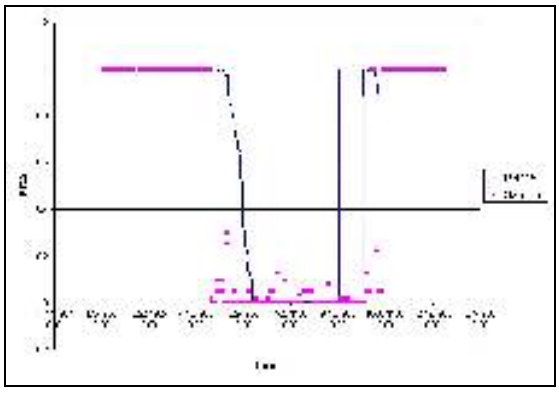


Figure A-70. HRU 14, 1985

HRU 15

Subwatershed D

Area: 3400 km²

Average Elevation: 1100 ft

Primary Vegetation: Croplands

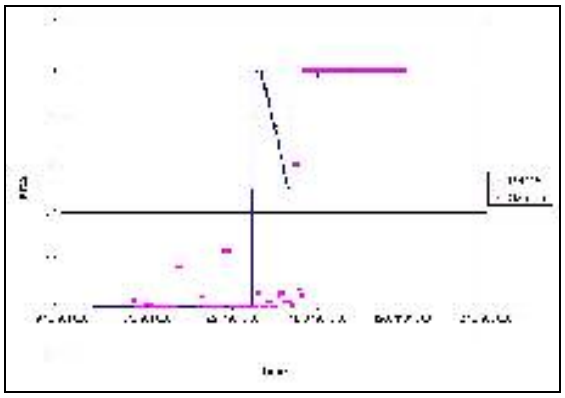


Figure A-71. HRU 15, 1981

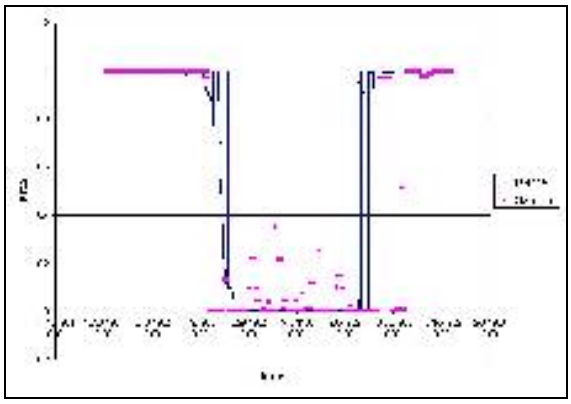


Figure A-72. HRU 15, 1982

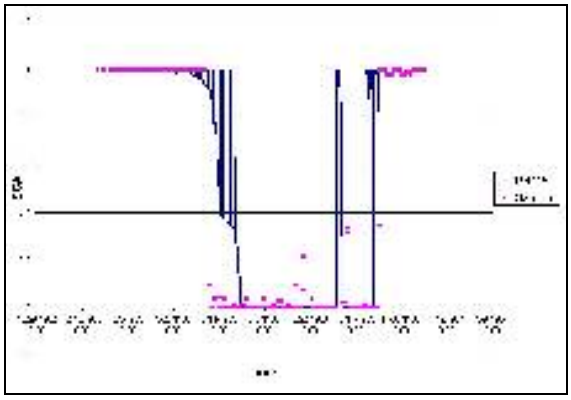


Figure A-73. HRU 15, 1983

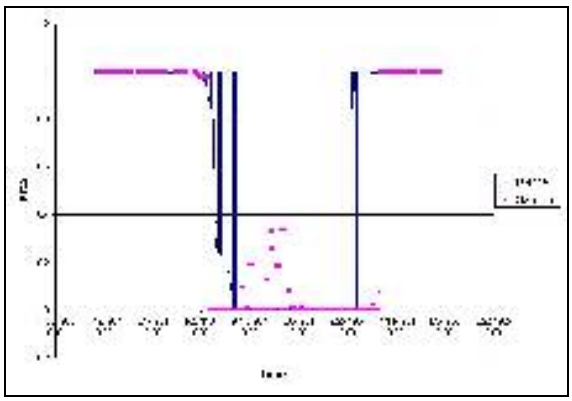


Figure A-74. HRU 15, 1984

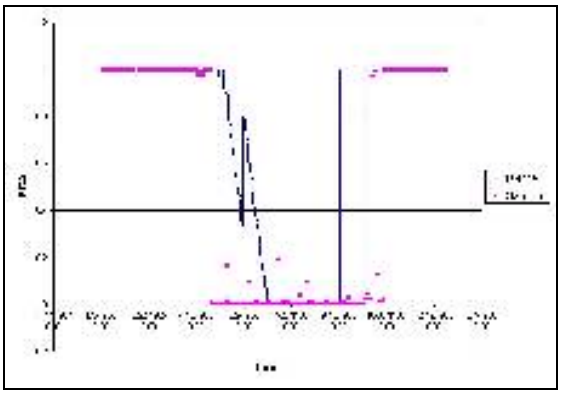


Figure A-75. HRU 15, 1985

HRU 16

Subwatershed E

Area: 1025 km²

Average Elevation: 1270 ft

Primary Vegetation: Deciduous Needleleaf

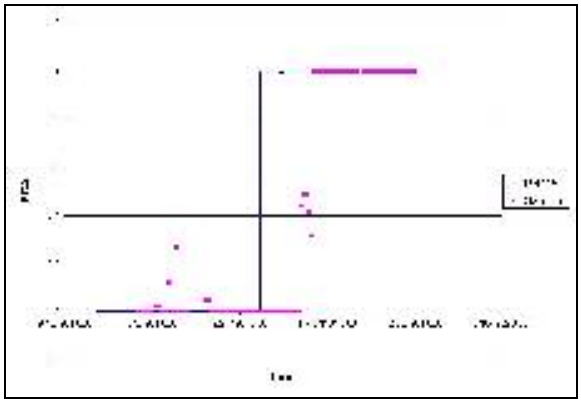


Figure A-76. HRU 16, 1981

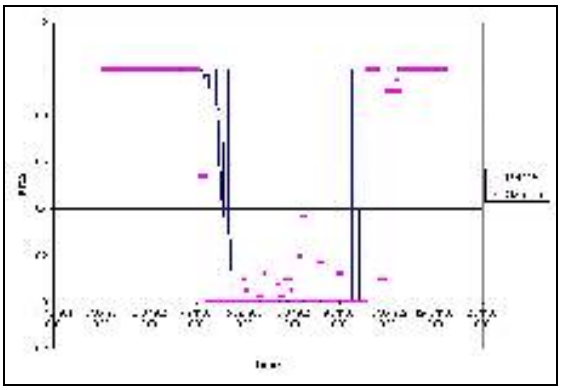


Figure A-77. HRU 16, 1982

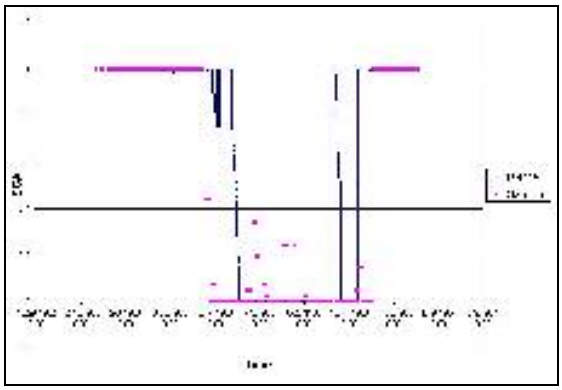


Figure A-78. HRU 16, 1983

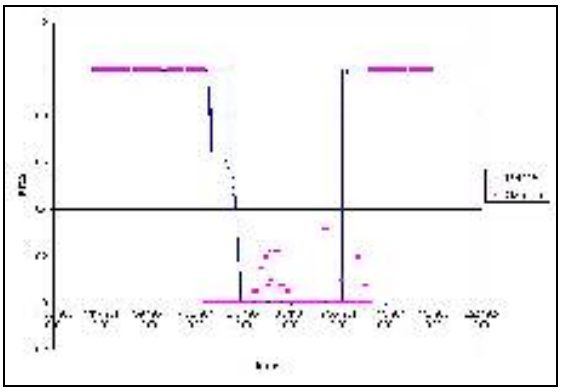


Figure A-79. HRU 16, 1984

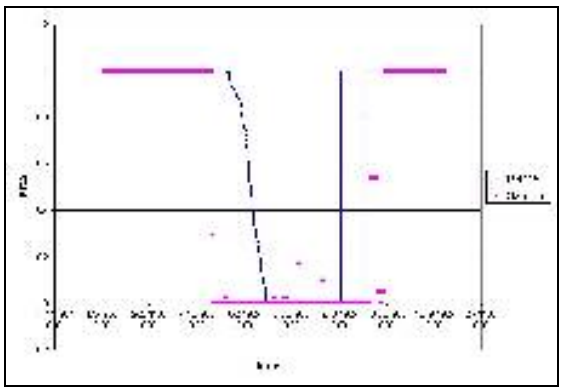


Figure A-80. HRU 16, 1985

HRU 17

Subwatershed E

Area: 6650 km²

Average Elevation: 1060 ft.

Primary Vegetation: Mixed Forest

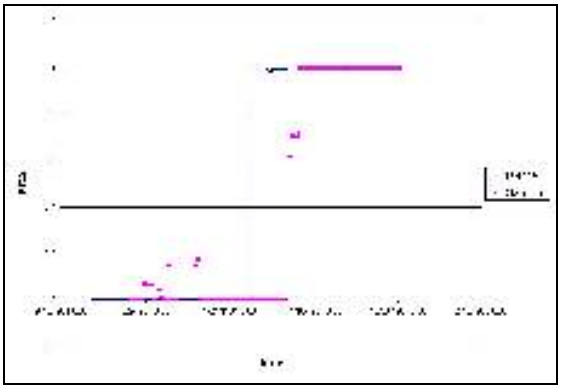


Figure A-81. HRU 17, 1981

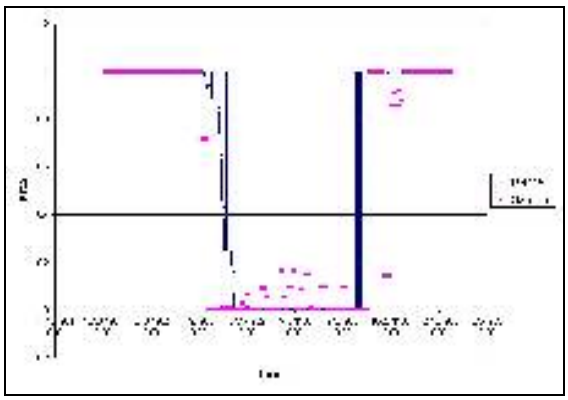


Figure A-82. HRU 17, 1982

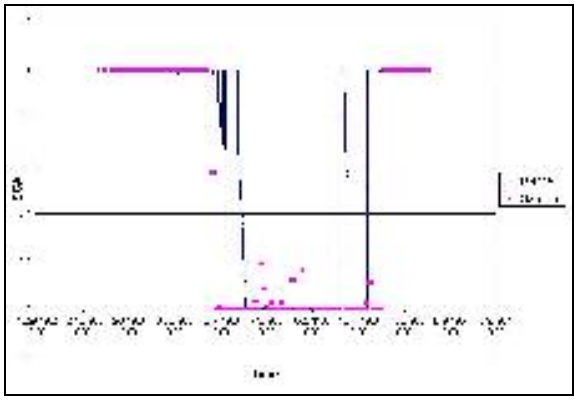


Figure A-83. HRU 17, 1983

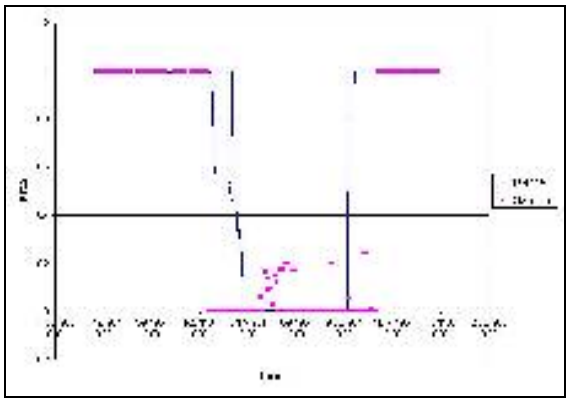


Figure A-84. HRU 17, 1984

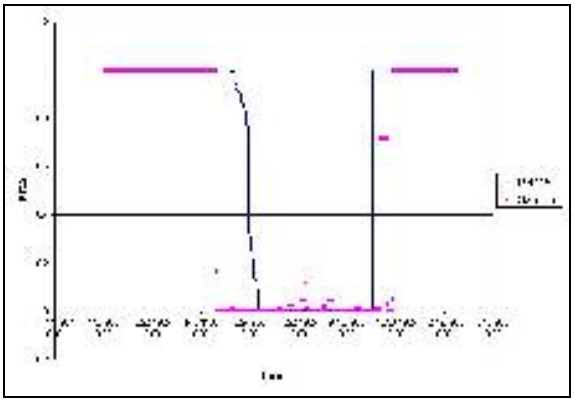
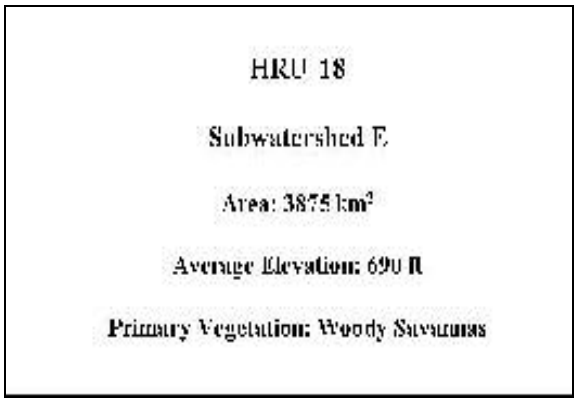


Figure A-85. HRU 17, 1985



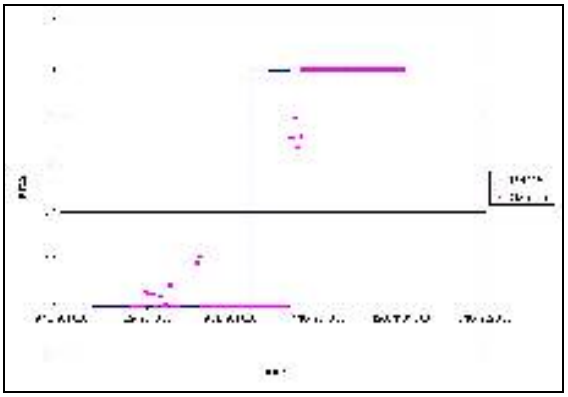


Figure A-86. HRU 18, 1981

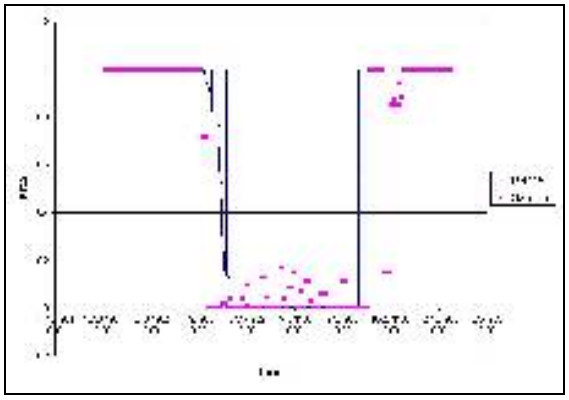


Figure A-87. HRU 18, 1982

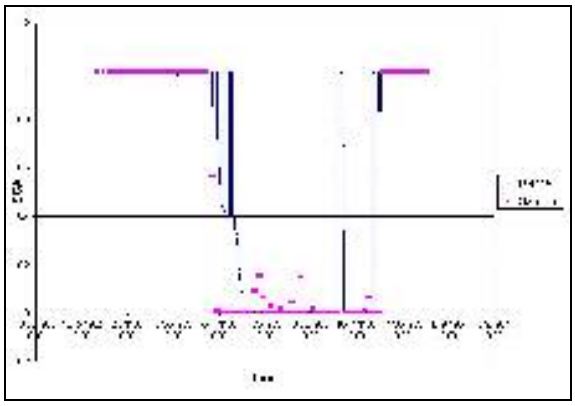


Figure A-88. HRU 18, 1983

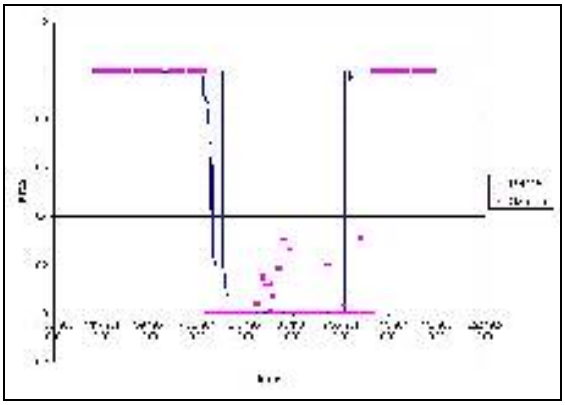


Figure A-89. HRU 18, 1984

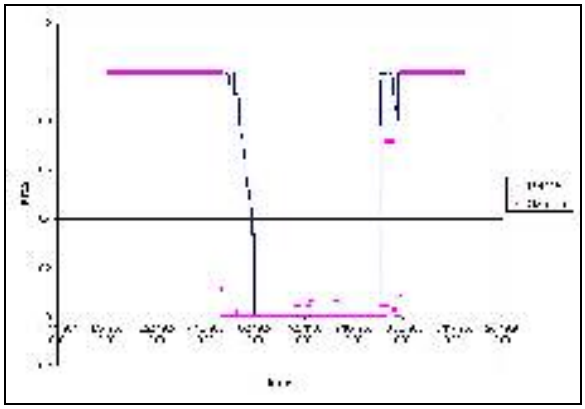


Figure A-90. HRU 18, 1985

HRU 19

Subwatershed E

Area: 3950 km²

Average Elevation: 350 ft

Primary Vegetation: Croplands

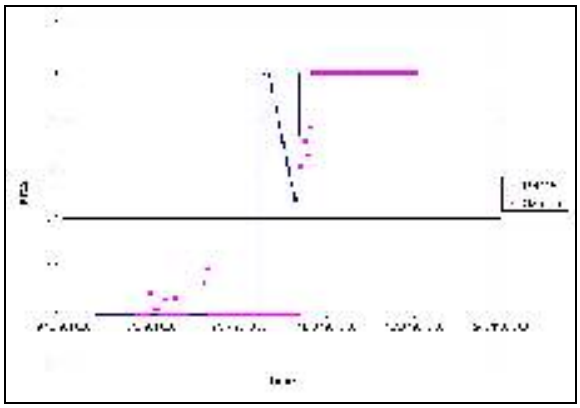


Figure A-91. HRU 19, 1981

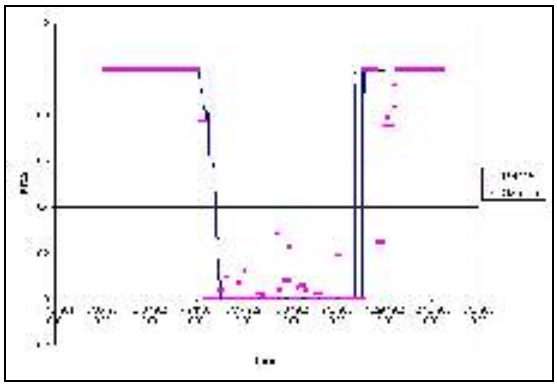


Figure A-92. HRU 19, 1982

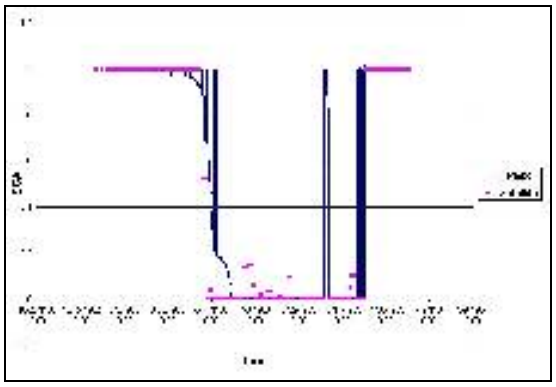


Figure A-93. HRU 19, 1983

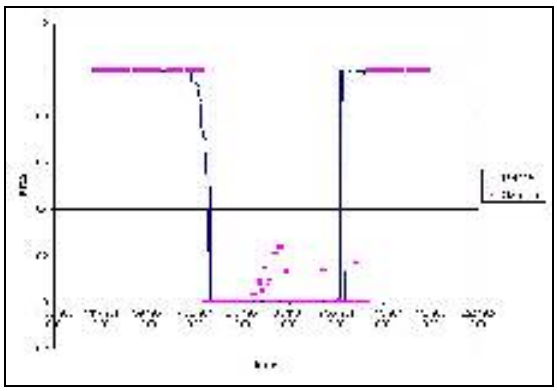


Figure A-94. HRU 19, 1984

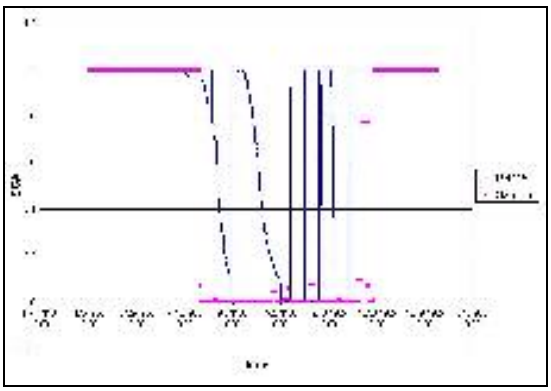


Figure A-95. HRU 19, 1985

HRU 20

Subwatershed B

Area: 450 km²

Average Elevation: N.A.

Primary Vegetation: Croplands

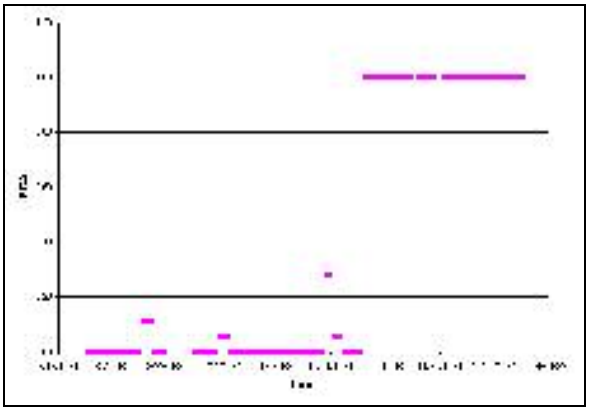


Figure A-96. HRU 20, 1981

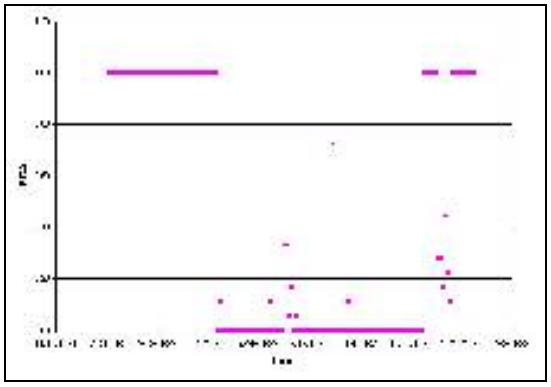


Figure A-97. HRU 20, 1982

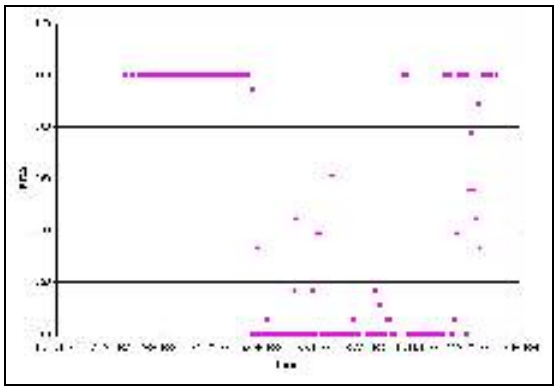


Figure A-98. HRU 20, 1983

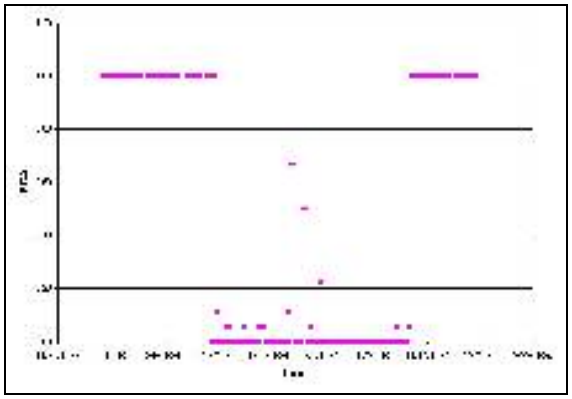


Figure A-99. HRU 20, 1984

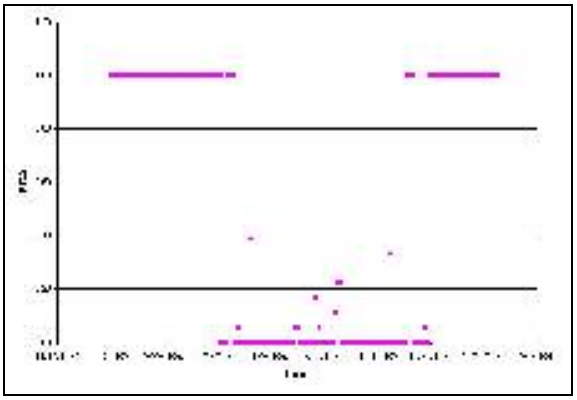


Figure A-100. HRU 20, 1985

APPENDIX B

3.5 Each data element in the correlogram curve in the plots shown in this appendix represents the value of the coefficient of cross-correlation (R) between the two time series at the lag-duration specified.

- The values on the right of the correlogram are not the lag-duration. Value 20 is for zero-lag. Each value more than 20 is a corresponding positive lag. E.g., value 26 is a 5-day positive lag. Similarly values less than 20 indicate a negative lag.
- The values on the left side of each plot are the actual values of R for that lag-duration, i.e., the value that the '*' is denoting in the plot.
- The output SCA time series from the PRMS model leads the cross correlation analysis.
- The correlogram analyses are only for the snow depletion periods of the time series: May 1 to July 31 (92 days) for the years 1982, 1983, 1984, 1985
- Data during that duration is not available from the satellite data source and so the year 1981 is omitted from this analysis.

HRU 1

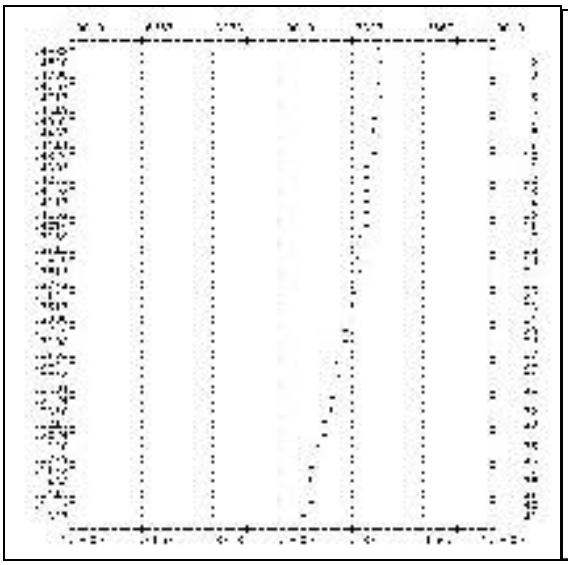


Figure B-1. HRU 1, 1982

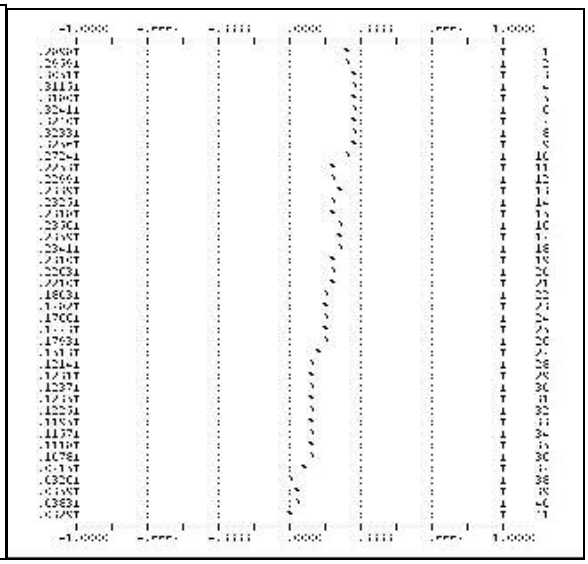


Figure B-2. HRU 1, 1983

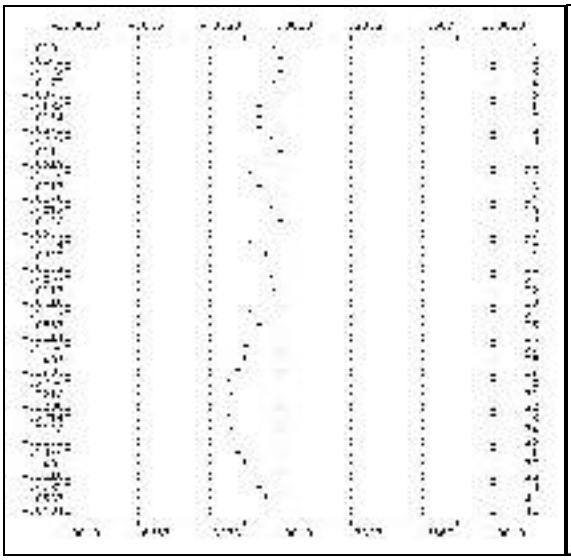


Figure B-3. HRU 1, 1984

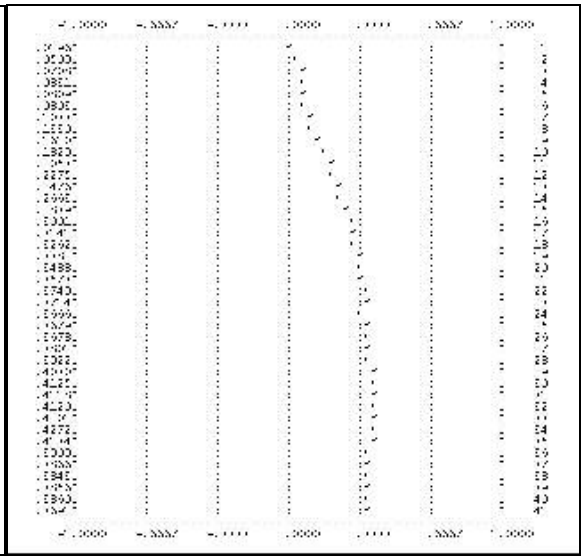


Figure B-4. HRU 1, 1985

HRU 2

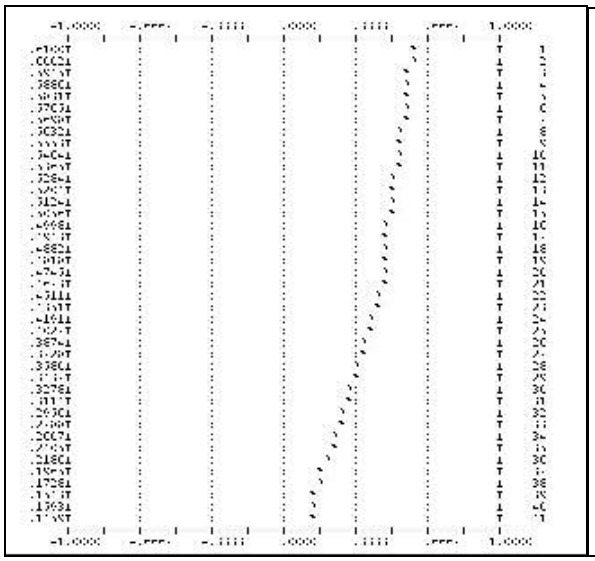


Figure B-5. HRU 2, 1982

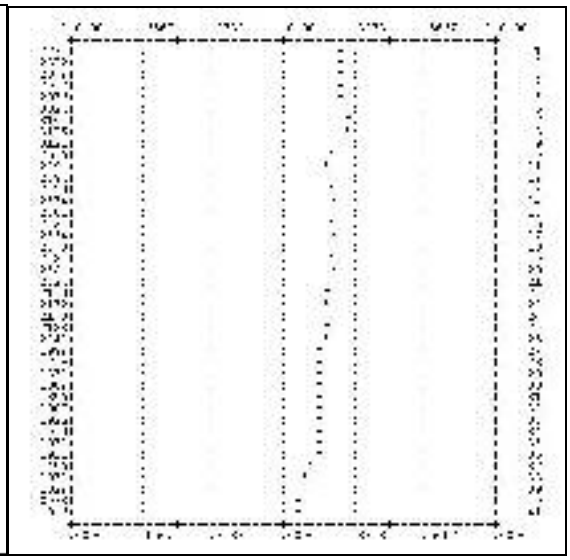


Figure B-6. HRU 2, 1983

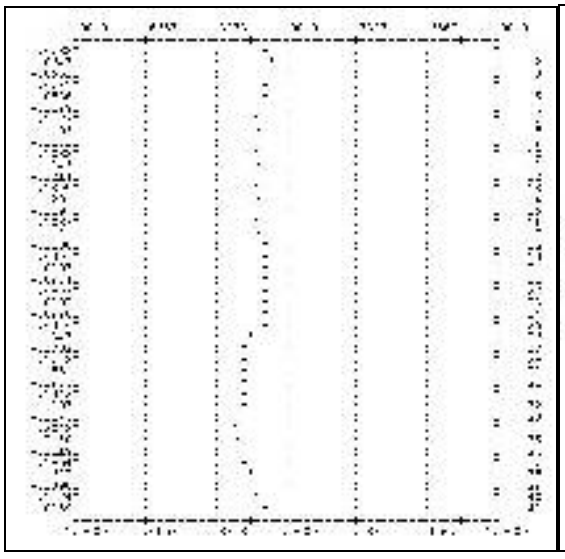


Figure B-7. HRU 2, 1984

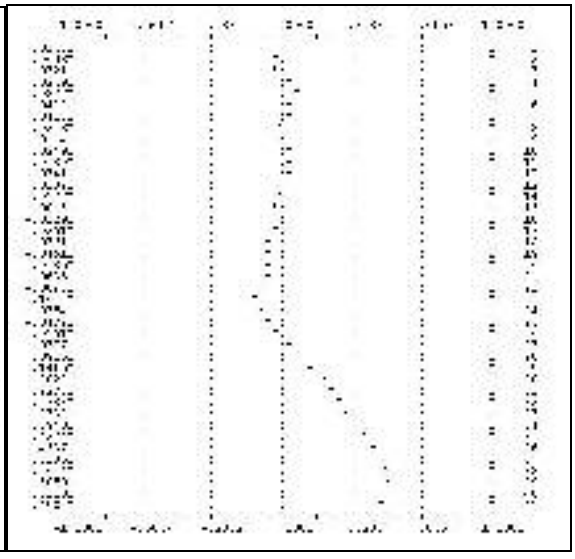


Figure B-8. HRU 2, 1985

HRU 3

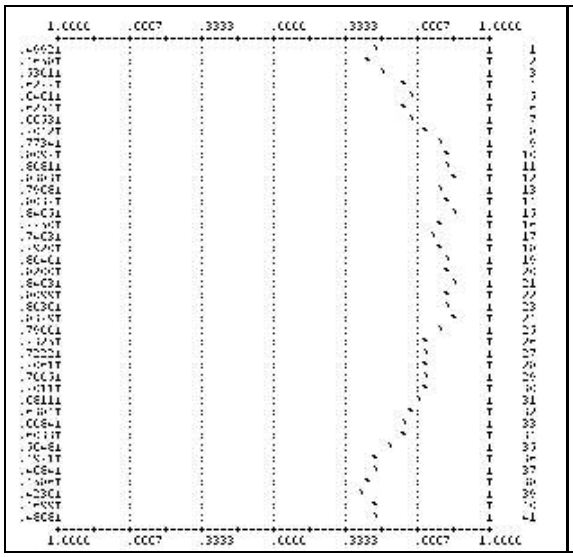


Figure B-9. HRU 3, 1982

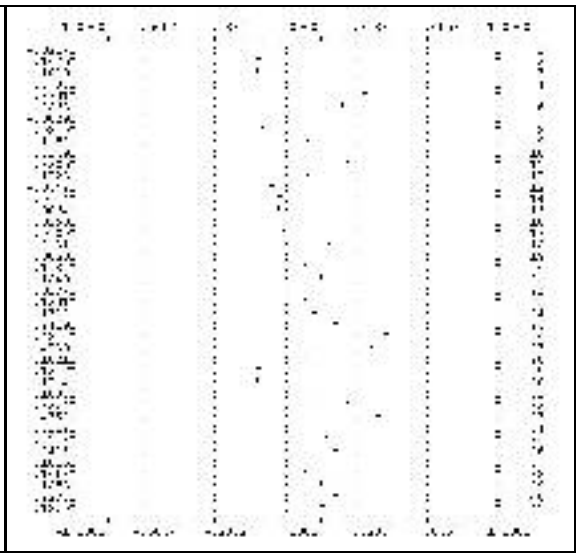


Figure B-10. HRU 3, 1983

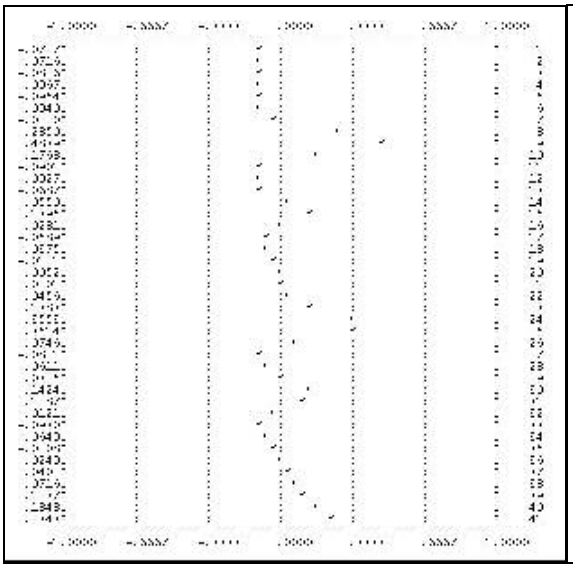


Figure B-11. HRU 3, 1984

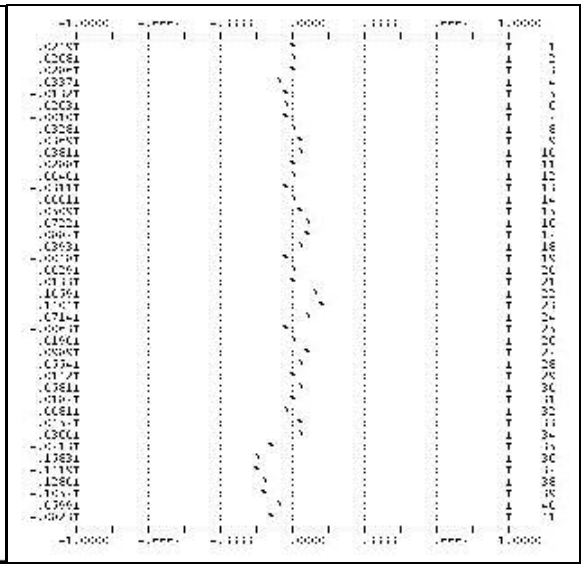


Figure B-12. HRU 3, 1985

HRU 4

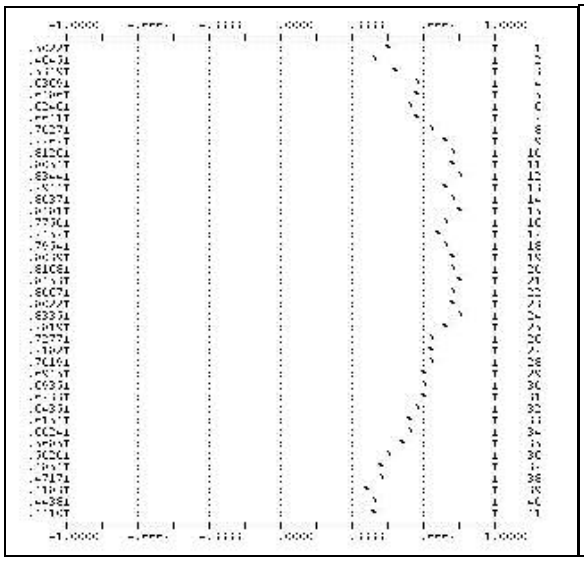


Figure B-13. HRU 4, 1982

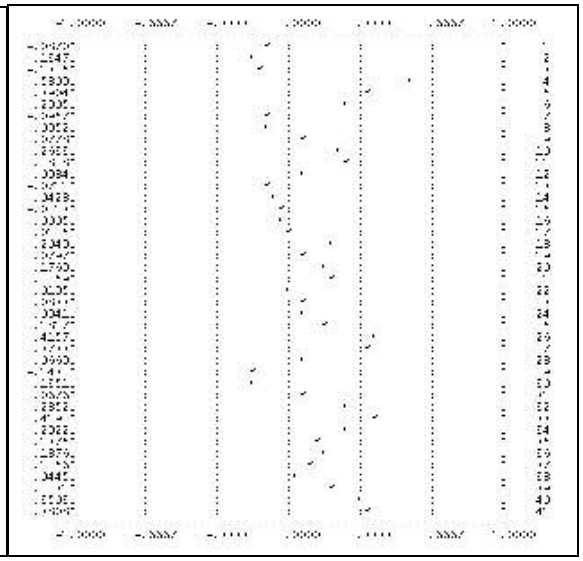


Figure B-14. HRU 4, 1983

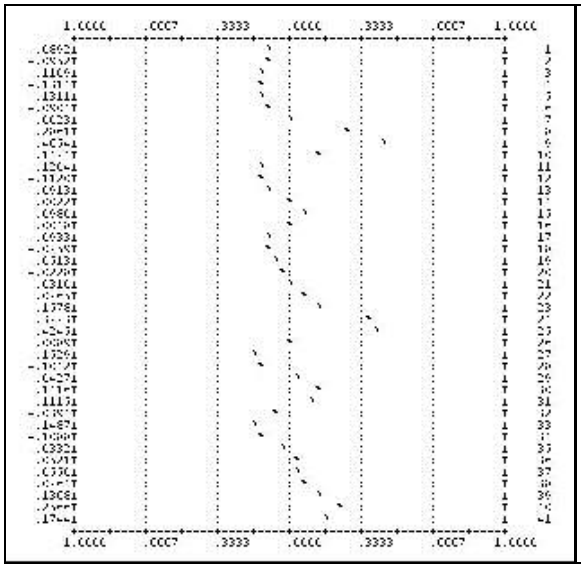


Figure B-15. HRU 4, 1984

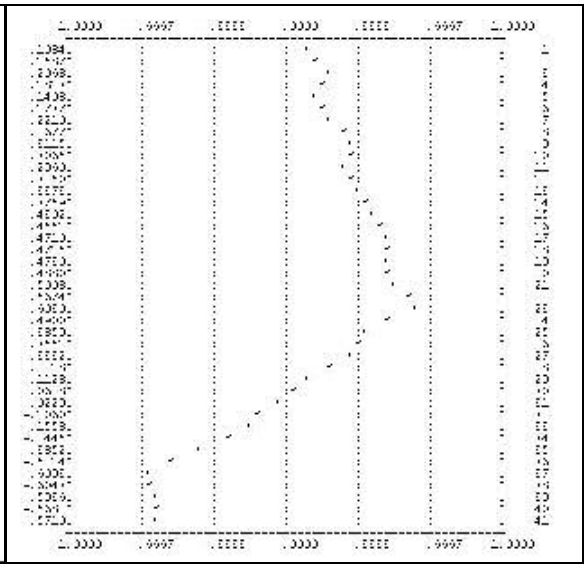


Figure B-16. HRU 4, 1985

HRU 5

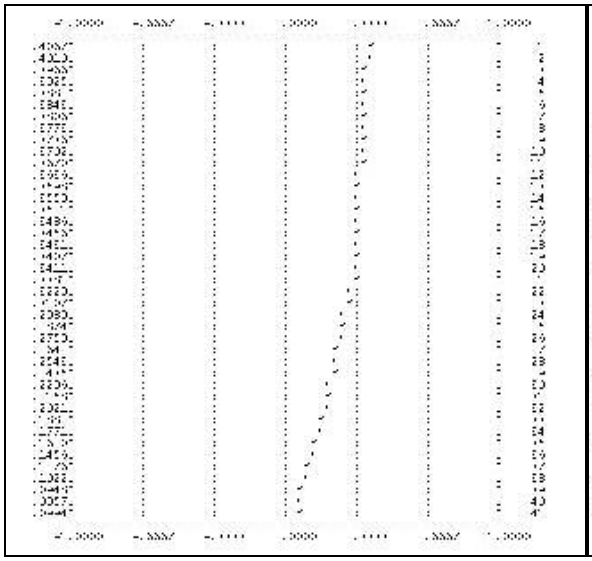


Figure B-17. HRU 5, 1982

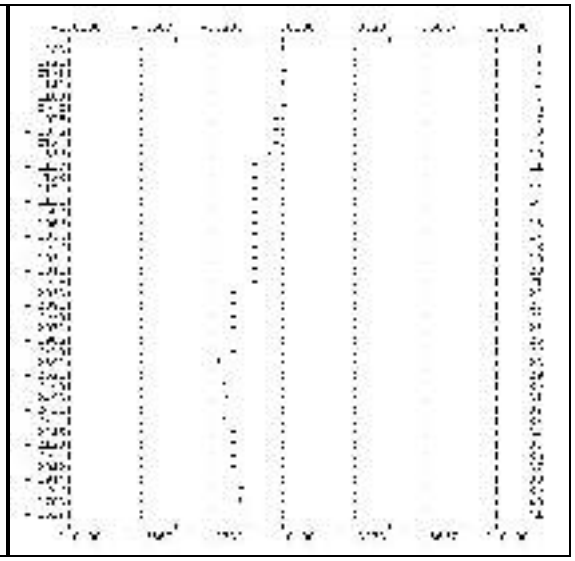


Figure B-18. HRU 5, 1983

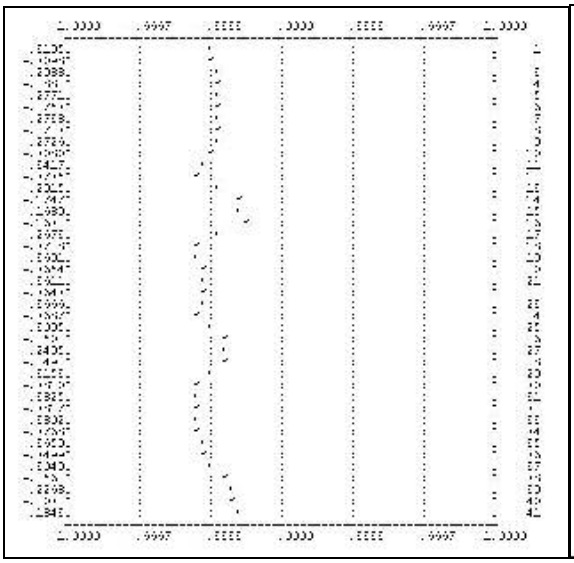


Figure B-19. HRU 5, 1984

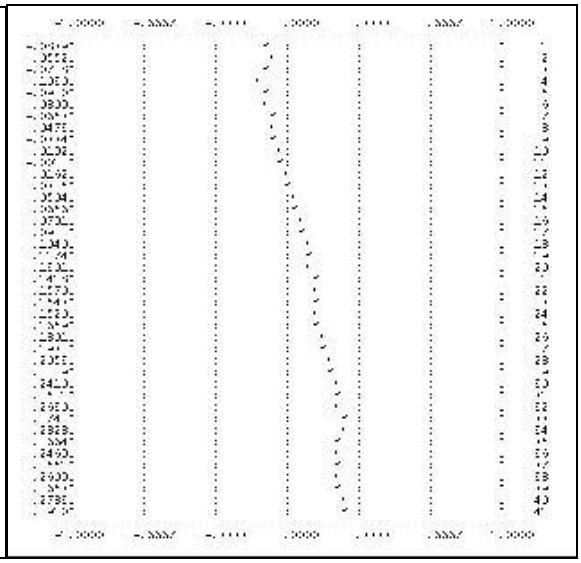


Figure B-20. HRU 5, 1985

HRU 6

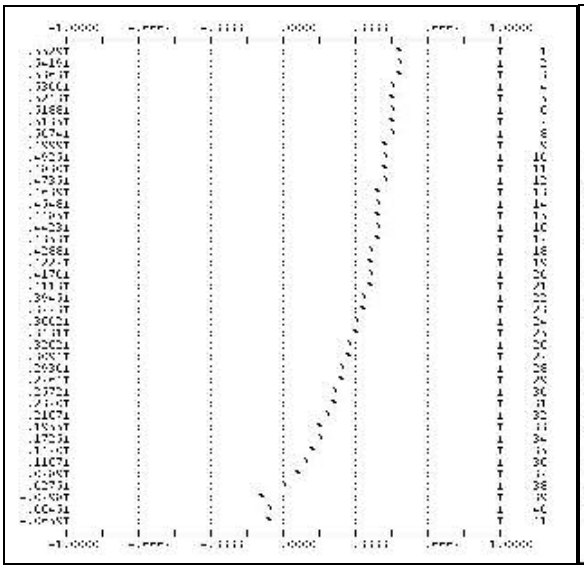


Figure B-21. HRU 6, 1982

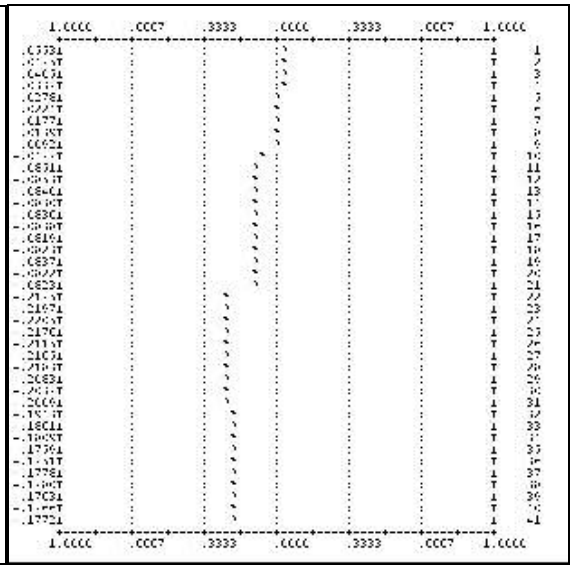


Figure B-22. HRU 6, 1983

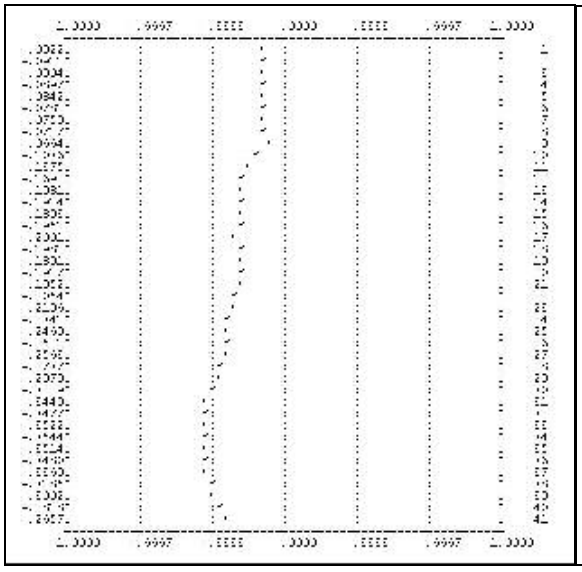


Figure B-23. HRU 6, 1984

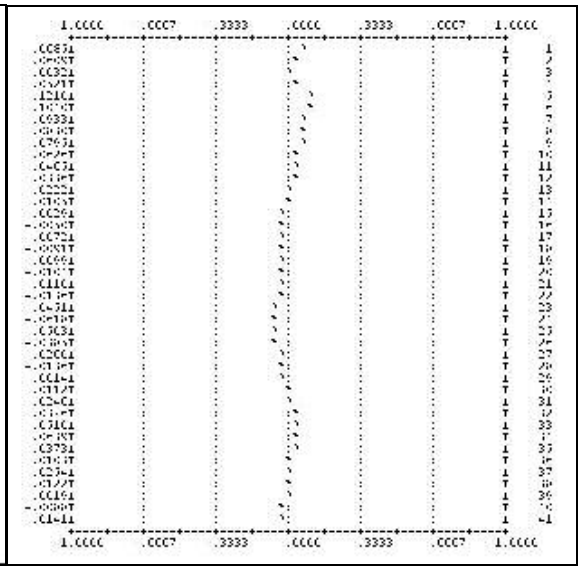


Figure B-24. HRU 6, 1985

HRU 7

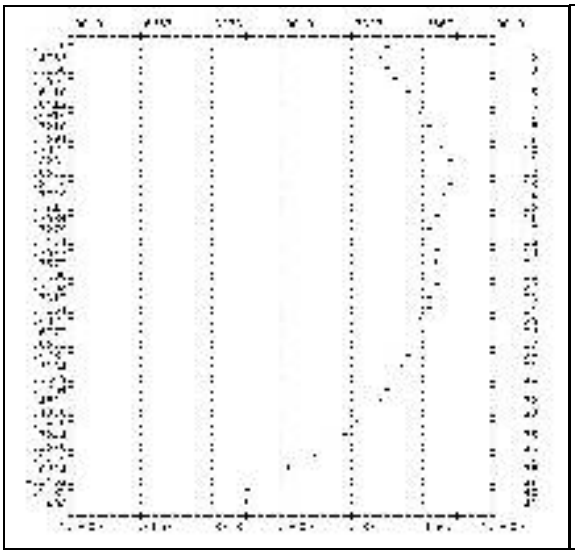


Figure B-25. HRU 7, 1982

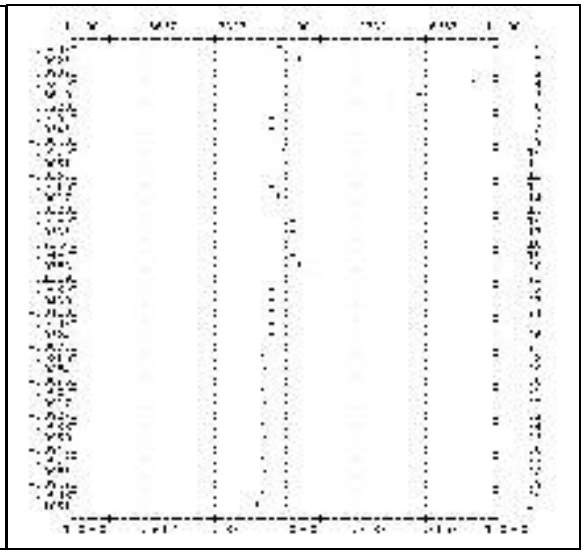


Figure B-26. HRU 7, 1983

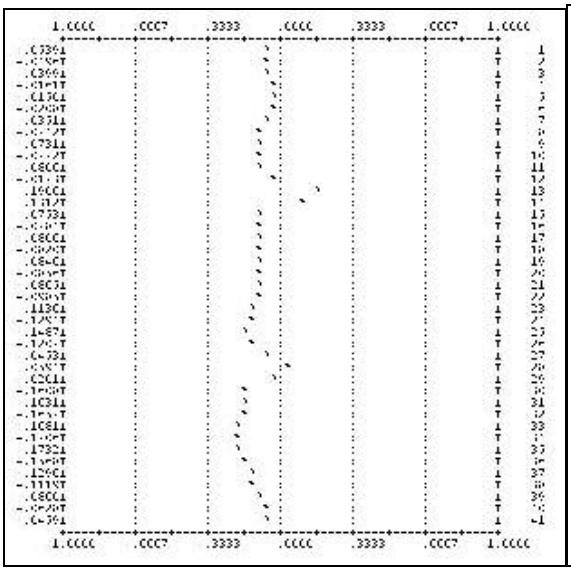


Figure B-27. HRU 7, 1984

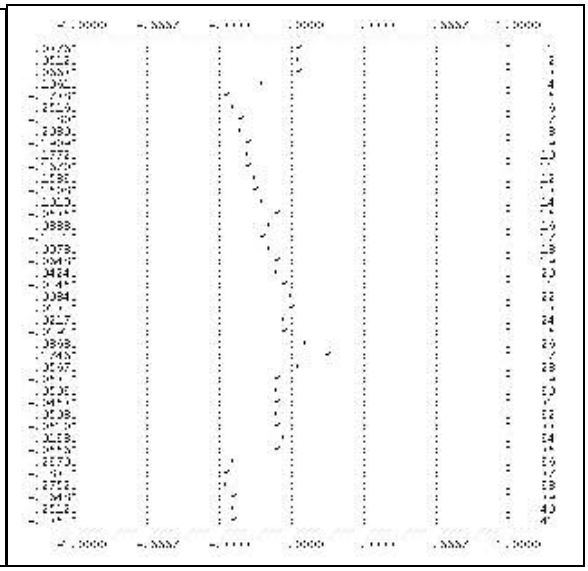


Figure B-28. HRU 7, 1985

HRU 8

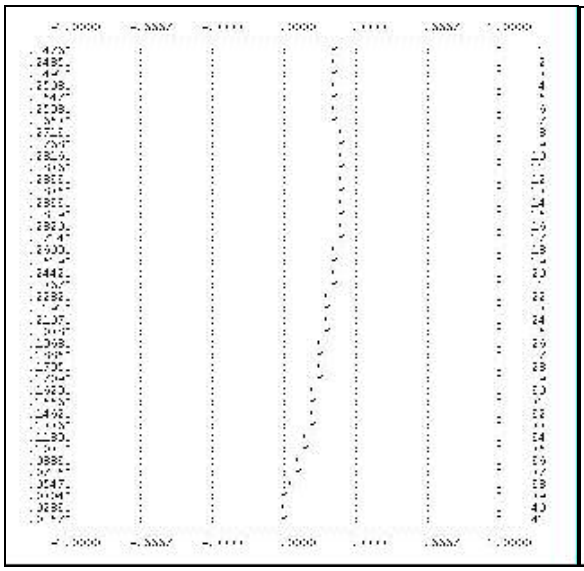


Figure B-29. HRU 8, 1982

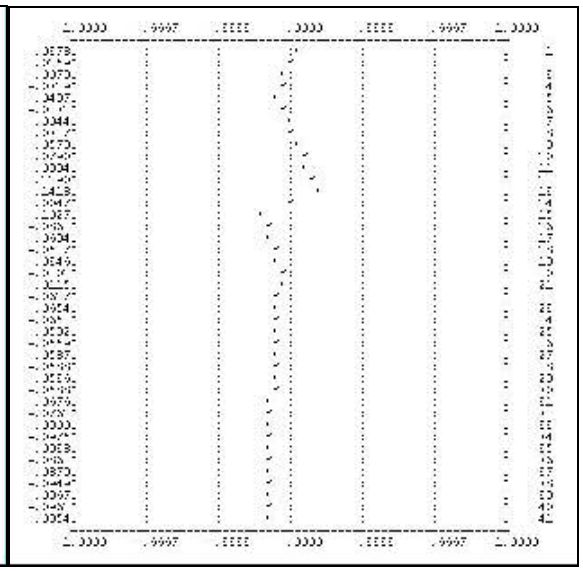


Figure B-30. HRU 8, 1983

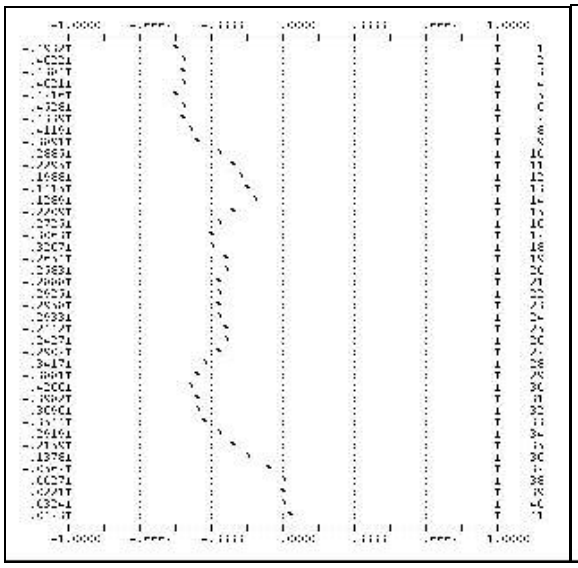


Figure B-31. HRU 8, 1984

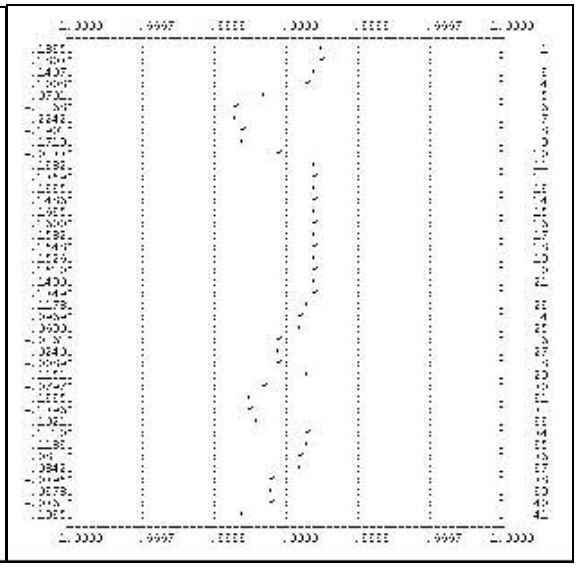


Figure B-32. HRU 8, 1985

HRU 9

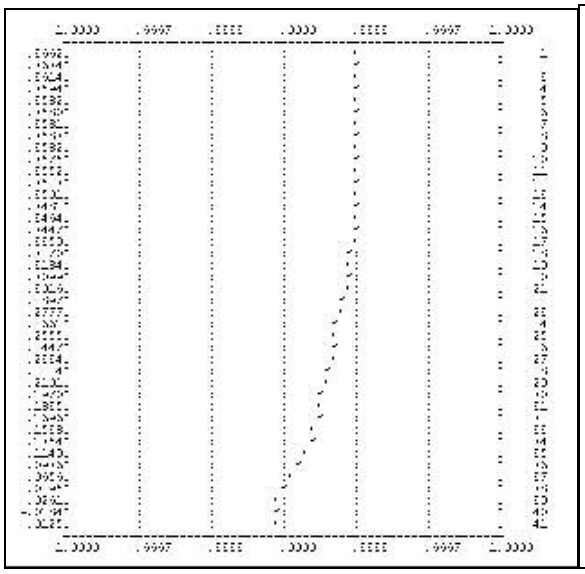


Figure B-33. HRU 9, 1982

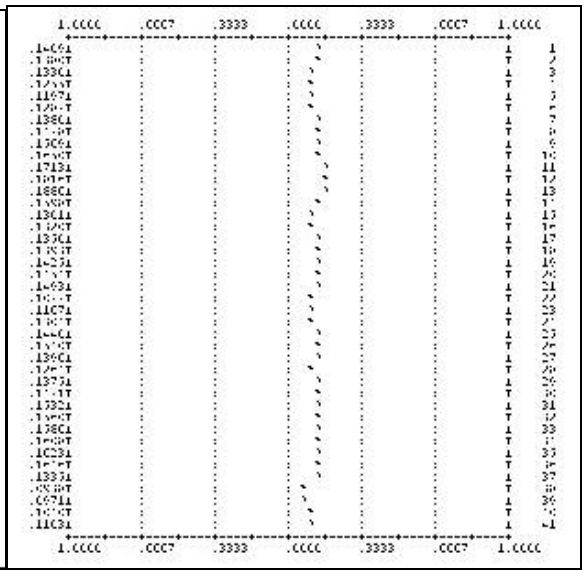


Figure B-34. HRU 9, 1983

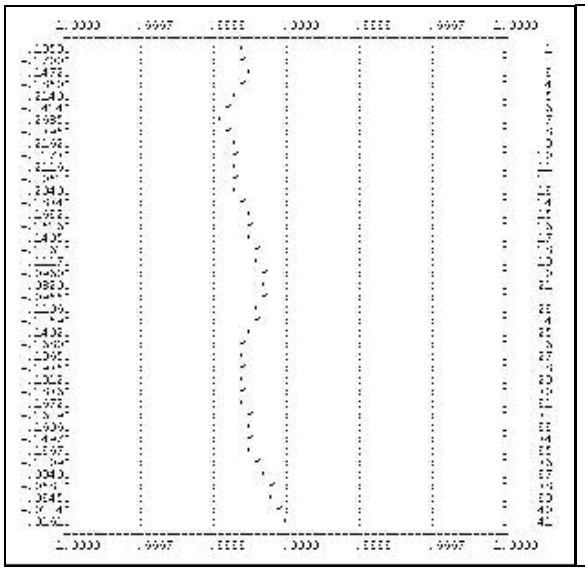


Figure B-35. HRU 9, 1984

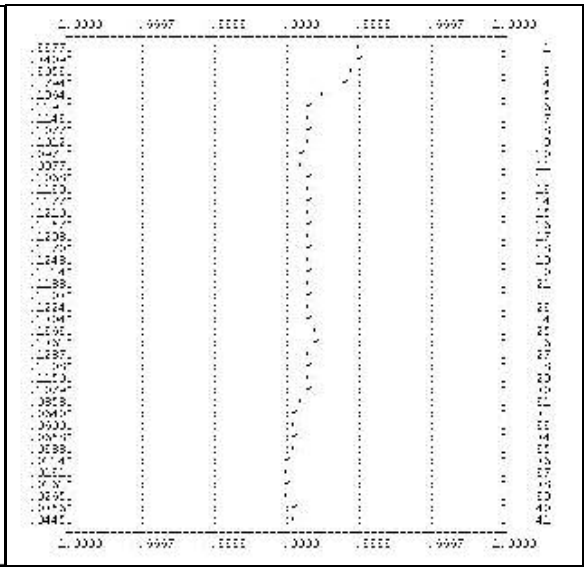


Figure B-36. HRU 9, 1985

HRU 10

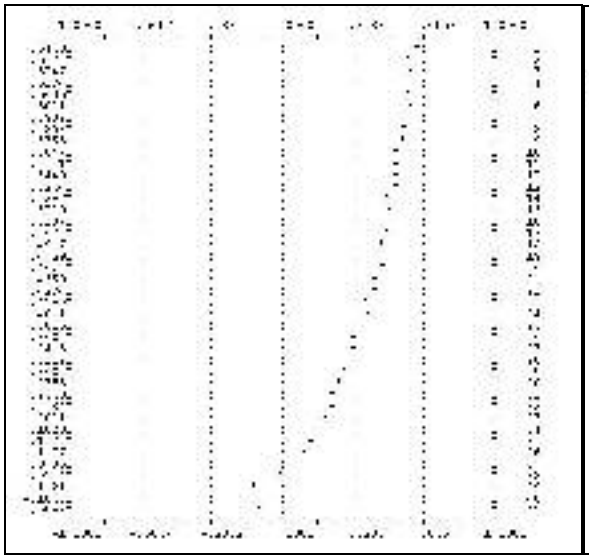


Figure B-37. HRU 10, 1982

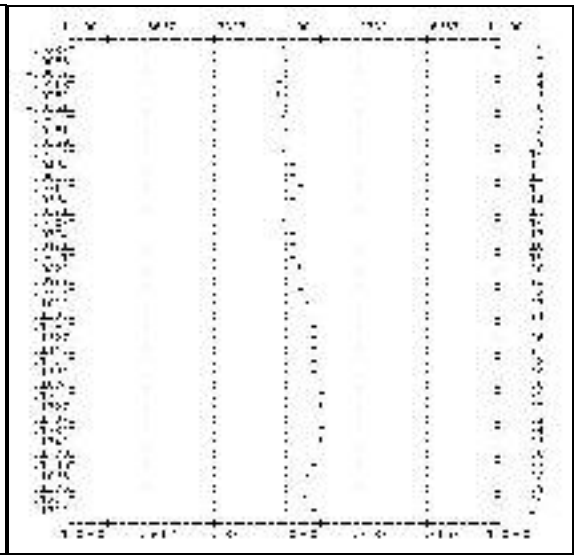


Figure B-38. HRU 10, 1983

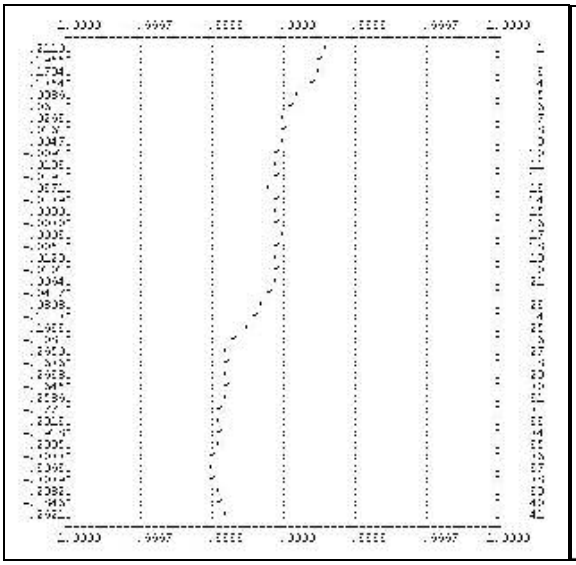


Figure B-39. HRU 10, 1984

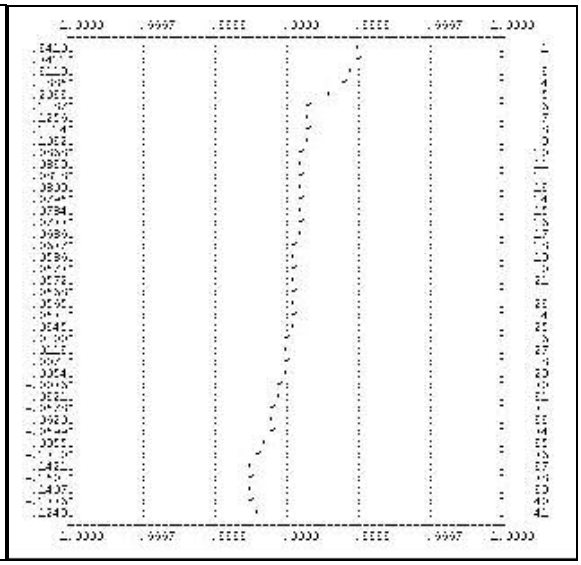


Figure B-40. HRU 10, 1985

HRU 11

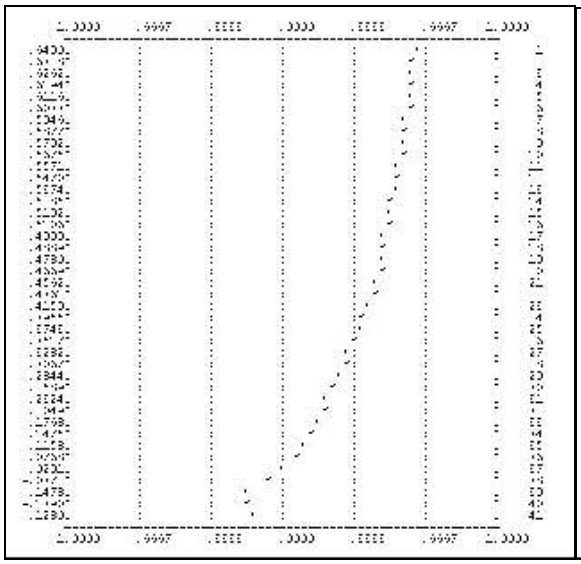


Figure B-41. HRU 11, 1982

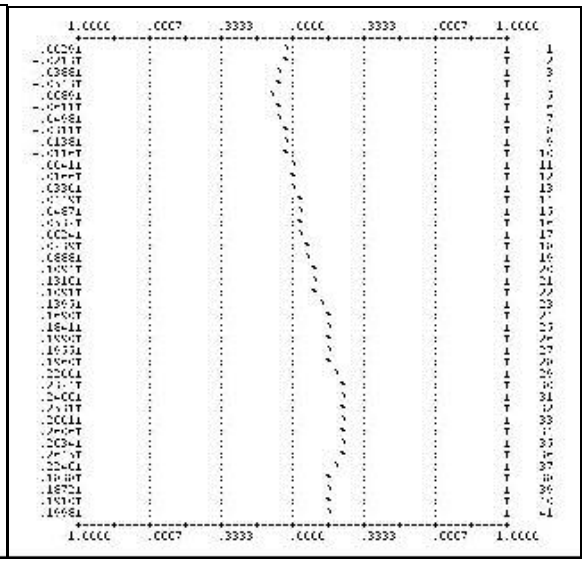


Figure B-42. HRU 11, 1983

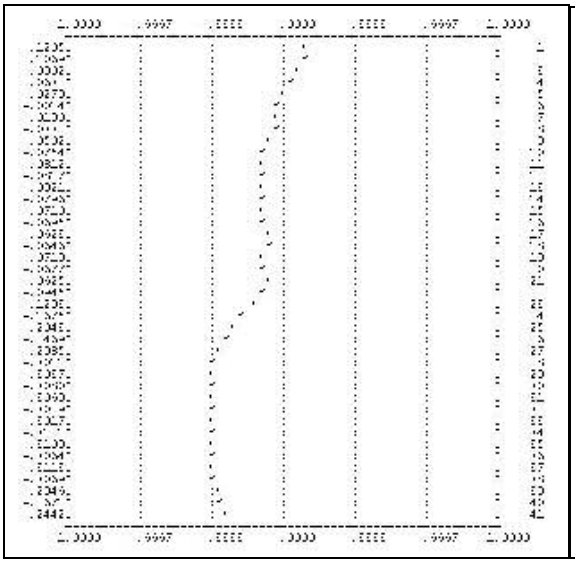


Figure B-43. HRU 11, 1984

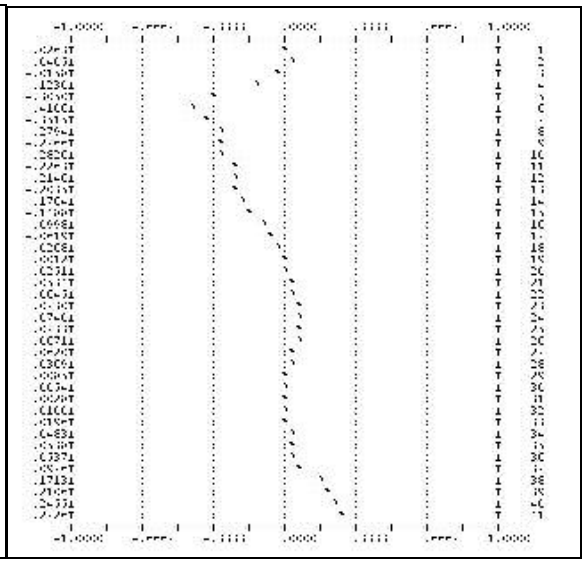


Figure B-44. HRU 11, 1985

HRU 12

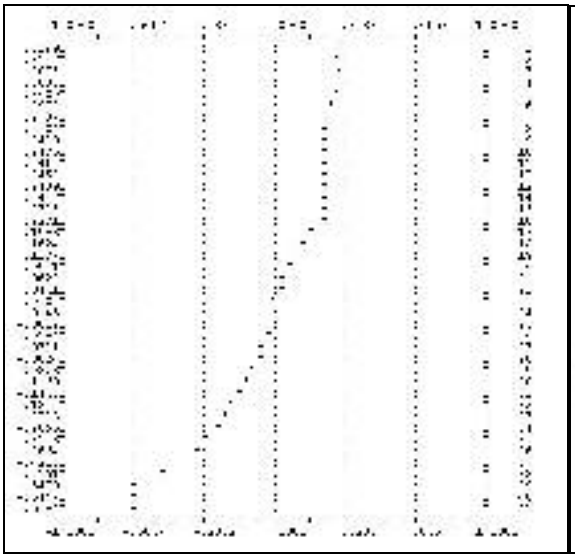


Figure B-45. HRU 12, 1982

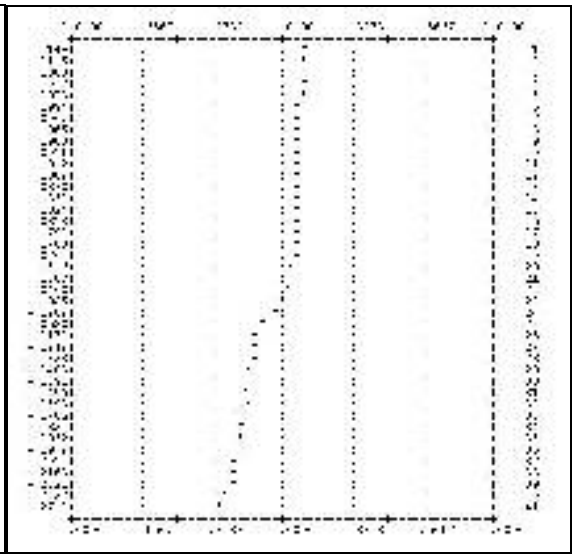


Figure B-46. HRU 12, 1983

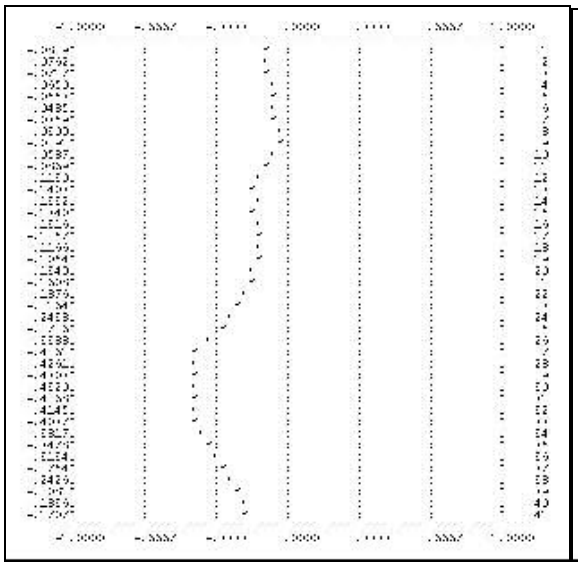


Figure B-47. HRU 12, 1984

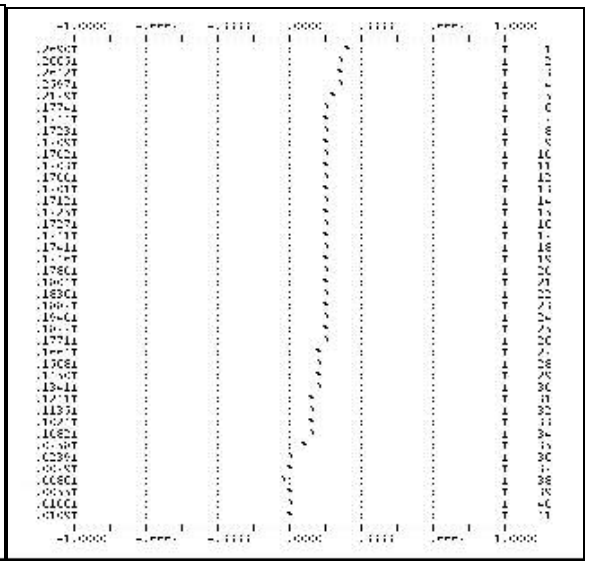


Figure B-48. HRU 12, 1985

HRU 13

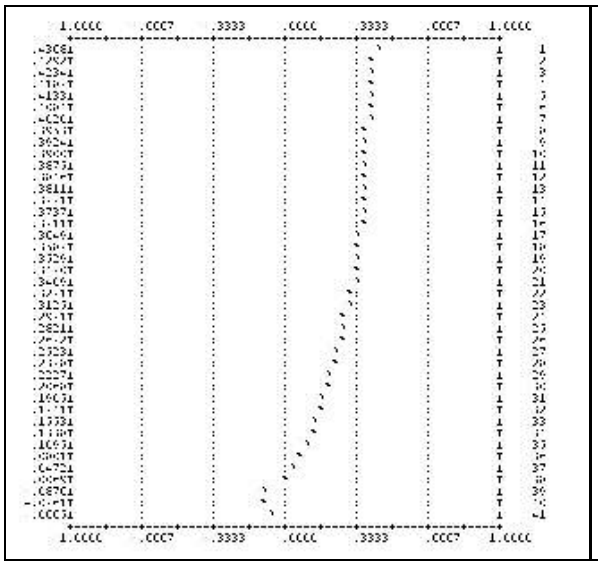


Figure B-49. HRU 13, 1982

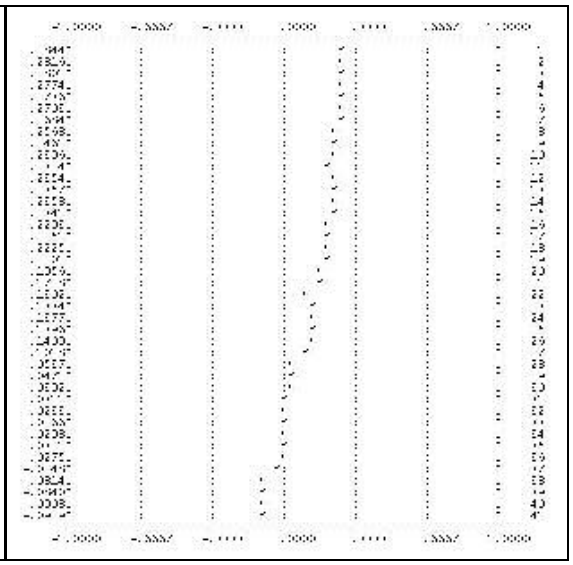


Figure B-50. HRU 13, 1983

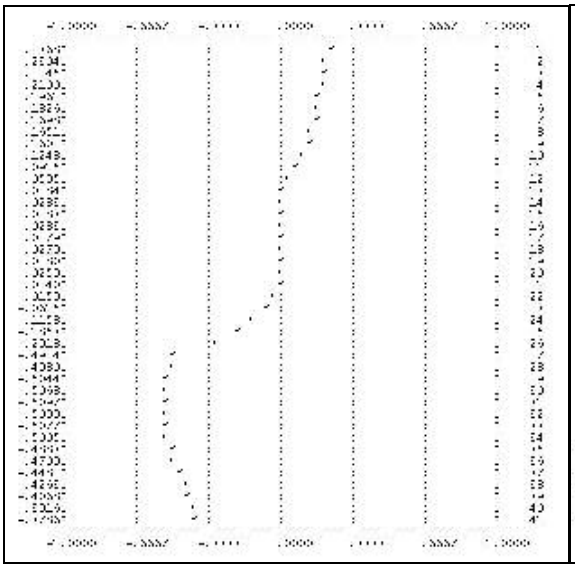


Figure B-51. HRU 13, 1984

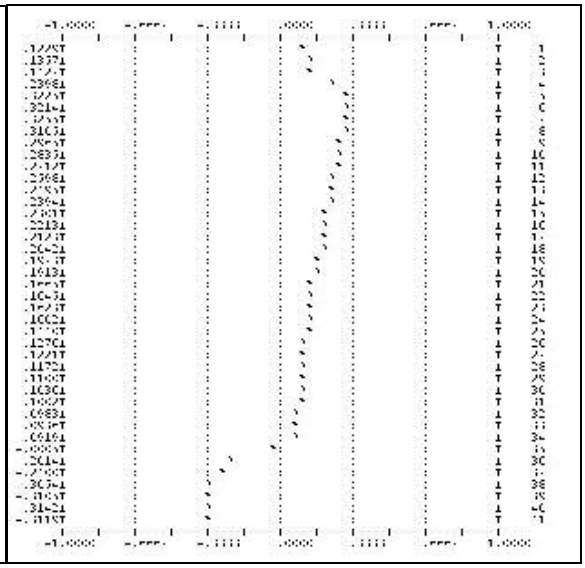


Figure B-52. HRU 13, 1985

HRU 14

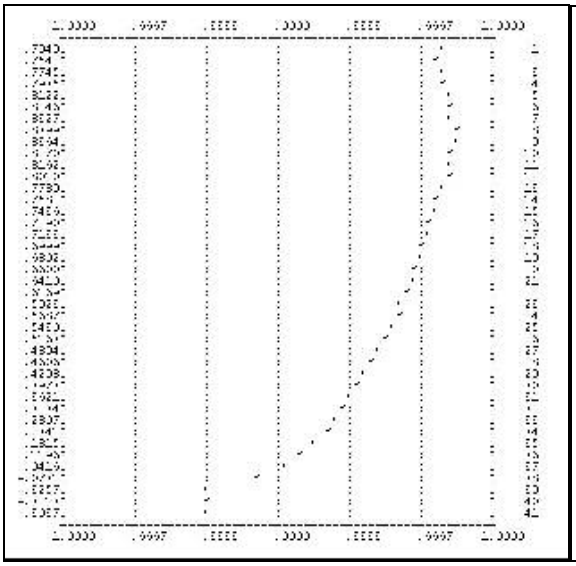


Figure B-53. HRU 14, 1982

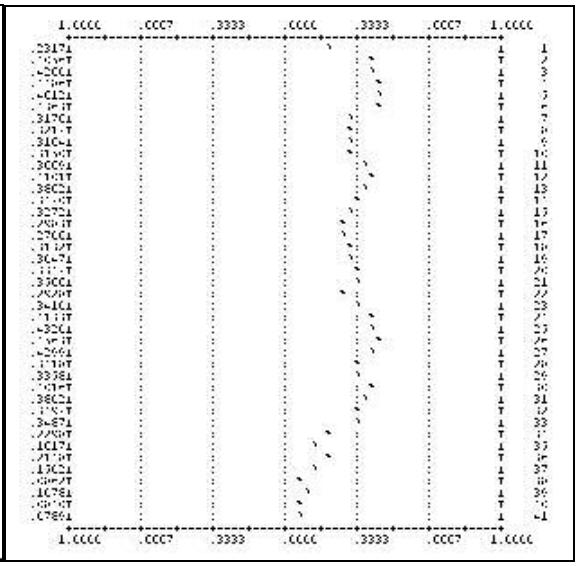


Figure B-54. HRU 14, 1983

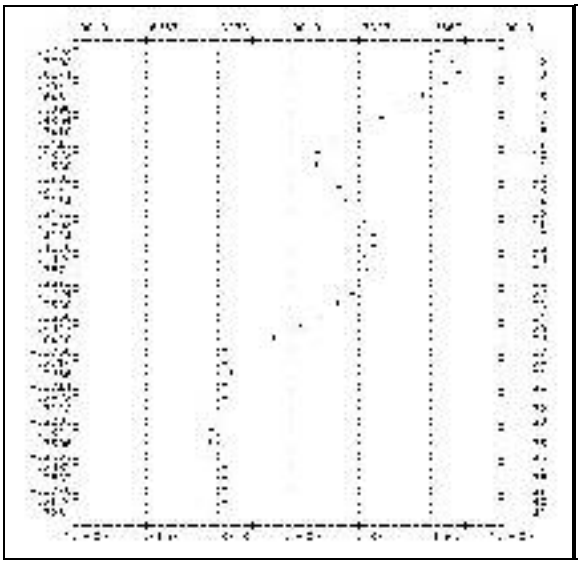


Figure B-55. HRU 14, 1984

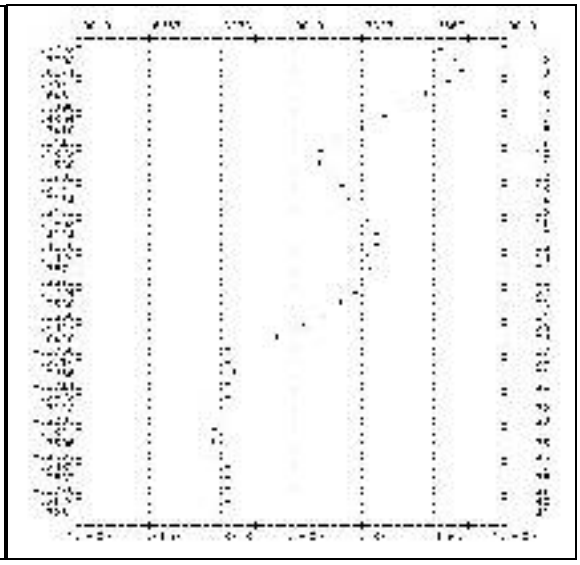


Figure B-56. HRU 14, 1985

HRU 15

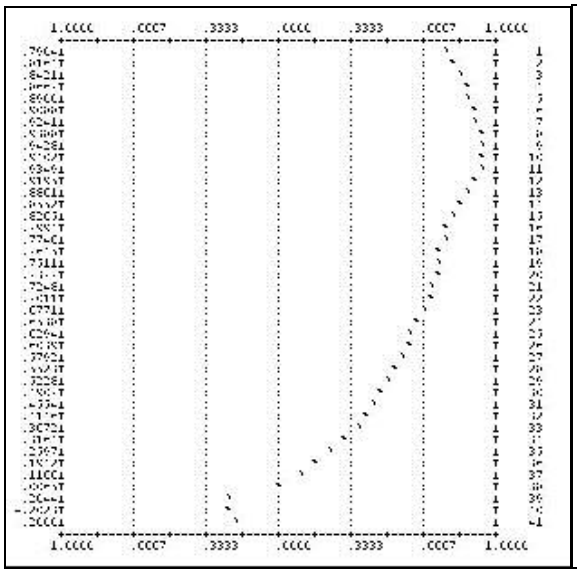
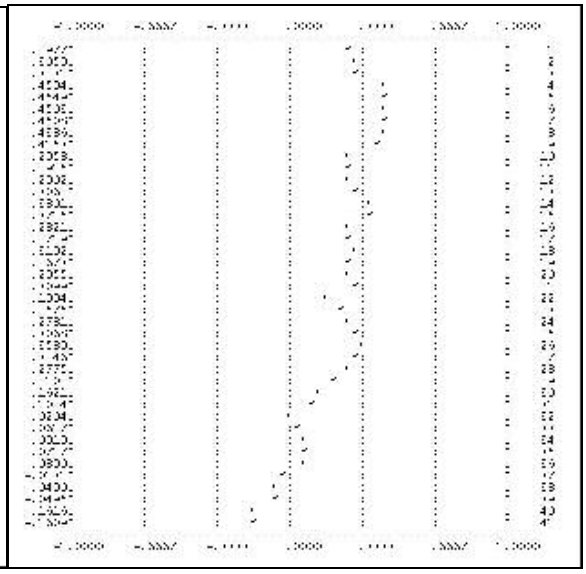


Figure B-57. HRU 15, 1982



HRU B-58. HRU 15, 1983

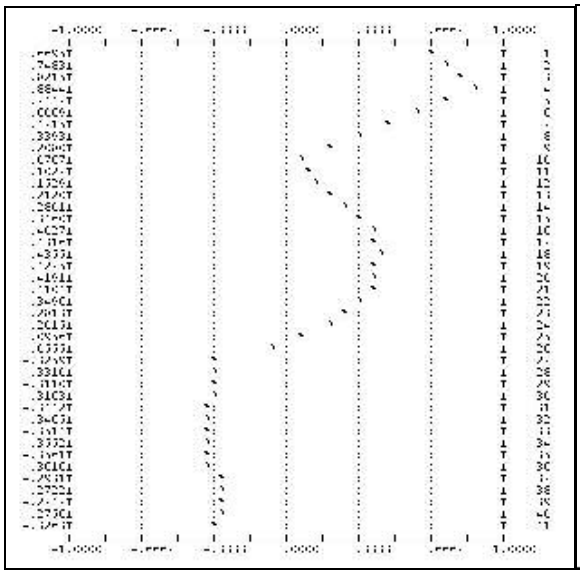


Figure B-59. HRU 15, 1984

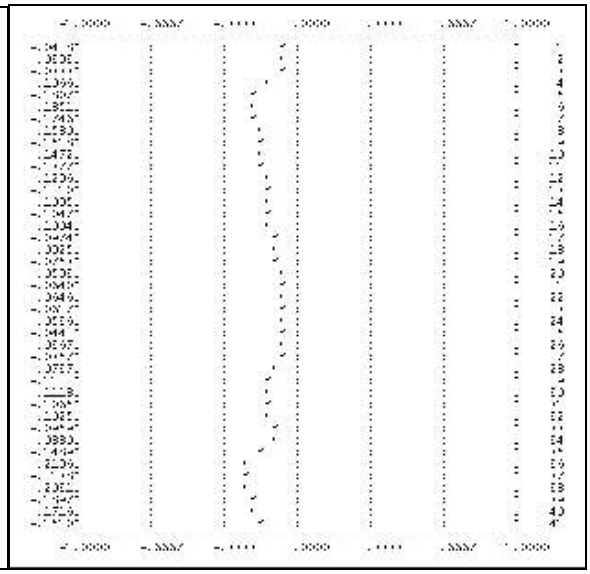


Figure B-60. HRU 15, 1985

HRU 16

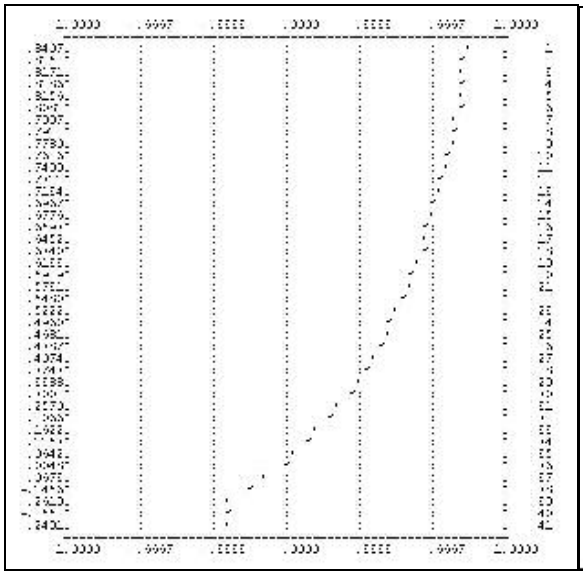


Figure B-61. HRU 16, 1982

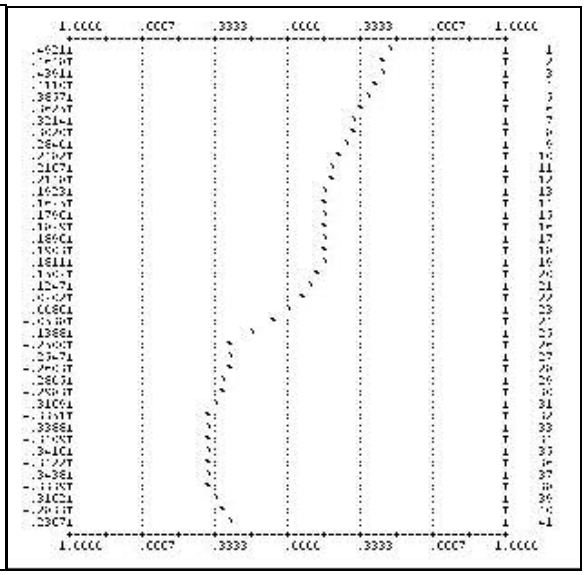


Figure B-62. HRU 16, 1983

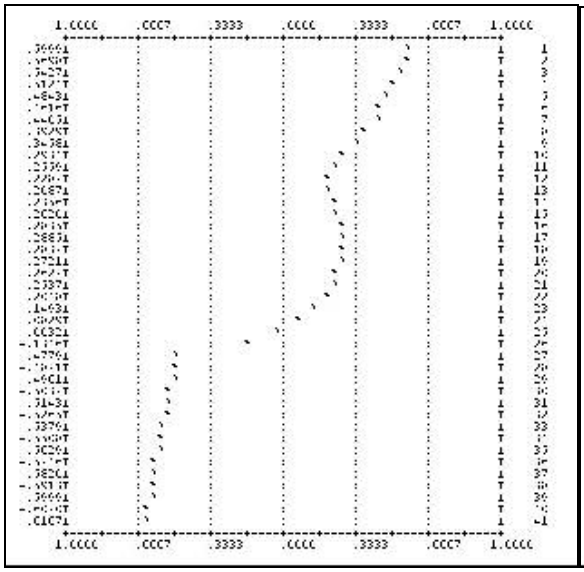


Figure B-63. HRU 16, 1984

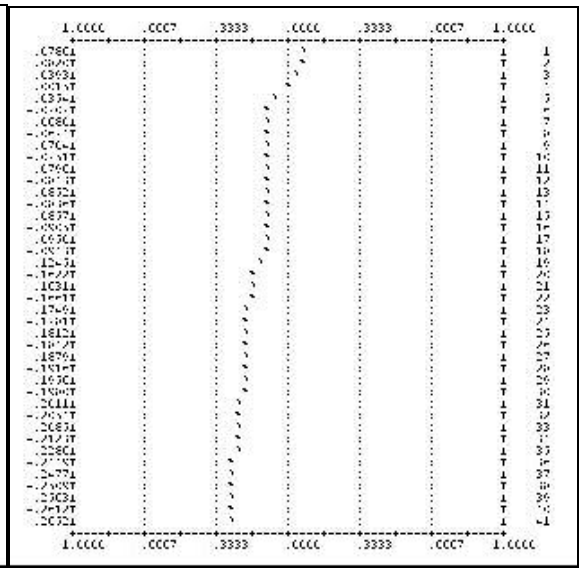


Figure B-64. HRU 16, 1985

HRU 17

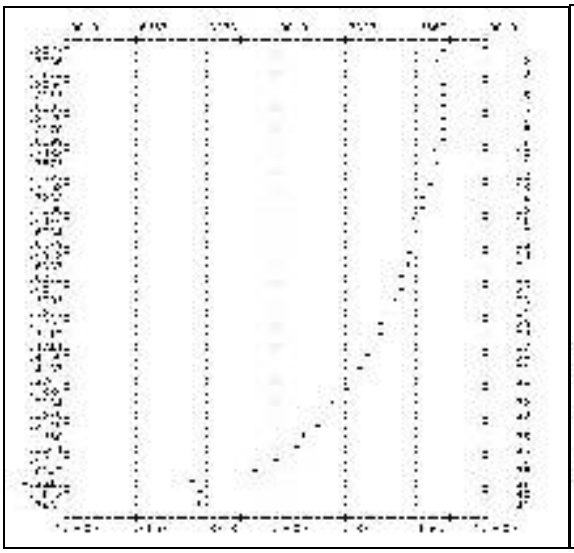


Figure B-65. HRU 17, 1982

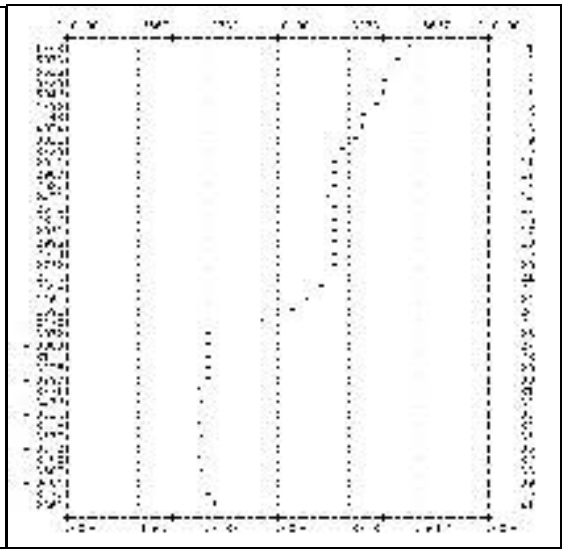


Figure B-66. HRU 17, 1983

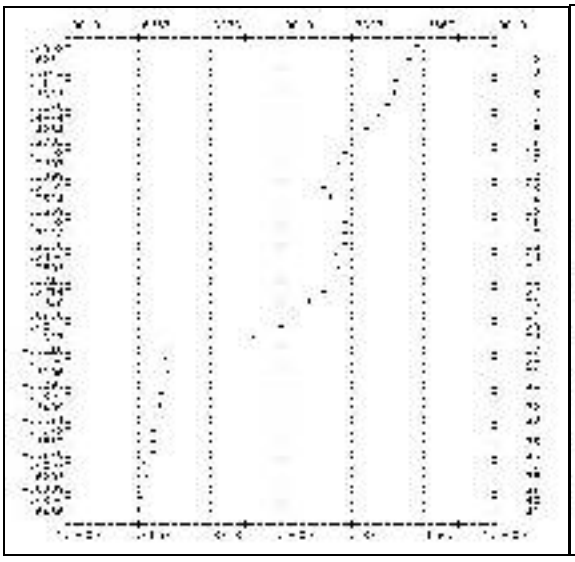


Figure B-67. HRU 17, 1984

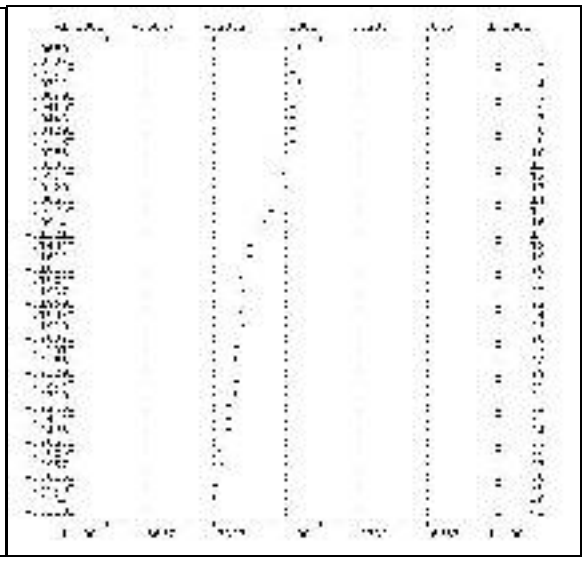


Figure B-68. HRU 17, 1985

HRU 18

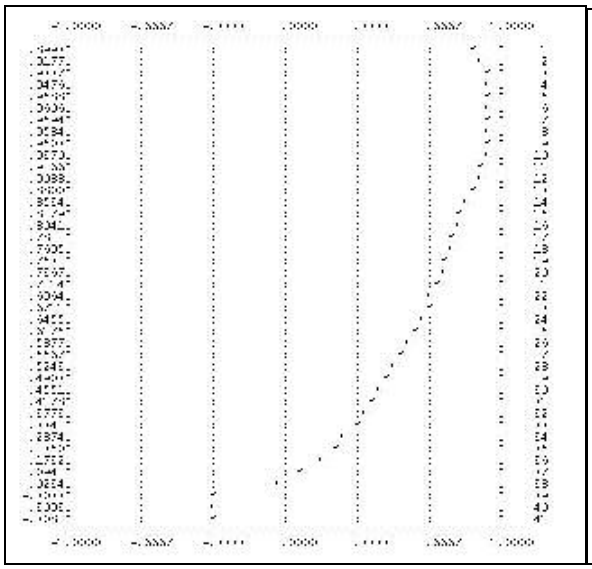


Figure B-69. HRU 18, 1982

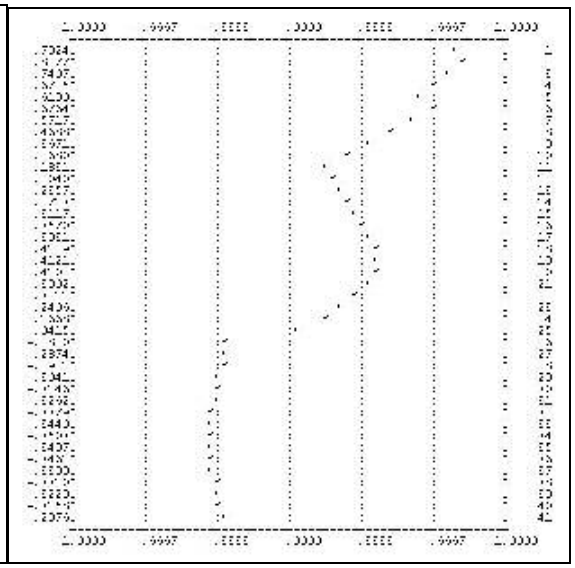


Figure B-70. HRU 18, 1983

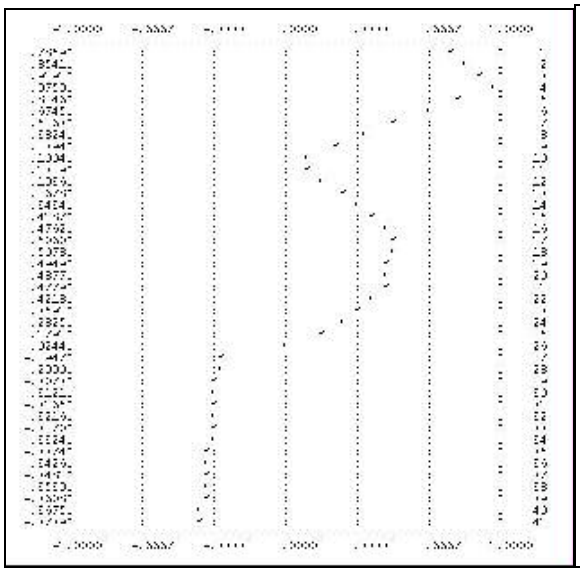


Figure B-71. HRU 18, 1984

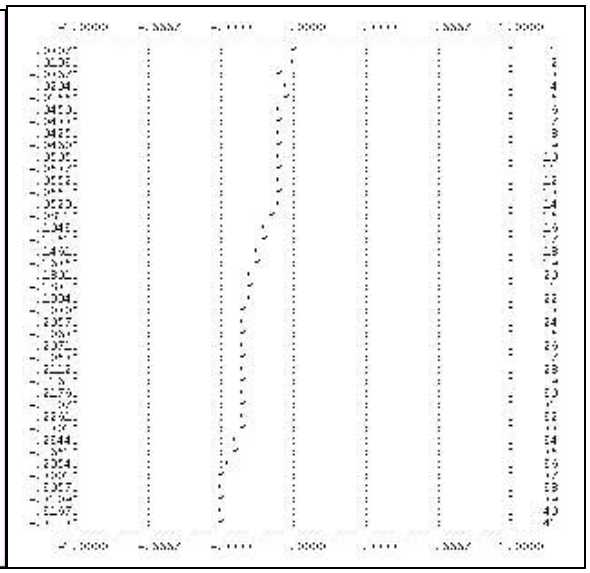


Figure B-72. HRU 18, 1985

HRU 19

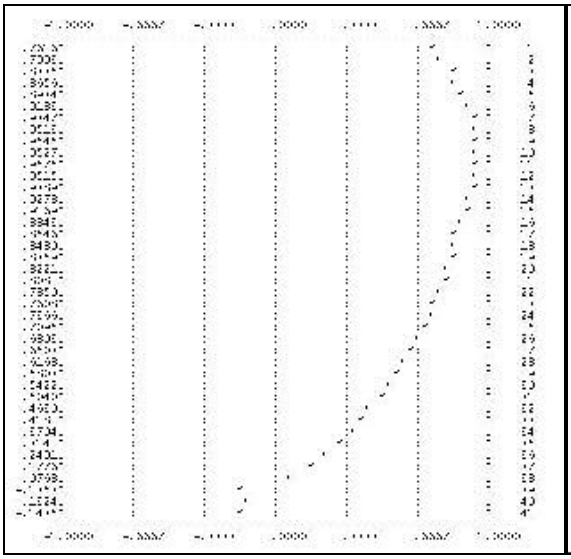


Figure B-73. HRU 19, 1982

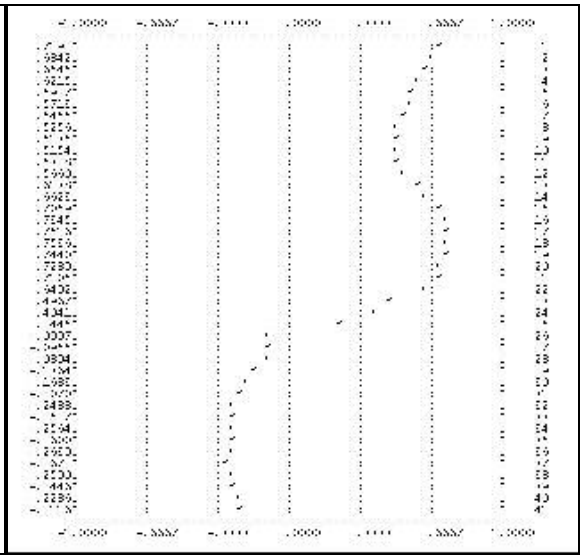


Figure B-74. HRU 19, 1983

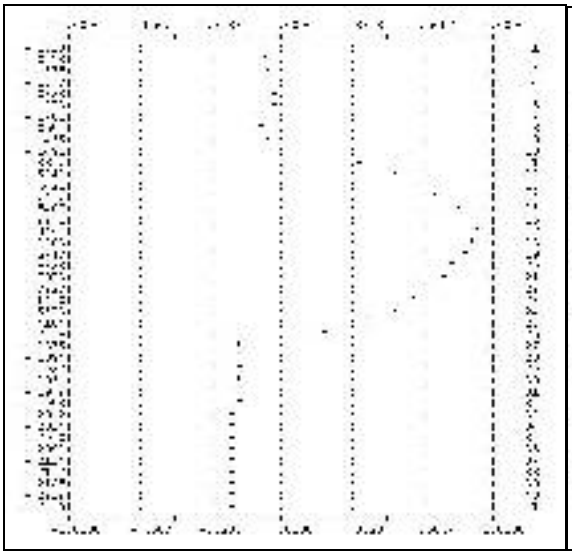


Figure B-75. HRU 19, 1984

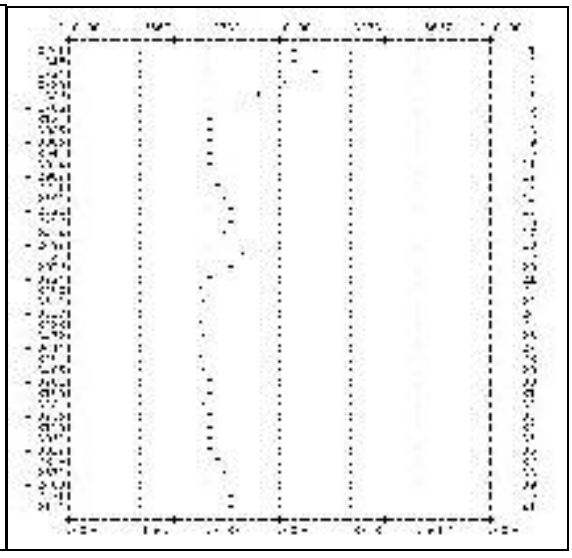


Figure B-76. HRU 19, 1985

APPENDIX C

Experiments on simulation of SCA (Section 5.3.2)

Assigned transition probabilities

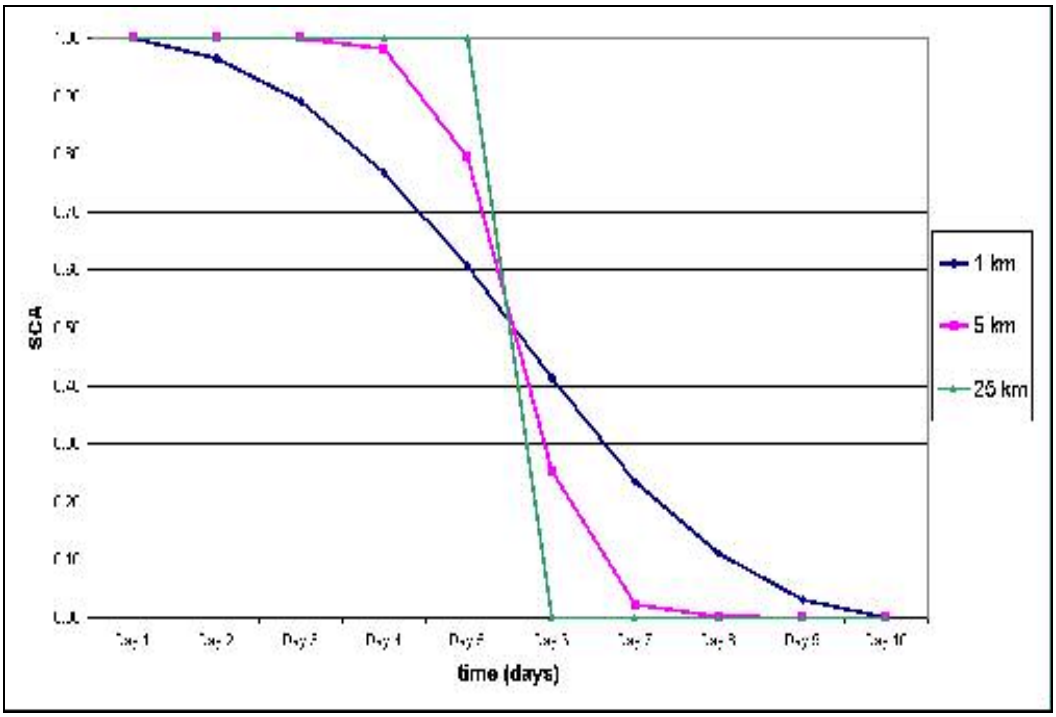
P_{11} : Neighboring pixel being snowcovered given primary pixel is snowcovered

P_{00} : Neighboring pixel snow-free given primary pixel is snow-free

Experiment 1

Simulated Sigmoidal reduction in SCA

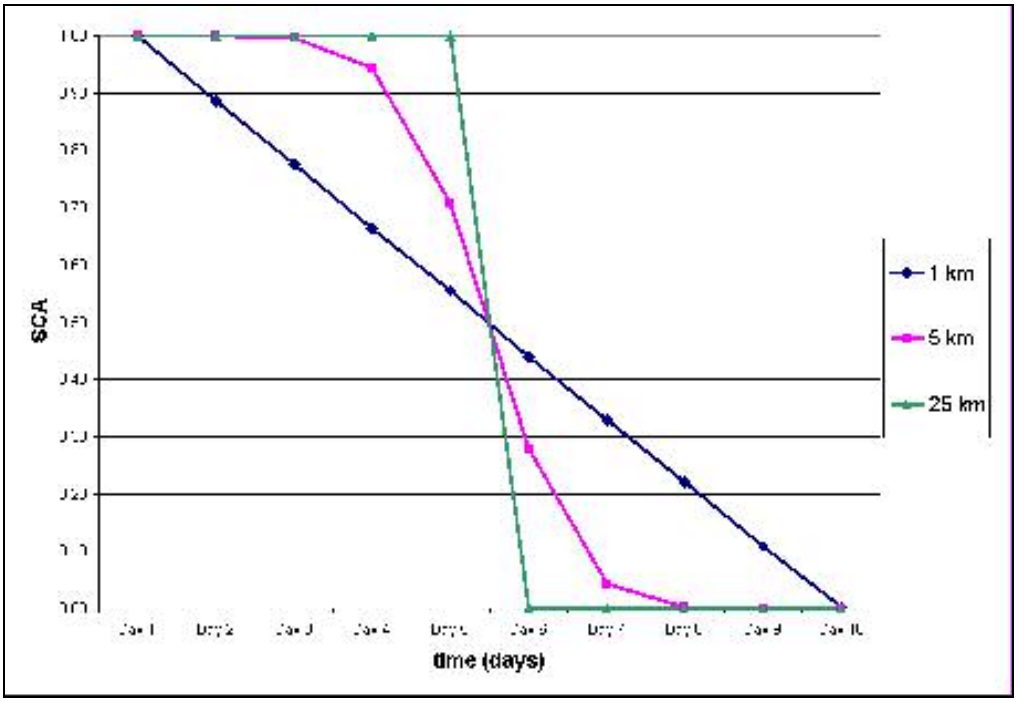
			Theoretical	Simulated Snow	Simulated Snow	Simulated Snow
	P(11)	P(00)	SCA fraction	1 km	5 km	25 km
Date 1	0.999	0.01	0.999	1.00	1.00	1.00
Date 2	0.97	0.18	0.965	0.97	1.00	1.00
Date 3	0.92	0.35	0.890	0.89	1.00	1.00
Date 4	0.85	0.5	0.769	0.77	0.99	1.00
Date 5	0.75	0.63	0.597	0.60	0.75	1.00
Date 6	0.63	0.75	0.403	0.41	0.26	0.00
Date 7	0.5	0.85	0.231	0.24	0.02	0.00
Date 8	0.35	0.92	0.110	0.11	0.00	0.00
Date 9	0.18	0.97	0.035	0.04	0.00	0.00
Date 10	0.01	0.999	0.001	0.00	0.00	0.00



Experiment 2

Simulated Linear reduction in SCA

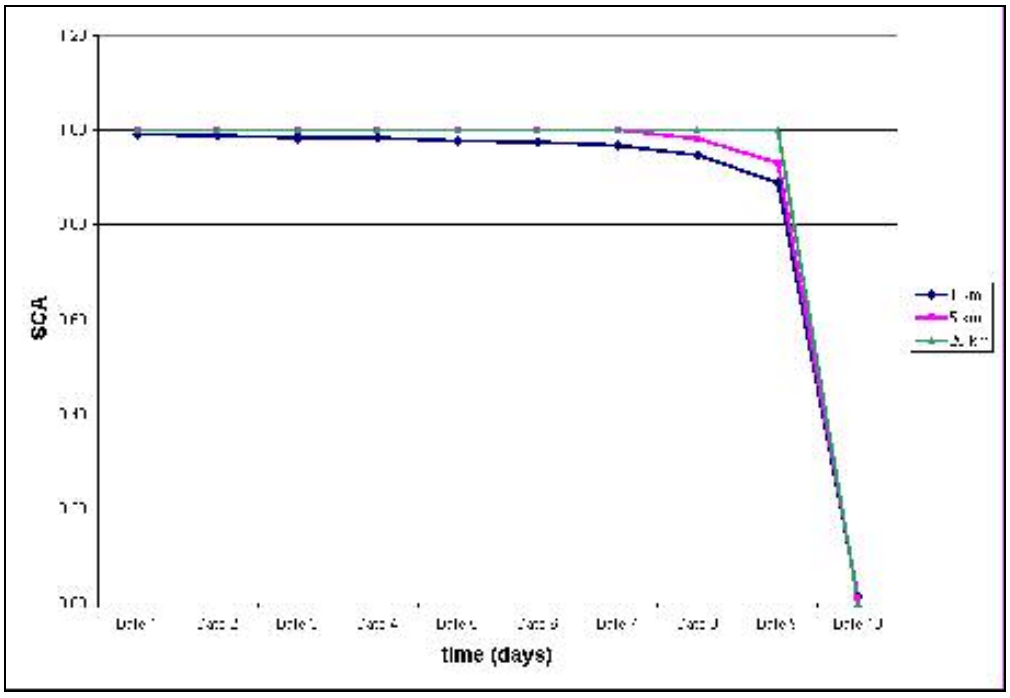
			Theoretical	Calculated Snow Fraction	Calculated Snow Fraction	Calculated Snow Fraction
	P(11)	**P(00)**	**SCA fraction**	**1 km**	**5 km**	**25 km**
Date 1	0.999	0.1	0.999	1.00	1.00	1.00
Date 2	0.9	0.2	0.889	0.89	1.00	1.00
Date 3	0.8	0.3	0.778	0.78	1.00	1.00
Date 4	0.7	0.4	0.667	0.66	0.94	1.00
Date 5	0.6	0.5	0.556	0.55	0.71	1.00
Date 6	0.5	0.6	0.444	0.44	0.28	0.00
Date 7	0.4	0.7	0.333	0.33	0.04	0.00
Date 8	0.3	0.8	0.222	0.22	0.00	0.00
Date 9	0.2	0.9	0.111	0.11	0.00	0.00
Date 10	0.1	0.999	0.001	0.00	0.00	0.00



Experiment 3

Simulation with constant $P_{11} = 0.99$

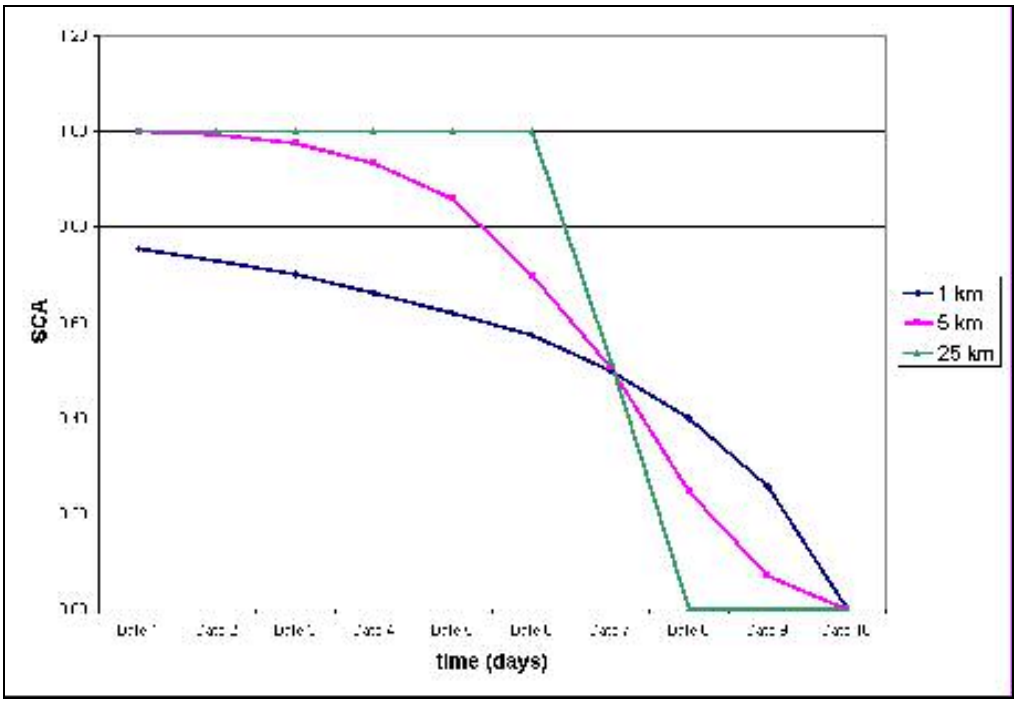
			Theoretical	Calculated Snow Fraction	Calculated Snow Fraction	Calculated Snow Fraction
	P(11)	**P(00)**	**SCA fraction**	**1 km**	**5 km**	**25 km**
Date 1	0.99	0.1	0.989	0.99	1.00	1.00
Date 2	0.99	0.2	0.988	0.99	1.00	1.00
Date 3	0.99	0.3	0.986	0.98	1.00	1.00
Date 4	0.99	0.4	0.984	0.98	1.00	1.00
Date 5	0.99	0.5	0.980	0.98	1.00	1.00
Date 6	0.99	0.6	0.976	0.97	1.00	1.00
Date 7	0.99	0.7	0.968	0.97	1.00	1.00
Date 8	0.99	0.8	0.952	0.95	0.98	1.00
Date 9	0.99	0.9	0.909	0.89	0.93	1.00
Date 10	0.99	0.99	0.500	0.01	0.01	0.00



Experiment 4

Simulation with constant $P_{11} = 0.70$

			Theoretical	Calculated Snow Fraction	Calculated Snow Fraction	Calculated Snow Fraction
	P(11)	P(00)	SCA fraction	1 km	5 km	25 km
Date 1	0.7	0.1	0.750	0.75	1.00	1.00
Date 2	0.7	0.2	0.727	0.73	0.99	1.00
Date 3	0.7	0.3	0.700	0.70	0.98	1.00
Date 4	0.7	0.4	0.667	0.66	0.93	1.00
Date 5	0.7	0.5	0.625	0.62	0.86	1.00
Date 6	0.7	0.6	0.571	0.57	0.70	1.00
Date 7	0.7	0.7	0.500	0.50	0.51	0.52
Date 8	0.7	0.8	0.400	0.40	0.25	0.00
Date 9	0.7	0.9	0.250	0.26	0.07	0.00
Date 10	0.7	0.999	0.003	0.00	0.00	0.00



APPENDIX D

Missing Dates of Coverage

AVHRR 5-km Polar Pathfinder Data Set

1981: 08/22-08/31, 11/29, 11/30

1982: 06/02, 08/11, 09/25, 09/26, 09/29, 11/26, 12/28-12/31

1983: 01/02-01/07, 01/09-01/13, 05/01-05/03, 08/06, 08/17-08/23, 09/21-09/26, 12/12, 12/13, 12/19, 12/20, 12/23-12/26, 12/28-12/31

1984: 01/01-1/10, 01/14, 01/15, 02/18-02/22, 02/24, 02/25, 03/23-03/30, 04/10, 04/14-04/19, 05/04, 06/04-06/06, 11/13, 11/18, 12/06, 12/09

1985: 01/01-01/03, 01/13, 01/17, 01/20, 01/23, 02/08-02/10, 03/16, 04/11, 04/12, 11/20, 12/07, 12/14

Coverage begins from 7/24/1981

REFERENCES

Armstrong, R. L. and M. J. Brodzik, 1995. An earth-gridded SSM/I data set for cryospheric studies and global change monitoring. *Advances in Space Research* 10:155-163.

Bloschl, G., D. Gutknecht, and, R. Kirnbauer. 1991. Distributed snowmelt simulations in an alpine catchment. 2. Parameter study and model predictions. Water Resources Research, 27:3181-3188

Brodzik, Mary J. 2003. All About EASE-Grid http://nsidc.org/data/ease/ease_grid.html (accessed: June 2003)

Brubaker, K.L. and E.G. Josberger. 2001. Fresh water Fluxes of the Ob River Basin, Russia: Contribution of Snowpack in Complex Terrain, Research proposal to the NASA Office of Earth Science.

Brubaker, K.L., M. Jasinski, A.T.C. Chang and E.G. Josberger, 2000. Interpolating sparse surface measurements for calibration and validation of satellite derived snow water equivalent in Russian Siberia, *Remote Sensing and Hydrology 2000* (Proceedings of a Symposium Held at Santa Fe, NM, USA)

CIRES 2002 Annual Report: Global Snow Cover Products Derived from Satellite Remote Sensing: Analysis and Validation of Long-Term Climatologies and Development of Near Real-Time Products from Current Satellite Sensors.

Chang, A.T.C., J.L. Foster, and D.K. Hall. 1987. Nimbus-7 derived global snow cover parameters, *Annals of Glaciology*, 9, 39-44.

Chang, A.T.C., J.L. Foster, D. Hall, A. Rango, and B. Hartline. 1982. Snow water equivalence determination by microwave radiometry. *Cold Regions Science and Technology*, 5, 259-267.

DeFries, R., Hansen, M., Townshend, J. R. G. and Sohlberg, R., 1998, Global land cover classifications at 8 km spatial resolution: The use of training data derived from Landsat imagery in decision tree classifiers, *International Journal of Remote Sensing;* 19 (16): 3141-3168.

Dozier, J., Snow reflectance from Landsat-4 thematic mapper. *IEEE Trans. Geosci. and Rem. Sens. GE-22(3):*323-328, 1984

Fowler C., J. Maslanik, T. Haran, T. Scambos, J. Key, and W. Emery. 2000. AVHRR Polar Pathfinder Twice-Daily 5 km EASE-Grid Composites. Boulder, CO, USA: National Snow and Ice Data Center. Digital media.

GLCF. 2003. Global Land Cover Facility. University of Maryland Institute for Advanced Computer Studies (UMIACS) (accessed: July 2003)

Goodison, B.E. 1989. Determination of areal snow water equivalent on the Canadian prairies using passive microwave satellite data. *IGARSS '89 Proceedings* 3:1243-1246.

Hansen, M., DeFries, R., Townshend, J. R. G. and Sohlberg, R., 2000, Global land cover classification at 1km resolution using a decision tree classifier, International Journal of Remote Sensing. 21: 1331-1365.

Hartman, G. 2000. Effects of GIS data resolution on hydrologic modeling. Masters thesis. University of Maryland, College Park.

Kirnbauer, R., G. Bloschl, and D. Gutknecht. 1994. Entering the era of distributed snow models. Nordic Hydrology, 25:1-24.

Klein, A.G., D.K. Hall and G. A. Riggs, 1998. Global Snow Cover Monitoring using MODIS, *International Symposium on Remote Sensing of Environment*, pp 363-366.

Knowles, Kenneth W. 1993. Points, pixels, grids, and cells – a mapping and gridding primer. Unpublished report to the National Snow and Ice Data Center, Boulder, CO.

Liston, G.E. 1999. Interrelationships among snow distribution, snowmelt, and snow cover depletion: Implications for Atmospheric, hydrologic, and ecologic modeling. *Journal of Applied Meteorology* 38(10):1474-1487

Maidment D., K. Asante, F. Olivera, J. Famiglietti. 2003. Terrain Analysis for Global Runoff Routing (accessed: November 2003). http://www.crwr.utexas.edu/gis/gishydro99/global/nsfhyd.htm#preprocessing

Menoes, M. 2003. Estimating the spatial and temporal distribution of snow water equivalent within a watershed. PhD Thesis. University of Maryland, College Park.

McCuen, R.H. 2003. “TSA.EXE” FORTRAN program. Personal Communication.

Moglen, G.E. 2003. “CE465Hydro” extension in ArcView 3.0. Personal communication.

NSIDC, 2003. National Snow and Ice Data Center http://nsidc.org (accessed: February 2003)

Nash, J.E. and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models Part 1 – A discussion of Principles. *Journal of Hydrology*, 10 (1970), 282-290.

National Air and Space Administration (NASA). 2003. MODIS Home Page. http://ltpwww.gsfc.nasa.gov/MODIS/MODIS.html (accessed: November 2003)

National Air and Space Administration (NASA). 2003. ASTER Spectral Library. http://speclib.jpl.nasa.gov (accessed: November 2003)

Smith, L.C., G.A. MacDonald, K.E. Frey, A. Velichko, K. Kremenetski, O. Borisova, P. Dubinin, and R.R. Foster, 2000. EOS – Russia Venture Probes Siberian Peatlands Sensitivity to Climate, EOS 81, 497, 503-4.

Suhov, P.N.1992. About stages of growing of new structures on management by water resources in river Tom basin. http://www.oieau.fr/ciedd/contributions/at2/contribution/souhov.htm (accessed: November 2003)

USGS. 2003. Land Processes Distributed Active Archive Center http://edcdaac.usgs.gov/gtopo30/gtopo30.html (accessed: April 2003)

Ward, T.J. 1992. Criteria for evaluation of watershed models. *Journal of Irrigation and Drainage Engineering*, Vol. 119, No. 3, May/June, 1993.

Wellesley College. 2003. Image Galleries http://www.wellesley.edu/Russian/Baikal/Gallery/maps/Ob_Enisei_basin_map.jpg (accessed: November 2003).