Air Force Institute of Technology

AFIT Scholar

3-2000

Testing of the New USGS K Index Algorithm at Bear Lake Observatory

Ariel O. Acebal

Follow this and additional works at: https://scholar.afit.edu/etd
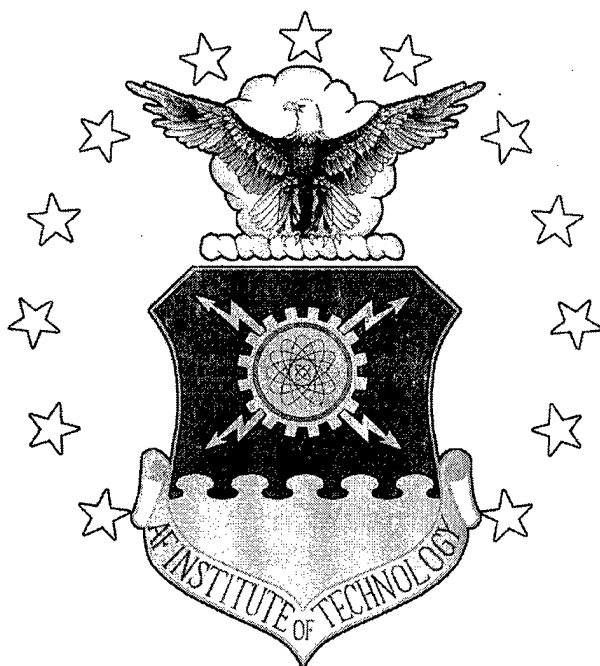
Part of the Other Earth Sciences Commons, and the Other Physics Commons

TESTING OF THE NEW USGS K INDEX
ALGORITHM AT BEAR LAKE
OBSERVATORY

THESIS

Ariel O. Acebal, Captain, USAF

AFIT/GAP/ENP/00M-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government

TESTING OF THE NEW USGS K INDEX ALGORITHM AT BEAR LAKE

OBSERVATORY

THESIS

Presented to the Faculty of the

Graduate School of Engineering Management

of the Air Force Institute of Technology

Air University

Air Education and Training Command in Partial Fulfillment of the

Requirements for the Degree of Master of Science in Applied Physics

Ariel O. Acebal, B.S.

Captain, USAF

March 2000
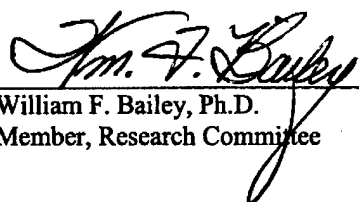
TESTING OF THE NEW USGS K INDEX ALGORITHM AT BEAR

LAKE OBSERVATORY

Ariel O. Acebal, B.S.
Captain, USAF

Approved:

_____   _____
Devin J. Della-Rose, Ph.D.   Date
Chairman, Research Committee

_____   _____
William F. Bailey, Ph.D.   Date
Member, Research Committee

_____   _____
David E. Weeks, Ph.D.   Date
Member, Research Committee

Acknowledgements

Abstract

The K index was developed by Bartels in 1939 as an estimate of the level of geomagnetic activity caused by the Sun. This index was computed manually every three hours at geomagnetic observatories using the magnetic traces of the surface planetary magnetic field. In 1991, the International Association of Geomagnetism and Aeronomy approved four additional methods to compute the K index; all of them were computer algorithms. One of the approved methods, the Wilson code, recently underwent some modifications. The new algorithm is now part of a Windows-based computer program being developed by the United States Geological Survey (USGS). After successfully evaluating a beta version of this new program, it was used to compute the K index for a new location. This new location is the Bear Lake Observatory (BLO), where the Utah State University has been collecting geomagnetic data from their magnetometer. Statistical techniques were applied to correlate K indices among existing stations in an effort to develop a test for the validity of the K index of a station. These statistical tests were applied to the BLO K index proving that the technique works and that the BLO K index was computed properly.

Table of Contents

List of Figures

x

List of Tables

TESTING OF THE NEW USGS K INDEX ALGORITHM AT BEAR

LAKE OBSERVATORY

1. Introduction

1.1 Problem Statement

The geomagnetic K index provides an objective and quantitative monitoring of the irregular variations of the Earth's magnetic field observed in a given location. This index can be computed using magnetometer observations and a computer algorithm. The Utah State University collects data from the Bear Lake Observatory (BLO) magnetometer, but does not compute the K index for this site. The United States Geological Survey (USGS) is developing a Windows based computer program to compute the K index for any site. This research will implement the official station K index at the BLO by using the data from the observatory's magnetometer and the USGS computer program. The new observatory's K index time series will be validated through correlation tests developed in this research project.

1.2 Background

The solar wind and the Interplanetary Magnetic Field (IMF) permeate the heliosphere. When they encounter a planet with a magnetic field, the solar wind and the IMF interact with the planet's magnetic field. This interaction leads to the generation of space currents. The space currents, in turn, affect the planetary magnetic field. By measuring the perturbations in the planet's magnetic field from ground-based observatories, a measure of the magnitude of the space currents can be determined. This leads to an indication of the solar activity and the amount of energy that it is being deposited into the planet's magnetosphere.

The amplitude of the perturbations of the magnetic field varies from observatory to observatory. Also, in order to determine the amount of solar activity, one would be forced to examine the traces of the magnetic field measurements. In order to simplify the process, geomagnetic indices were introduced. These indices provide an estimate for the level of activity in the interaction between the Earth's magnetic field and the solar wind. By comparing indices' values, the relative activity level of the magnetosphere and ionosphere (MI) system can be assessed. Increased activity in the MI system can lead to disturbances which often cause operational impacts such as satellite drag, temporary loss of high frequency communications between airplanes and ground controllers, and loss of imagery from weather and reconnaissance satellites [Shea and Smart, 1998].

Many indices have been developed and used over the years. One of the most commonly used indices is the K index. This index was first introduced in 1938 and adopted internationally in September 1939. The K index is calculated at magnetic

observatories throughout the world every three hours using ground-based measurements of magnetic variations due to fluctuations of MI currents.

Economic factors and trends to automate magnetic observatories created a need to have computer algorithms calculate the K index. An internationally accepted algorithm was designed by Wilson (1987) and a Windows-based computer program is currently under development by the USGS. This program computes the Sq curve based on raw magnetometer data. The code was recently modified and must be tested before the scientific community can use it.

1.3 Research Objective

There are three goals to this project. The first goal is to establish the accuracy and reliability of the new USGS algorithm. The program was evaluated using over forty years of archived data. The second goal is to determine if a correlation exists between the K index time series of the different observatories. If a correlation exists, then it can be used to evaluate the K index time series of a new observatory. Correlations among the observatories were computed using the K index time series. The last goal was to compute the K index at Bear Lake Observatory. The K index was computed using the new USGS algorithm. The validity of the K index time series was determined using the correlation tests developed as part of the second goal.

1.4 Air Force Impact

Today's Air Force is responsible for operating and controlling the airspace from the mud to the sun. This airspace includes the near-Earth space, where the Air Force commands and controls over 120 operational satellites supporting National Command Authorities, federal and civilian agencies and all U.S. and allied military forces. Interactions between these space platforms and energetic particles in this hostile environment can lead to hardware outages or data corruption [Tascione, 1994]. The frequency of these operations impacts increases during geomagnetic storms. Space operators can use advance warning of geomagnetic storms to change operations in an effort to minimize these impacts.

Single station K indices can be used to correlate the physical process they represent with other MI observations in order to increase the understanding of the solar-terrestrial interaction. Advancing this understanding will lead to better models, which will aid space weather officers in forecasting geomagnetic storms.

Finally, results of the new USGS program can be used to compare and evaluate the USAF K index program with the possibility of the USAF adopting the USGS algorithm. Currently, the USAF computes an hourly K-like index for various observatories. From these indices, a planetary index is computed and based on the results, worldwide warnings and advisories are disseminated. By using an internationally recognized algorithm, the USAF would have additional data from which to compute the planetary index. This increase could translate to a more accurate tool from which to forecast disturbances in the MI system.

1.5 Sequence of presentation

Chapter 2 is an overview of the space environment and its effects on the planet's magnetic field. It also reviews the Earth's magnetic field, how it is measured, and the coordinate systems used to describe the field. Finally, this chapter ends by describing the K index and how it is computed (both manual and computer methods). The next chapter discusses the methods used to test the new USGS algorithm, the evaluation of correlation between stations' K index time series as a function of distance and steps taken to compute the K index for BLO. Chapter 4 examines the results of all the tests and presents some of the data produced during the project. The last chapter serves as a conclusion and includes recommendations for the new USGS code, the correlation tests and the BLO K index. Finally, appendices A through D include some of all the data generated during the project.

2. Literature Review

2.1 Introduction

In order to understand the concepts explored in this document, a brief explanation of the interaction between the Sun, the IMF, and the Earth's magnetic field is necessary. This interaction generates current systems that affect the planet's magnetic field. The fluctuations in the surface magnetic field are measured with a magnetometer. Finally, the geomagnetic observations are used to compute the K index using different methods.

2.2 The Earth's Magnetic Field

The planet's magnetic field has been a subject of study since the Chinese first used magnets as navigational aids in 250 B.C. In 1600, William Gilbert published the first textbook in geomagnetism and concluded that the Earth behaved as a great magnet [Hine, 1968]. Gauss introduced improvements in magnetic field observations in 1848. The current theories on geomagnetism where introduced by Chapman and Bartels in 1940. All along, the magnetic field was measured and charted using different instruments and coordinate systems. Further studies have shown that the planet's magnetic field can be divided into internal and external magnetic fields. The following sections examine the coordinate systems used to chart the field, the instruments used to measure it, and the internal and external components of the field.

2.2.1 Coordinate systems

There are two coordinate systems used to describe the planet's magnetic field [Figure 1]. One of these systems is the XYZ system. In this system, the positive X component of the magnetic field points to the geographic north. Since this is a right-hand coordinate system, the positive Y-axis points east and the positive Z-axis points down towards the Earth. In this coordinate system, the individual components are usually measured in nanoteslas (1 Tesla = 10^4 Gauss = 10^9 Nanoteslas or Gammas). The other coordinate system is the HDZ system. The Z coordinate is the same as in the XYZ system. The H component of the magnetic field is the horizontal component (the component in the X-Y plane). The D component is the declination angle and it is measured clockwise from the X-axis to the H vector. In this system, the H and Z components are measured in nanoteslas, while the D component is measured in arc minutes (1 degree = 60 arc minutes). The following equations allow the conversion from one coordinate system to another:

$$D = \text{ArcTan}\left(\frac{Y}{X}\right)$$

and

$$H = \sqrt{X^2 + Y^2}$$

Finally, the total field component is F and it is given by

$$F = \sqrt{X^2 + Y^2 + Z^2} = \sqrt{H^2 + Z^2}$$
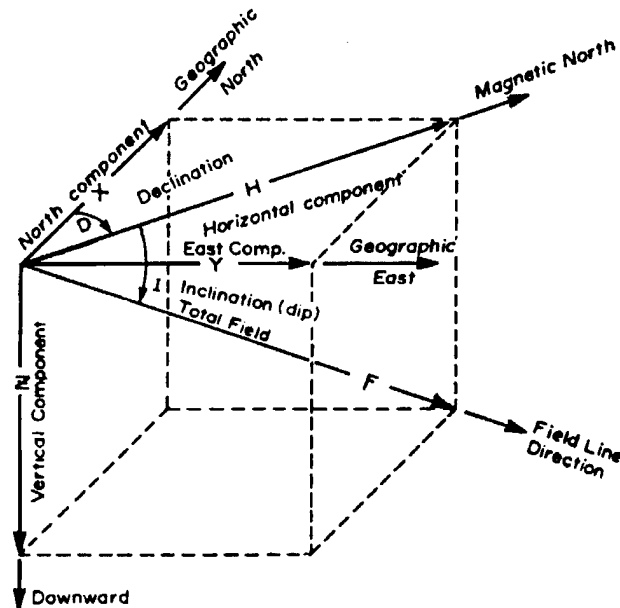
Figure 1 – Coordinate systems used to measure the planetary magnetic field (taken from Campbell, 1997).

2.2.2 Measuring the magnetic field.

Early magnetic field measuring instruments consisted of a magnetized spoon-shaped pointer that was allowed to move and point southward. Today's instruments are capable of measuring not only the direction of the field, but the magnitude as well. One such instrument is the magnetometer. A simple magnetometer consists of a pivoted-needle, a deflecting magnet and a scale as shown in Figure 2. The magnetometer is first lined up so that the needle lies in a north-south direction when the deflecting magnet is not present. Once the deflecting magnet is installed, the needle will take up a position in the resultant field of the earth and the deflecting magnet, giving the inclination angle.

Then by knowing the magnetic moment of the deflecting magnet and the distance from

the needle, H can be calculated according to

$$M = \frac{1}{2} H d^3 \operatorname{Tan} \theta$$

where M is the magnetic moment of the magnet,

θ is the measured inclination angle, and

d is the distance from the magnet to the needle [Hine, 1968].

This type of magnetometer is prone to errors due to the way that the inclination

angle is measured.
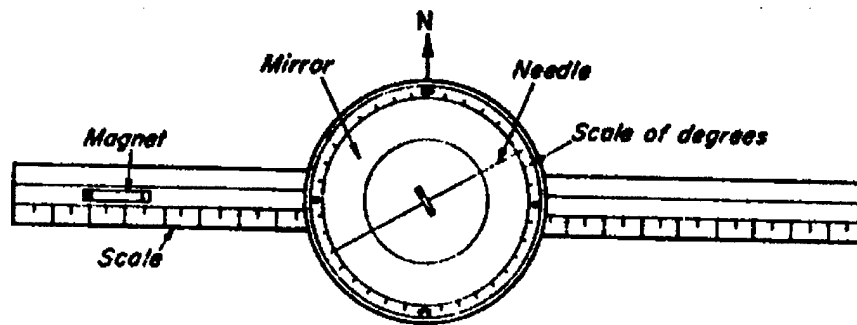
Figure 2 – Diagram of a simple magnetometer (taken from Hine, 1968).

A more accurate magnetometer is the fluxgate magnetometer. Many satellites and

most geomagnetic observatories use this type of instrument. The BLO uses a fluxgate

magnetometer. Figure 3 shows a diagram of the fluxgate magnetometer. The primary

winding contains a highly permeable ferromagnetic core. A high frequency alternating

current is run through the winding, generating a strong oscillating field. In the absence of

any external field, the oscillating field is symmetrical in the positive and negative

9

directions. If there is an external field (the Earth's field), the oscillating field is offset.

This offset is measured by the secondary winding which surrounds the primary coils.

The geomagnetic field amplitude and phase are then computed from these distortions

[Campbell, 1997]. Three sets of coils are used, one for each of the three coordinate axis.

A magnetometer trace is shown on Figure 4.
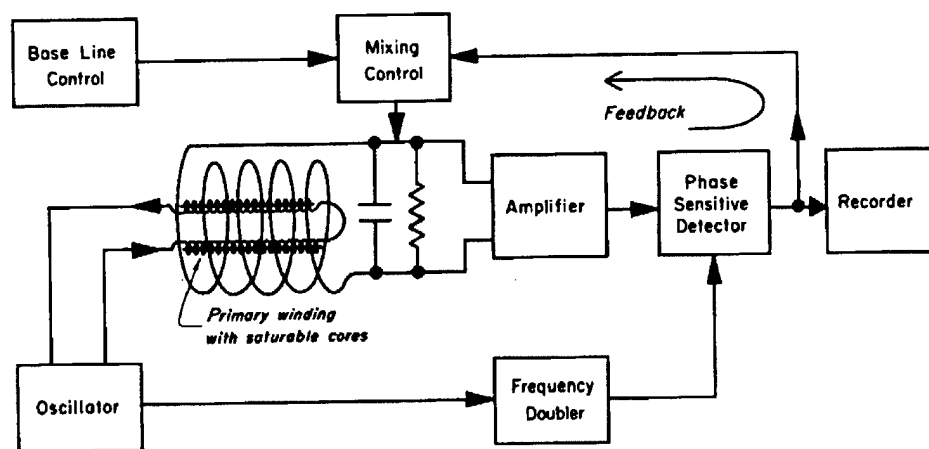
Figure 3 – Diagram of a fluxgate magnetometer (taken from Campbell, 1997).
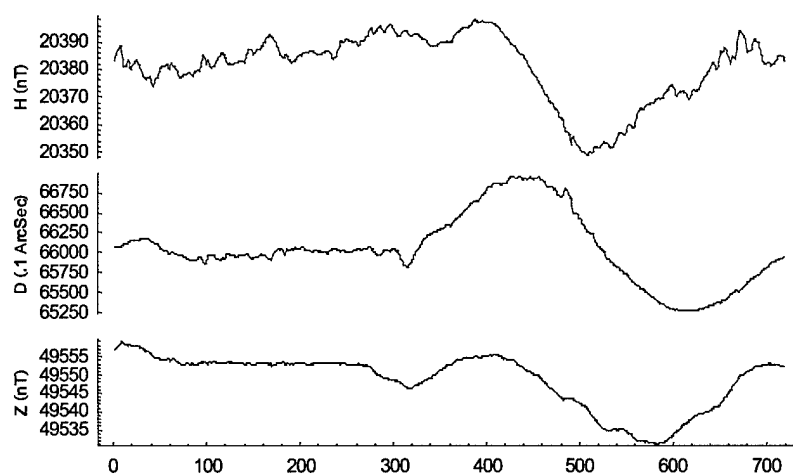
Figure 4 – Magnetometer trace for Boulder on 14 August 1999.

10

2.2.3 The Internal Field

The internal field can be subdivided into a dipole and non-dipole field. The internal field's dipole source is induced by currents due to the fluid motion in the liquid portion of core [Tascione, 1994]. To a first approximation, this magnetic field is similar to that of a uniformly magnetized sphere and has a direction of a centered dipole axis displaced 11.3° from the geographic pole (Figure 5). The magnetic properties of the Earth's crust and eddies in the internal current system are the main non-dipole sources. The average value of the internal field is 0.5 Gauss (50,000 nT) at the surface (for comparison purposes, a small bar magnet is about 1000 Gauss or 2,000 times stronger than the planet's magnetic field at the surface).
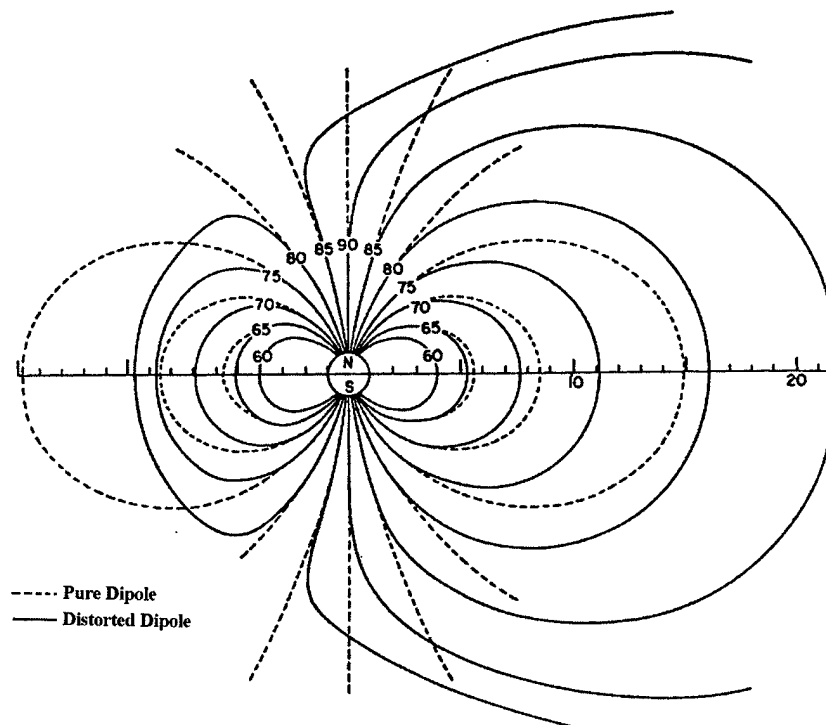
Figure 5 – Earth's magnetic field shown as a pure and distorted dipole (taken from McPherron, 1991).

2.2.4 The External Field

The external field's sources are the effects of various current systems in the earth's space environment and are often called the Magnetosphere and Ionosphere (MI) current system. These currents are described in the following sections and either enhance or diminish the magnetic field measured on the surface. How do the currents change the magnetic field? Ampere's law states that all currents have a magnetic field associated with them. This field is circular for a line current and its direction can be determined using the "right-hand rule". Pointing the thumb in the direction of the current, the fingers curl in the direction of the field. This simple explanation shows that the magnetic field changes directions about the current and that points in a perpendicular plane to the current that are 180° apart will have magnetic fields that point in opposite directions (Figure 6).
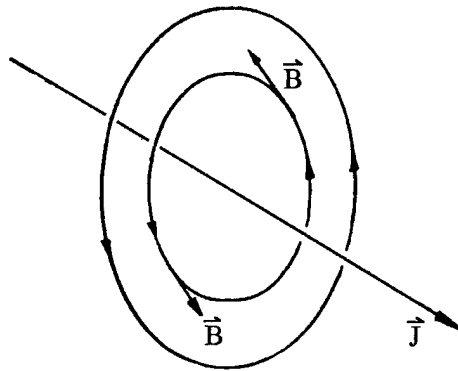
Figure 6 – The magnetic field around a current.

Now, the principle of superposition allows magnetic field to be added or subtracted, depending on the orientation of the two fields. So if two fields point in the

same direction, the net field will be the sum of the two, while if the fields point in opposite directions, the net field will be the difference of the two. This is what happens with the MI currents and the planetary field. Figure 7A shows how currents can enhance the H component of the magnetic field measured at the surface (in location marked X), while Figure 7B shows how a current flowing in the opposite direction can detract from the H component of the field measured at the surface.
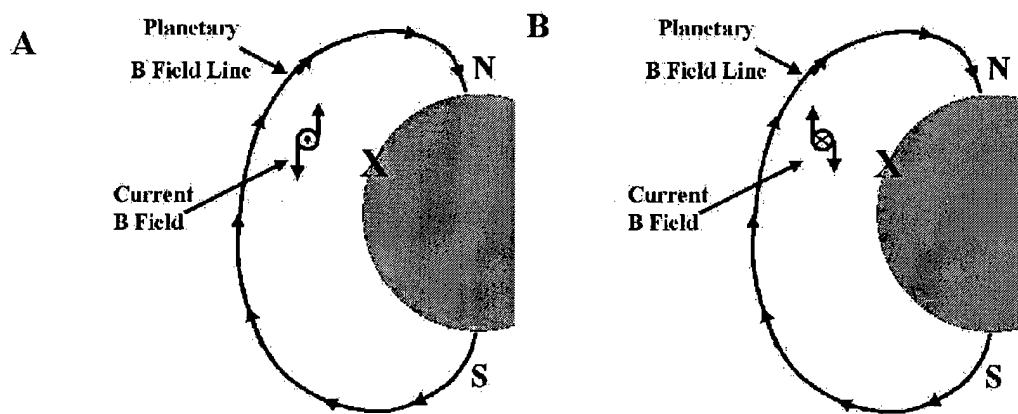
Figure 7 – Superposition of the planetary and current magnetic field.

One last point needs to be made before discussing the individual currents. The plasma content affects the magnitude of the MI currents. A change in the plasma content (caused by increase activity in the Sun) will alter the magnitude of the currents, which in turn will alter the magnetic field measured at the surface. Measuring the changes in the magnetic field lead to the different geomagnetic indices.

2.3 Magnetospheric Current Systems

The expansion of the Sun's corona is called the solar wind. This moving plasma has a magnetic field associated with it, and it's called the IMF. The solar wind and the IMF permeate interplanetary space to about 150 Astronomical Units (AU) forming the heliosphere. When the IMF encounters another planet's magnetic field, the two interact and generate local currents. The region of interaction is a definable boundary called the magnetopause (Figure 8). The cavity formed within the magnetopause is called the magnetosphere. Using a very simple analogy, the magnetosphere is bullet-shaped, similar to water flowing in a stream around a boulder. The blunt end of the magnetosphere faces the sun and, in the Earth's case, is located at a distance of about 10 Earth radii (R_e). This distance fluctuates depending on solar wind speed and density, and the orientation and strength of the IMF. During solar active times, this solar-side of the magnetosphere can be compressed earthward to distances within 7 R_e. On the antisunward side, the magnetosphere is stretched out into a tail, the magnetotail, to distances well beyond 60 R_e.
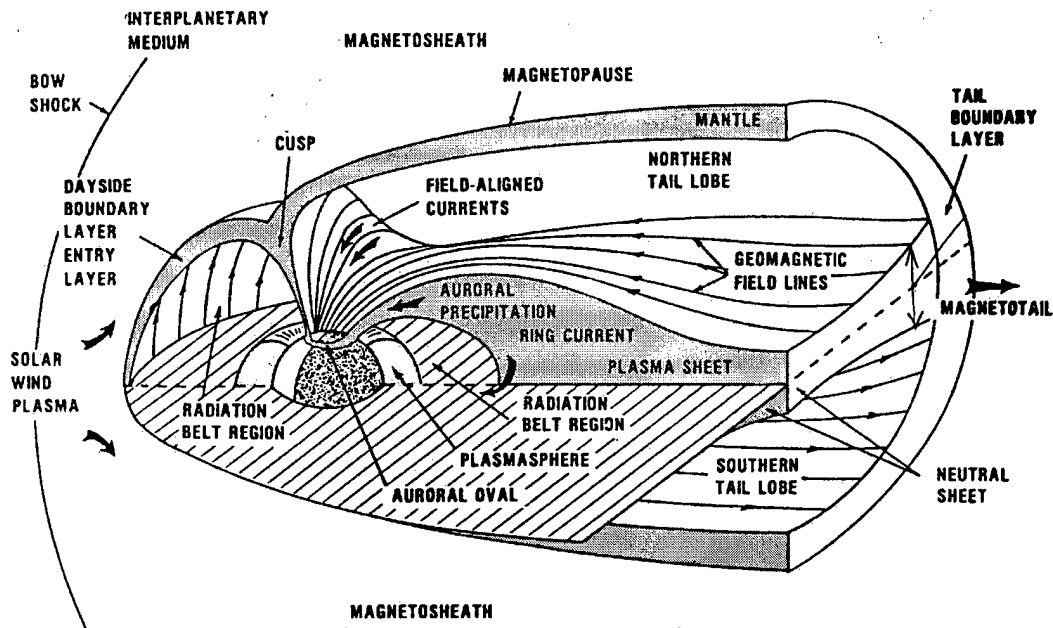
Figure 8 – The magnetosphere and its components (taken from Tascione, 1994).

2.3.1 Chapman-Ferraro Current

The region ahead of the magnetopause is the bow shock. In this region, the solar wind plasma is heated and compressed and it interacts with the geomagnetic field. In a simplified model of the boundary interaction, as the solar plasma flows around the magnetic field, positive and negative ions penetrate the magnetopause. Once inside the magnetopause, the ions experience the planet's magnetic field and a Lorentz force ($\vec{V} \times \vec{B}$). This Lorentz force forces the positive and negative ions to drift in opposite directions. Due to their larger mass, the protons have a much larger turning radius than the electrons. For that reason, the number of protons crossing an arbitrary line (line QS in Figure 9) is greater than the number of electrons crossing that same line. This net flux of protons represents a current.

15
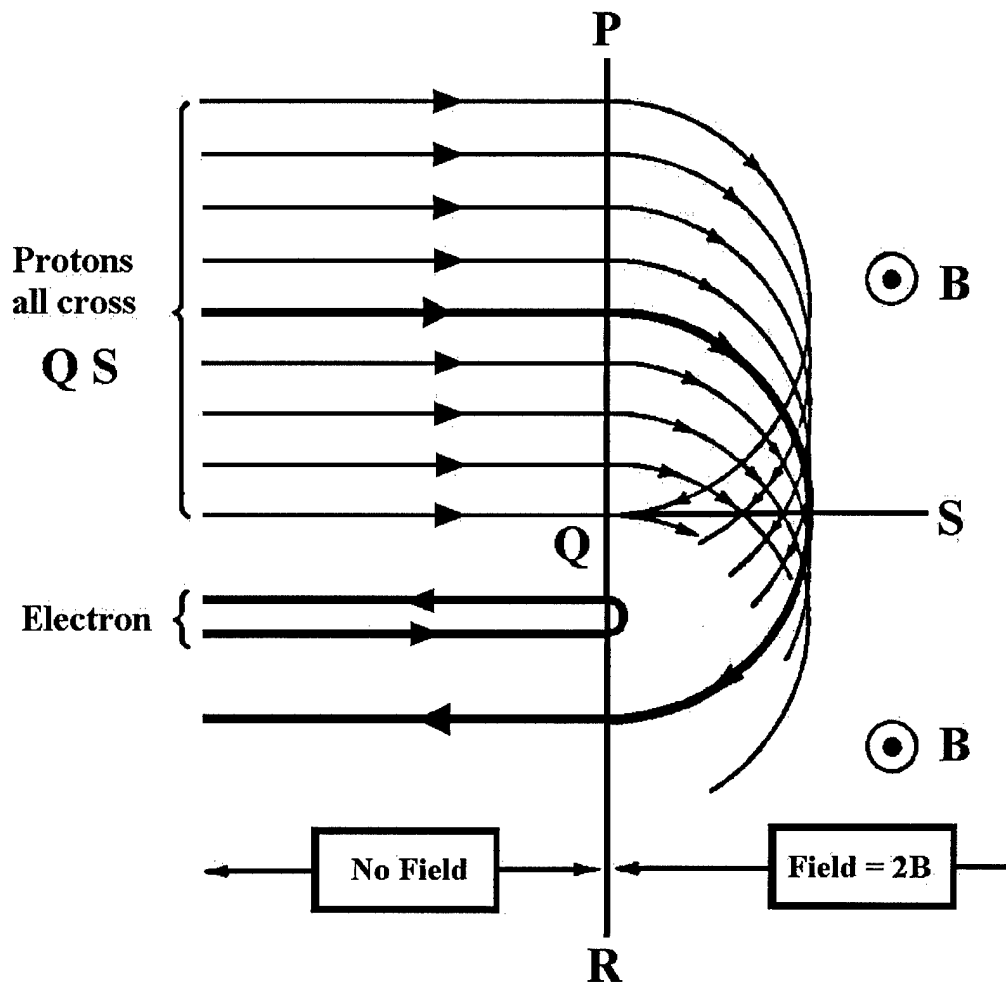
Figure 9 – A simple model of the magnetospheric boundary leading to the generation of the Chapman-Ferraro current (taken from Raddcliffe, 1972).

This current, the Chapman-Ferraro current, flows from the Earth's dawn side to the dusk side (Figure 10). The magnetic field associated with this current enhances the Earth's magnetic field earthward of the boundary. This enhancement translates to an increase of about 25nT to the H component of the Earth's magnetic field at the geomagnetic equator [Della-Rose, 1999].

16
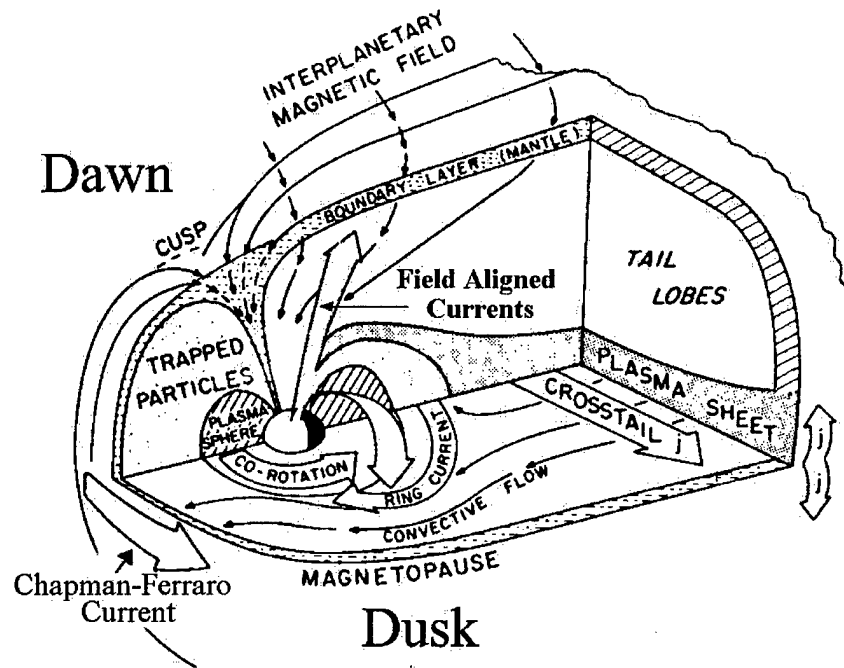
Figure 10 – Magnetospheric current systems (taken from Tascione, 1994).

2.3.2 Ring Current

Another current system, the ring current, is located closer to the Earth, at about 3 to 6 R_e. The source of this current is the eastward orbit of electrons and westward orbit of protons. The charged particles drift in opposite directions due to the inward-pointing planetary magnetic field gradient. The net result of the different orbits is a westward current. The magnetic field associated with this current is opposite the direction of the H component of the planetary field at the surface. Reductions to the geomagnetic field from this current system are on the order of 50 nT at the surface, but are greatly enhanced during geomagnetic storms causing up to 2% field reduction [Tascione, 1994].

2.3.3 Tail Current

On the anti-sunward side, the cross tail current flows from dawn to dusk. This current, like the ring current, detracts from the Earth's magnetic field, but only on the night side [Figure 10].

2.4 Ionospheric Current Systems

The Chapman-Ferraro, the ring and the cross-tail currents are considered magnetospheric currents. The ionosphere is also a source of currents that affect the overall strength of the geomagnetic field. These currents flow in a confined layer 90 to 150 km above the Earth's surface and can be regarded as low, mid, and high latitude current systems. Table 1 shows the latitudes associated with these different regions.

Table 1 – Approximate latitude for ionospheric regions used to describe current systems.

Region	Latitude
Low Latitude (Includes equatorial zone)	0° to 30°
Mid-Latitude (Sub-auroral)	30° to 70°
High Latitude (Auroral zone and polar cap)	70° to 90°

2.4.1 High Latitude Systems

At high altitudes, the magnetosphere is considered a collisionless plasma and is governed by the following equations:

$$\vec{E} + \vec{V} \times \vec{B} = 0$$

and

$$\rho \frac{\partial \vec{V}}{\partial t} = -\nabla p + \rho \vec{B} + \vec{J} \times \vec{B}$$

where E = electric field,

V = Plasma drift,

B = magnetic field,

g = gravity,

p = pressure, and

ρ = plasma density,

Taking the cross product of this last equation with $\bar{B}$ yields

$$J_\perp = \frac{1}{B^2}\rho\bar{B}\times\frac{\partial\bar{V}}{\partial t} + \rho\bar{g}\times\bar{B} + \bar{B}\times\nabla p$$

which shows that cross-field (Hall) currents are governed by pressure gradient and time-dependent plasma flow and not the electric fields [Kelly, 1989]. At these high latitudes, the magnetosphere can be divided into two regions: open and closed field lines. The open configuration indicates that the magnetic field lines originating at the surface of the planet connect with the solar wind and the IMF. The closed configuration indicates that the magnetic field line originates and ends on the surface of the planet. These two configurations give rise to different current systems.

Over the polar cap, the Earth's magnetic field topology is considered open. This allows for the interaction of the ionosphere with the solar wind and the IMF though the mapping of magnetospheric features. In this region of the magnetosphere, the plasma flows in an anti-sunward direction. This creates an electric field directed from dawn to dusk over the polar cap, which maps down to ionospheric levels (E_{PC}). The $\bar{E}$ x $\bar{B}$ plasma drift is then also in the anti-sunward direction in the ionosphere. At E-region

altitudes, the ions collide with neutrals, resulting in a slower and random-like drift for the

ions. This sets up a current flowing on the opposite direction of the plasma flow.

Therefore, there is a current directed sunward over the polar cap. This current results in a

surface magnetic field of 60 nT [Della-Rose, 1999].

The plasma flow over the auroral zone is in a sunward direction. Similar to the

polar cap region, the current flow is opposite the plasma flow (Figure 11). As a result, a

current flows in an anti-sunward direction.
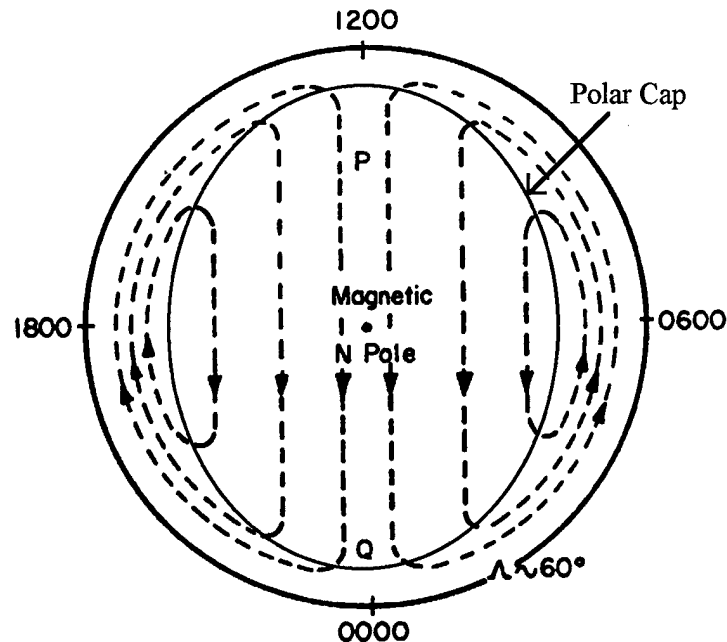
Figure 11- High latitude ionospheric plasma convection (taken from McPherron, 1991).
The times shown indicate the local time.

But the story doesn't end there. The dayside and the nightside have different

conductivities. The dayside conductivity results from photo-ionization. Particle

precipitation takes place in the night side auroral zone. The precipitation causes

ionization, which in turn enhances the night side conductivity. This conductivity gradient

sets up a polarization electric (E_{pol}) field over the polar cap aligned sunward (Figure 12). This field summed with the E_P field yields a net electric field (E_{net}) pointing to mid-afternoon. The effect of this net field is to rotate the location of the current system clockwise [McPherron, 1991]. As a result of this rotation, there is a westward flow on the dawn side called the westward auroral electrojet and an eastward flow on the dusk side, called the eastward auroral electrojet (Figure 13). It is important to stress that the rotation itself does not create the auroral electrojets, but it does affect their location. The magnetic field due to the eastward jet enhances the H component of the planetary surface magnetic field on the nightside while the westward jet's magnetic field detracts from the surface magnetic field. The effects on the surface field can be as great as 1000 nT during magnetic substorms [Della-Rose, 1999].
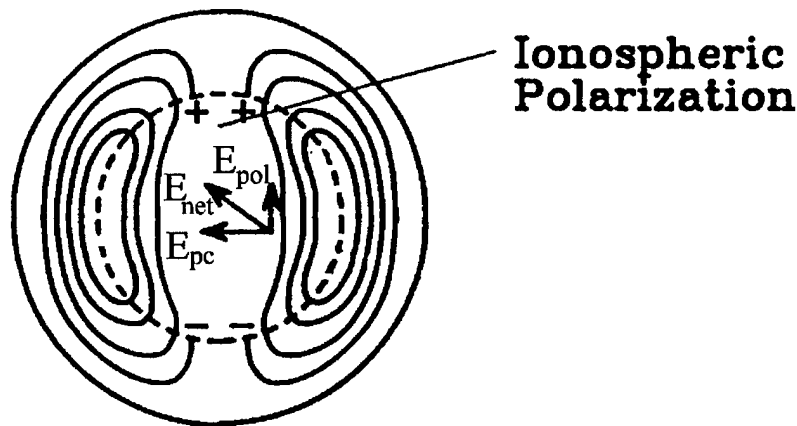
Figure 12 - Polarization, polar cap, and resulting net electric fields (taken from McPherron, 1991).
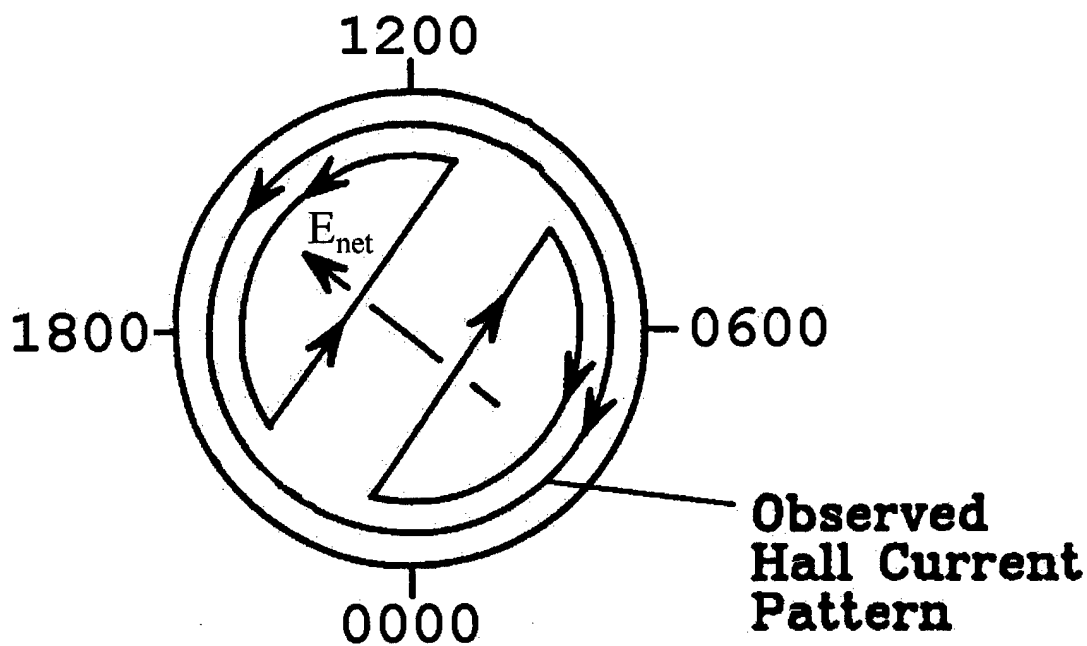
Figure 13 – Diagram of the auroral electrojet (taken from McPherron, 1991). The times shown indicate the local time.

Finally, the Birkeland or field-aligned currents connect the flow between the convection ionospheric and magnetospheric currents. The Birkeland currents cannot be uniquely distinguished from ground observations alone [Hargreaves, 1992].

2.4.2 Mid-Latitude Currents

Daytime heating creates a thermal high pressure cell centered at low latitudes. This high pressure cell drives the thermospheric wind polewards on the dayside and equatorward on the night side. This wind flow pushes the conducting ionospheric plasma into regions of increasing magnetic flux in the dayside and into regions of decreasing magnetic flux on the nightside. Faraday's law states that changes in the magnetic flux will generate an electromagnetic force (EMF) and Lenz's rule states that the EMF will be

22

set up in a direction to oppose the increase in the magnetic flux on the dayside (or to

oppose the decrease in the magnetic flux on the nightside). On the dayside, this current

will flow counter clockwise in the Northern Hemisphere (Figure 14), and clockwise in

the Southern Hemisphere. This is called the Sq (solar quiet) current and it is a very

important factor in the computation of the K index. The characteristics of this current are

defined during geomagnetic quiet periods. At nighttime, the flow pattern is reversed, but

since the ionosphere's conductivity is decreased due to recombination, the current's

magnitude is much smaller.
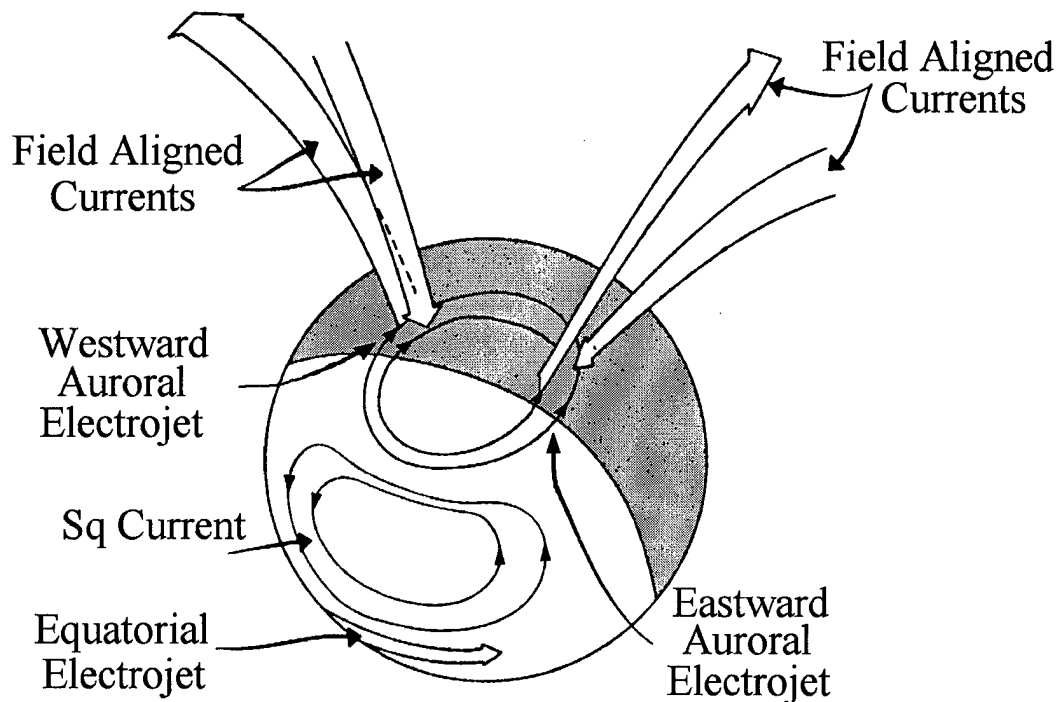
Figure 14 – Ionospheric current systems (taken from Tascione, 1994).

These patterns are fixed in relation to the sub-solar point, while the Earth rotates

underneath them. This creates a secondary Sq current underground since the Earth is a

conductor. This current is about a third of the magnitude of the ionospheric Sq current [Della-Rose, 1999]. The effects of both currents are added (Figure 15). The "total" Sq current is measured by ground-based instruments and is a function of local time, latitude, longitude, season, and solar cycle. The local time dependency (Figure 15) exhibits a diurnal and a semidiurnal variation. The diurnal variations are less than 50 nT and the semidiurnal variation, associated with the lunar position, variation is very small ($\pm$ 4 nT) [Rishbeth and Garriott, 1969]. Figure 16 shows the latitude and seasonal dependency. The solar cycle variation is larger: the solar cycle maximum Sq current is between 1.6 and 3.0 times larger than the solar minimum Sq current [Della-Rose, 1999].
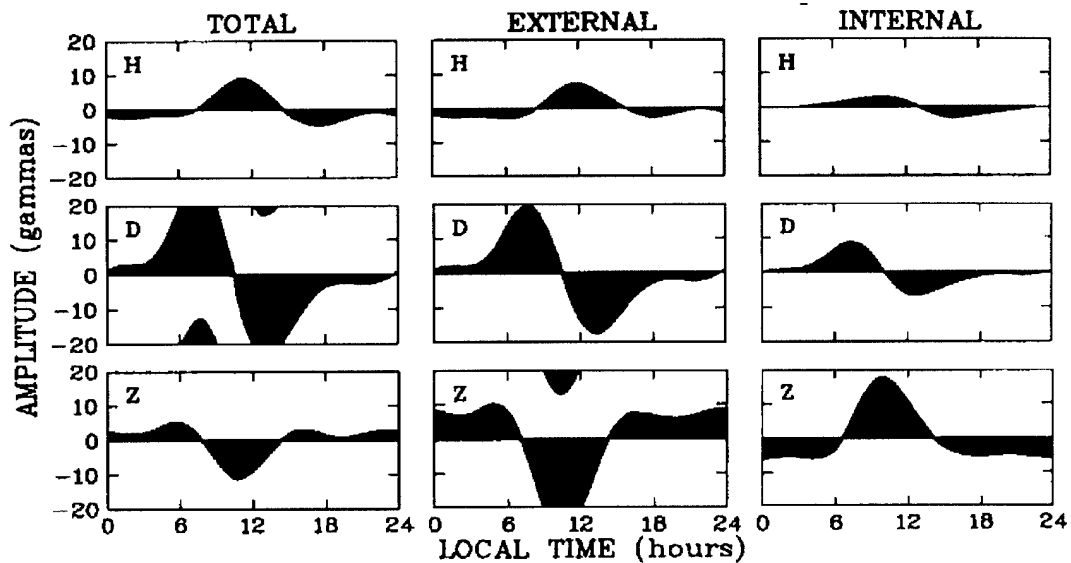
Figure 15 – Internal and external Sq Components at 35°N. (taken from Campbell, 1997).
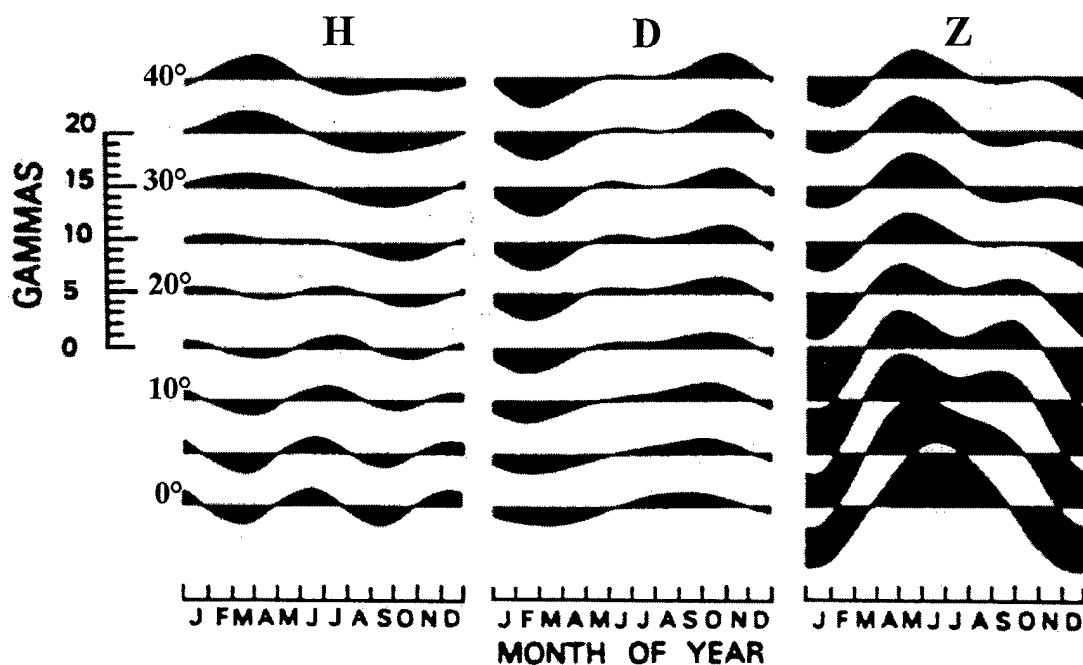
Figure 16 – Sq variations base on latitude and season (taken from Campbell, 1997).

2.4.3 Low Latitude Currents

The electric field over the geomagnetic equator is oriented east-west. In this region, the ionosphere is horizontally stratified. This leads to vertical electric fields that enhance conductivity in this region. The enhanced conductivity is also in the east-west direction and is called the cowling conductivity. The cowling conductivity is responsible for the current jet located at the equator. Generally, this jet flows eastward during the day (Figure 14) and westward at night. The electrojet is about 450 to 500 km wide [Campbell, 1997], but its effects fall off rapidly with latitude. Thus, this current does not significantly affect the K index.

2.5 The K index

Geomagnetic indices provide an estimate for the level of activity in the interaction between the Earth's magnetic field and the solar wind. By comparing indices' values, the relative activity level of the MI system is determined. A more active Sun will increase magnetospheric and ionospheric activity which often leads to operational impacts such as satellite drag, temporary loss of high frequency communications between airplanes and ground controllers, and loss of imagery from weather and reconnaissance satellites [Shea and Smart, 1998].

Many geomagnetic indices have been developed and used over the years. One of the most commonly used indices is the K index. Dr. J. Bartels first computed this index in 1938 at the Niemegk Observatory in Germany. The index was adopted internationally in September 1939 and has been in use ever since. The K index is calculated at magnetic observatories throughout the world every three hours based on universal time (UT) using ground-based measurements of the planet's magnetic field. This index measures the magnetic perturbations to the planetary field. The amplitude of the perturbation is assigned a class value using a quasi-logarithmic scale (Table 2). The values of the K index range from 0 (indicating "quiet" conditions) to 9 (indicating "severe storm" conditions). Each observatory has a scale like the one shown for the Niemegk Observatory.

Table 2 – K index class limits for the Niemegk Observatory.

Range (nT)	K Value	Relative Level of Activity
0 – 4	0	Quiet
5 – 9	1	
10 – 19	2	
20 – 39	3	Unsettled
40 – 69	4	Active
70 - 119	5	Minor Storm
120 – 199	6	Major Storm
200 – 329	7	Severe Storm
330 – 499	8	
500+	9	

In order for the K index to be meaningful, each K index category must have the same meaning at each observatory. In other words, the frequency distribution of the K index values must be the same at all the sites [Menvielle and Berthelier, 1991]. This is achieved by proportionally adjusting the range limits of the K index categories at each observatory. In practice, the scale is defined by the lower limit of a station's K = 9 value. Then, all other classes are scaled so that they are proportional to the Niemegk scale. For example, the lower limit of the K = 9 class for an equatorial station might be 250 nT, while it remains 500 nT for a mid-latitude station [Della-Rose, 1999]. The K index is more relevant at mid-latitudes where the effects for the auroral electrojets can be measured.

Before discussing how the K index is actually computed, it is necessary to discuss the importance of the K index. The United States Air Force uses a planetary "average" of station K indices, or K_P index to issue alerts and warnings about increased geomagnetic activity. These warnings are important to the satellite operator's community since geomagnetic storms have been associated with satellite surface charging and increased

atmospheric drag. The USAF has divided the K_P index values into three main categories: quiet ($K_P = 0$-3), active ($K_P = 4$), and storm-level ($K_P = 5$-9). Based on the level of activity reported by the USAF, satellite operators take actions to protect their assets.

2.5.1 Hand-scaled K index

As stated in the previous section, the K index is based on the amplitude of the perturbations of the planetary field. The method used to compute the amplitudes is rather simple, but has one complicated step. For a three-hour period, the H and D components measured at the observatory are separately plotted, along with a plot of the H and D magnetic signature of the Sq current for that period. Next, the Sq curve is moved up until it touches the maximum value of the H component. The position of the Sq curve is plotted, and the Sq curve is the moved down until it touches the minimum value of the magnetometer trace. Once again, the Sq curve's new position is plotted, resulting in two parallel Sq curves. The vertical distance between the Sq curves is the three-hour range value in nanoteslas for the H component (Figure 17). The same procedure is repeated for the D component. Finally, the larger of the two range values is converted into a K index using the observatory's scale (Table 2). Two important points need to be emphasized at this time. First, computing the Sq curve is not a simple process. Several methods have been established by experts in the field to compute the daily Sq current [Mayaud, 1967]. In the end, it comes down to subjectivity by the individual observer computing the K index. Studies showed that different observers draw different Sq curves for the same observatory [Wilson, 1986 and Riddick and Stuart, 1984]. The second point is that in the process described above, the effects of the Sq current on the magnetic field are removed.

This means that for sub-auroral observatories, the auroral electrojets and the Birkeland currents are the main currents affecting the K index.

Figure 17 shows six different intervals where the K index is computed. The dashed line is the Sq curve and the dotted lines are the upper and lower bound of the Sq curve. Each amplitude a is then converted to a K index using the scale for the observatory (assuming that the H component amplitude is larger than the D component amplitude).
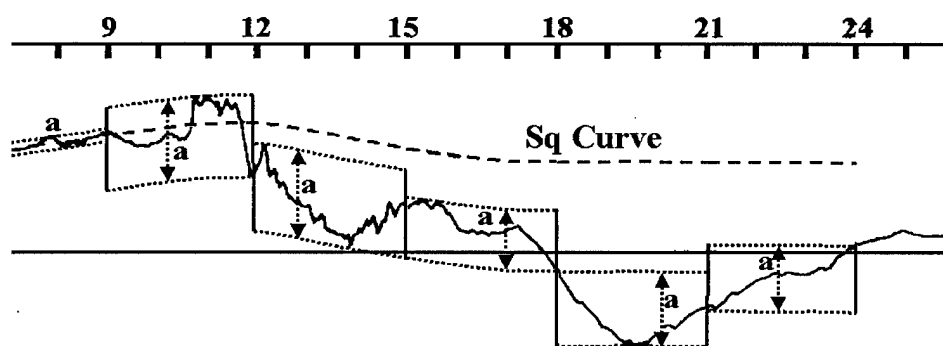
Figure 17 – Magnetometer trace of the H component for Guam as a function of local time. The Sq is plotted along with the magnetometer trace. The 3-hour amplitude is converted to a K index value using the station's table. (taken from Mayaud, 1980).

2.5.2 Computer generated K index

Economic factors and trends to automate magnetic observatories created a need to have computer algorithms calculate the K index. According to Della-Rose (1999), the International Association of Geomagnetism and Aeronomy (IAGA) approved the following K index methods in 1991:

29

1.   "USGS": the method developed by the United States Geological Survey (USGS) is described in Wilson (Journal of Geomagnetism and Geoelectricity, 39, 97, 1987).

2.   "AS": the Adaptive Smoothing method is described in Jankowski, et al. (Annales Geophysicae, 6, 589, 1988) and in Novozynski, et al. (Geophysics Journal International, 104, 85, 1991).

3.   "FMI": the method developed by the Finnish Meteorological Institute is described in Sucksdorff, et al. (Geophysical Transactions, 36, 344, 1991).

4.   "LRNS": the Linear-phase Robust Non-Linear Smoothing method is described in Hattingh, et al. (Annales Geophysicae, 6, 611, 1988) and in Hattingh, et al. (Geophysics Journal International, 99, 533, 1989).

5.   "HS": The Hand Scaling traditional derivation process remains the basic reference and a still accepted method, following the rules described in the "Atlas des Indices K" by Mayaud (IAGA Bulletin no. 21, 113pp., 1967).

The main difference among these five methods revolves around the calculation of the Sq curve.

The USGS algorithm computes the daily Sq curve by fitting a curve to a set of mean hourly values (MHV) computed from the magnetometer readings. The code uses the 24 MHV for a given day, as well as the last two MHV from the previous day and the first two MHV from the next day. If an MHV value is missing, the code uses the daily or monthly MHV.

A linear least-squares fit is used to generate a line to a sliding set of three MHVs. The intersections of the resulting consecutive lines are computed and then a cubic spline

30

is fitted to these points. The resulting spline is the Sq curve for that day. The smoothing of the digitally derived Sq curve is statistical and depends on the time interval between mean values used to fit the curve. In the manual method, the smoothing of the Sq curve is determined by the observer and is subject to variation from observer to observer. After computing the digital Sq curve, the K index is determined in the same manner as described for the manual K index.

The USGS algorithm has undergone some changes. These changes were incorporated into the new USGS K index Windows based program (referred to as the new USGS code throughout this document). The differences between the "new" and the old (referred to as the old Wilson code throughout this document) USGS algorithms are subtle. The main difference between the two algorithms is how the Sq curve is calculated. In the old Wilson code, the Sq curve was computed one day at the time, regardless of the amount data available. This means that for a 20-day calculation, there are 20 spline fits with a total of 40 anchor points. In the new USGS code, the Sq is computed for all the available data. In the case of a 20-day calculation, there is only one spline fit, with only two anchor points. The result of this change is an improvement on the spline fit.

3. Methodology

3.1 Introduction

This project was divided into three parts; all parts involved the comparison of data between the 13 USGS geomagnetic observatories and the Bear Lake Observatory (BLO) in Utah (Figure 18 and Table 3). Part one tested the new algorithm by generating K indices for 13 geomagnetic observatories. These results were then compared to K indices generated by hand-scaled analysis and other computer algorithms. Parts two and three were closely related. In the second part, different methods were tested to try to establish a correlation between two stations' K-index time series and the distance separating the two observatories. Part three computed and established the reliability of the K index for BLO. The validity of BLO's K index was verified using the results of part two.
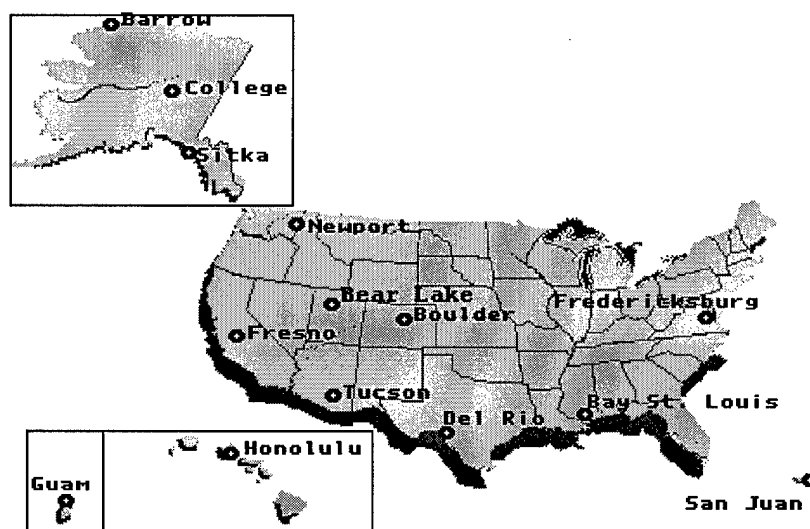
Figure 18 – Map of the geomagnetic observatories used in this research.

Table 3 – Location of the geomagnetic observatories used in this research.

Station ID	Location	Geographic		Geomagnetic	
		Latitude	Longitude	Latitude	Longitude
BLO	Bear Lake, UT	41.90N	248.60W	49.75N	248.60W
BOU	Boulder, CO	40.14N	254.76W	49.14N	319.58W
BRW	Barrow, AK	71.32N	203.38W	70.05N	250.88W
BSL	Bay St. Louis, MI	30.40N	270.60W	41.50N	340.74W
CMO	College, AK	64.86N	212.16W	65.06N	263.71W
DLR	Del Rio, TX	29.49N	259.08W	38.93N	326.30W
FRD	Fredericksburg, VA	38.21N	282.63W	49.32N	357.90W
FRN	Fresno, CA	37.09N	240.28W	43.07N	303.58W
GUA	Guam, M. Islands	13.56N	144.87W	5.82N	215.88W
HON	Honolulu, HI	21.32N	201.94W	21.61N	269.83W
NEW	Newport, WA	48.26N	242.88W	55.03N	303.22W
SIT	Sitka, AK	57.05N	224.67W	59.81N	279.98W
SJG	San Juan, PR	18.38N	293.88W	28.51N	010.50E
TUC	Tucson, AZ	32.25N	249.17W	39.90N	314.45W

3.2 Data Availability

Two different types of data were used in this project: magnetometer data and K index time series. The magnetometer data came from four different sources: the National Geophysical Data Center (NGDC), the USGS Golden Geomagnetic Information Node (GIN), the USGS Data CD-ROM, and the BLO magnetometer. The K index time series were generated for each observatory using three different K index calculation methods. The sources specific to each observatory and dates available for the data are listed in Table 4 (excluding BLO, which is discussed later).

The magnetometer data were actual measurements from the magnetometers and thus considered raw data. This type of data were in the form of one-minute averaged (30 second-averaged for the BLO data) geomagnetic digital recordings from the magnetometers. The magnetometer data from NGDC and GIN were downloaded from

their respective sources. Unlike the CD-ROM magnetometer data, no quality control was performed by the USGS on these data sets. Therefore, quality control had to be performed to ensure accurate results. The data sets were checked for bad data points by establishing a maximum value that the components could change between measurements. These threshold values were different for each of the three magnetic components (H, D, and Z) and were large enough to allow for real physical phenomena that causes spikes in the data, while being low enough to discard instrument measuring errors. If a datum point was identified as bad (improper format or unreliable value), it was assigned the average value of the six closest observations (three preceding and three following the observation in question). If a large sequence of observations were identified as bad data points (more than 5 bad data points in a row), they were assigned a special value identifying them as missing data. These measurements were not used in any of the calculations or comparisons. Finally, the data were plotted for a final qualitative check of each data set. Any remaining bad data points were coded as missing values.

If the geomagnetic data were provided in the X, Y, and Z coordinates, as in the case of the BLO data, a coordinate change was made using the following equations:

$$D = \operatorname{ArcTan}\left(\frac{Y}{X}\right)$$

and

$$H = \sqrt{X^2 + Y^2}$$

The transformed BLO data were then checked for bad data points using the procedures previously described. The magnetometer data for BLO were available from July 1999 through October 1999.

The K index time series were derived using raw magnetometer data and three

different methods of computing the K index. Therefore, this type of data were considered

derived or processed data. The three type of K index time series were: hand scaled K

indices (HK) from NGDC, old Wilson code K indices (OK) from the USGS CD-ROM,

and new USGS code (NK). The hand scaled and the old Wilson time series were

archived data. The computations of these K indices were performed by the USGS. The

NK time series were computed using all the quality controlled magnetometer data

available and the new USGS program. This gave most observatories three different types

of K index time series. Since the only data available for BLO were 30-second averaged

magnetometer measurements, only the NK time series was available for BLO.

Table 4 – Data available for the project. BLO data is not included in this table.

Station ID	K Index		Geomagnetic 1.0 minute averaged	
	Hand-Scaled	Old Wilson and New USGS Code	USGS CD-ROM	NGDC and GIN
BOU	01/70-02/78	01/90-12/94	01/90-12/94	08/99-10/99
BRW	01/70-03/75	01/90-12/94	01/90-12/94	
BSL		01/90-12/94	01/90-12/94	
CMO	01/57-06/91	01/90-12/94	01/90-12/94	
DLR		01/90-12/94	01/90-12/94	08/99
FRD	01/57-12/92	01/90-12/94	01/90-12/94	08/99-10/99
FRN		01/90-12/94	01/90-12/94	08/99
GUA	01/70-12/86	01/90-12/94	01/90-12/94	
HON	01/47-12/90	01/90-12/94	01/90-12/94	
NEW	01/70-04/82	01/90-12/94	01/90-12/94	08/99-10/99
SIT	01/64-12/86	01/90-12/94	01/90-12/94	
SJG	01/70-12/87	01/90-12/94	01/90-12/94	
TUC	01/70-12/86	01/90-12/94	01/90-12/94	08/99

3.3 Testing the New Code

The first part of the research was designed to establish the accuracy and reliability of the new K-index algorithm. This part compared the HK (when available), OK, and NK K index time series for each individual station. To carry out the comparison, one of the time series was considered the true K index (TK) and the other one was considered the K index time series to be tested (QK). Table 5 lists the different ways that the three time series were compared during this set of tests.

Table 5 – Different type of K index comparisons made during the first part of this project.

TK	vs.	QK
Hand-Scaled (HK)		Old Wilson code (OK)
Hand-Scaled (HK)		New USGS code (NK)
Old Wilson (OK)		New USGS code (NK)

In all cases, the difference between TK and QK were examined. A difference of 0 indicated a perfect match between values being compared, +1 indicates the QK value was one K index category smaller than the TK, where as -1 indicates the QK values to be one K index category larger than the TK value, and so on. The differences were reported according the value of the TK category in question. For example, if TK = 1 and QK = 2, then the comparison would be reported as a –1 in the K = 1 category.

When comparing computer-generated K indices with hand-scaled K indices, the common practice is to truncate the computer-generated value [Della-Rose, 1999]. This allows the comparison of two integers, but more importantly, it is an attempt at preserving the integrity of the long-running K index database for each observatory. The idea is that a K index value computed with a computer-algorithm should have the same meaning as a value computed with the manual method. The first test compared the three

36

methods of generating the K index for each observatory by truncating, rounding to the

nearest integer, and using the actual decimal values for the OK and NK values. The

rounding and truncating comparisons were simple since the difference between two

integers yields an integer. A more complicated scheme had to be employed when

comparing the hand-scale method with the actual value of the computer-generated

methods. Since the difference between the two values could lead to a decimal value, the

differences had to be binned as shown on Table 6. The results of this test were used to

determine which comparison method was more effective.

Table 6 – Range of values for the differences resulting from the comparison of hand-scaled to computer-generated K index values.

Actual Difference	Reported Difference
0	0
0.1 to 1.0	1
1.1 to 2.0	2
2.1 to 3.0	3
3.1 and higher	4

The second test in this phase of the project involved the comparison of the K

index distributions created for each observatory using the three different computing

methods. This test used all the data available for each observatory, which in most cases

did not cover the same time period. For almost all of the observatories, there was more

hand-scaled K index data than either the old Wilson code or the new USGS program K

indices. The idea behind this test was that if two computing methods are similar, then the

distributions of their results should be equal, provided that the distributions are large

enough. Were the distributions large enough? The smallest distribution in this

comparison test was five years, or 14,600 K index observations worth of data (Table 4).

Based on this, it was determined that the distributions were large enough for these tests. The distributions were compared using three different methods: comparisons of the plots of the three distributions, the two-sample Kolmogorov-Smirnov (KS) test, and comparison of the distributions' moments.

The first distribution comparison test simply binned the data according to the K index class and method used to compute the K index. Then, the distributions were normalized by dividing each K index category by the total number of K index values for that distribution (i.e. a percentage distribution). This allowed a direct comparison of distributions of different size. This test was a direct and common sense approach to comparing the distributions. This comparison test could only have two basic results; either the distributions were exactly alike or they were different. If the distributions were identical, there would be no difference in any K index category. If this was not the case, then the distributions were different. But, how different is different?

To answer this question, the qualitative comparison had to give way to quantitative tests. The first of the quantitative tests was the two-sample Kolmogorov-Smirnov test as described by Connover (1971) and Wilks (1995). For this test, a null hypothesis is proposed. This hypothesis is then either accepted or rejected at a confidence level based on the statistical test. In this case, the null hypothesis was that the two distributions were samples from the same parent distribution. This test was run three times for each observatory that had all three distributions (see Table 4 and Table 5). Once the null hypothesis is established, then the computations began. The first step was to order the two distributions, $G(x)$ and $F(x)$, from smallest to largest. Then, each ordered number in a distribution was assigned a value denoting the proportion of values

that are less than or equal to it. Suppose the distribution G(x) consists of 10 values, then

after ordering the values, the lowest value would have a rank of 1/10 and the highest

value would have a rank of 10/10 or 1. This creates two "new" distributions, $S_G(x)$ and

$S_F(x)$. The next step was to compute the test statistic according to

$$D_{max} = Max \mid S_G(x) - S_F(x) \mid$$

This is simply the absolute value of the largest difference between each new

"pair" of data. An example of the ranking and D_{max} computation is shown on Table 7.

Table 7 – An example of the two sample Kolmogorov-Smirnov test.

G(x)	F(x)	D=|(S_G(x) − S_F(x))|	G(x)	F(x)	D=|(S_G(x) − S_F(x))|
	5.2	0-1/15=0.067			5/10-8/15=0.033
	5.7	0-2/15=0.133	9.9		6/10-8/15=0.067
	5.9	0-3/15=0.200	10.1		7/10-8/15=0.167
	6.5	0-4/15=0.267	10.6		8/10-8/15=0.267
	6.8	0-5/15=0.333			8/10-9/15=0.200
7.6		1/10-5/15=0.233	11.2		9/10-9/15=0.300
	8.2	1/10-6/15=0.300		11.3	9/10-10/15=0.233
8.4		2/10-6/15=0.200		11.5	9/10-11/15=0.167
8.6		3/10-6/15=0.100		12.3	9/10-12/15=0.100
8.7		4/10-6/15=0.000	12.5	12.5	10/10-13/15=0.133
	9.1	4/10-7/15=0.067		13.4	10/10-14/15=0.067
9.3		5/10-7/15=0.033		14.6	10/10-15/15=0.000

The null hypothesis is then rejected at the (100-α)% level if

$$D_{max} > \left[\frac{1}{2} \left(\frac{1}{n} + \frac{1}{m} \right) \ln \left(\frac{\alpha}{2} \right) \right]^{1/2}$$

where n = total number of data point in the G(x) distribution,

m = total number of data points in the F(x) distribution, and

α = significance level expressed as a percentage

For the example on Table 7, D_{max} = 0.333 and occurs between $F(x)$ = 6.8 and

$G(x)$ = 7.6. The null hypothesis is accepted at the 10% level since D_{max} = 0.333 < 0.366.

One final point needs to be emphasized about this test. The result of this test does not say

that the two distributions are equal. What it does say is that if the null hypothesis is

accepted, then there is an α% chance that the distributions are not related.

Another quantitative test involved the moments of the distributions. The mean

and variance for each distribution type within an observatory were computed and

compared. Then, the difference of each of these values was examined. Larger

differences meant that the distributions were "more" different than distributions that

yielded smaller differences.

The next batch of tests binned the K index differences by K index class, time of K-

index, day of the month, and month of calculation. In these tests, the data were for the

same time period. This meant that most of the comparisons were between the K indices

generated from the old Wilson code and the new USGS code.

3.4 Testing for a Correlation Between Stations

The second part of the project was to establish a correlation between the K-indices

of different stations. If some relationship could be established between existing

observatories, then that relationship could be used to establish the accuracy of the K-

index of a new station (BLO in this case). The rationale is that the MI current dynamics

affect the planetary magnetic field. The changes they induce on the surface magnetic

field should be similar for locations in close proximity, leading to similar K index values. The similarity should decrease as the distance between observatories increases, leading to different K index values.

The comparison between any two stations was done using the same K-index computation method for the same time period. For example, Boulder's NK time series from January 1990 to December 1994 was compared to Newport's NK time series for the same time period. Four different comparison tests were done: plotting the difference between station A and station B as a function of time, scatterplots, Root Mean Squared Error (RMSE), and Spearman rank correlation coefficient.

In the first method, the difference of the K-index time series of two nearby stations should hover around zero with few fluctuations of ±1. As the distance is increased, the difference in the K indices should increase, leading to more fluctuations of greater magnitudes. This test was not designed to compute a numerical correlation between the stations, but to qualitatively show the effects of distance on the K index difference between two observatories.

The second test in this section was to graphically display the paired data using a scatterplot. A scatterplot is a collection of points in the plane whose Cartesian coordinates are the values of each member of the data set. These plots can show trends, clustering of values, and outlier points. The scatterplot for the same station (K-index of station A plotted on both axes) would show a straight line with a slope of one going through the origin. If different data were used on either axis, then the shape of the scatterplot would change from a straight line to a surface area. The shape of this area would depend on how different the two data sets are to each other. If the differences are

small, then the area of the scatterplot would be similar to the straight line indicating a tight clustering of the data. If the difference is large, the area covered by the scatterplot would also be large (loose or no clustering). This means that stations that are in close proximity (small differences in the data sets) should show a tight clustering of data about a 45° line with few outlying points. As the distance is increased, the clustering should broaden and the number of outlying points should also increase. Figure 19 shows the effect of increasing differences between the two data sets being plotted in the scatterplot. As in the case of the previous test, this test was not designed to compute the correlation between stations but to qualitatively show whether a correlation exists between stations and how it changes based on distance.
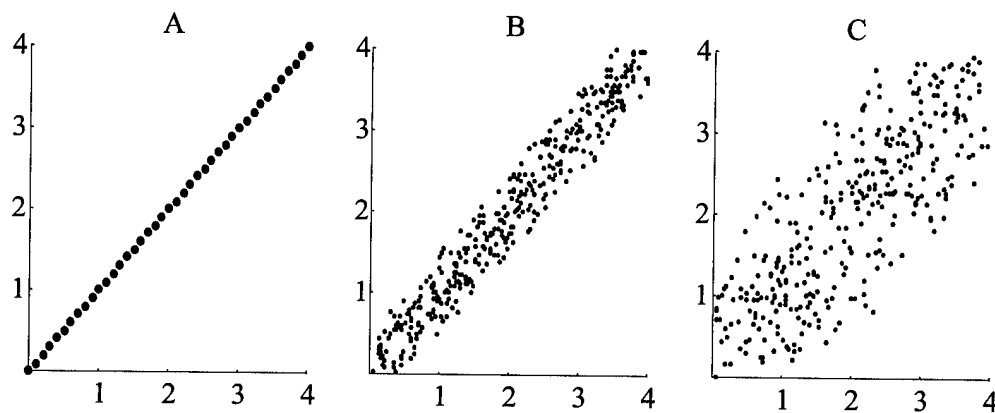
Figure 19 – Examples of scatterplots with increasing differences being plotted. In A, there is no difference between the two data sets, B shows an increase in the difference (but still highly correlated), and C shows large differences between the data sets.

The previous two tests were qualitative and were used as common sense tests. The RSME is a quantitative test used to measure the differences between two data sets. In this case, the RMSE was used to measure the similarity between the K index time

series between stations. If two time series are identical, the RMSE value is zero. As the differences increase, so does the RMSE value. The RMSE was computed using the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2}$$

where X_i and Y_i are the ith K-index value at stations X and Y, respectively.

Another quantitative measure is the Spearman rank correlation coefficient and is described in great detail by Wilks (1995) and Conover (1971). The Spearman rank correlation coefficient, or rank correlation, is similar to the often used (or misused) Pearson Ordinary Correlation Coefficient. Instead of using the actual data, the rank correlation uses the rank of the data. The data, K index in this case, are ordered according to their values from smallest to largest. The lowest value is assigned a rank of 1, and so on until the Mth data point is assigned the largest rank of M. If two or more data points have the same value, then the rank for the repeated data points is the average rank of those values. Table 8 shows an example of a data set and how it is ranked.

Table 8 – An example of a data set and its rank.

X	Y	Rank X	Rank Y	(Rank X-Rank Y)^2
2	8	1	8	49
3	4	2	5	9
4	9	3	9	36
5	2	4	2	4
6	3	5	3.5	2.25
7	6	6	6	0
8	3	7	3.5	12.25
10	1	8.5	1	56.25
10	7	8.5	7	2.25

The rank correlation coefficient, r, is calculated using

$$r = \frac{\sum_{i=1}^{n} (\text{Rank} (X_i) * \text{Rank} (Y_i))}{\sqrt{\sum_{i=1}^{n} (\text{Rank} (X_i))^2} * \sqrt{\sum_{i=1}^{n} (\text{Rank} (Y_i))^2}}$$

where X_i and Y_i represent the rank of the K-index value for that particular station. The rank correlation coefficient can also be computed using

$$r = 1 - \frac{6 \sum_{i=1}^{n} (\text{Rank} (X_i) - \text{Rank} (Y_i))^2}{n (n^2 - 1)}$$

where n is the number of pairs of data points. Using the sample data set in Table 8, the rank correlation coefficient would be

$$r = 1 - \frac{6 \sum_{i=1}^{n} (\text{Rank} (X_i) - \text{Rank} (Y_i))^2}{n (n^2 - 1)} = 1 - \frac{6 * 171}{9 (81 - 1)} = -0.425$$

The rank correlation has several advantages over the Pearson correlation coefficient. The Pearson correlation coefficient tests the linear correlation between two variables. As a result, depending on the data distribution of the two stations, the Pearson correlation coefficient could over- or underestimate the actual correlation between the stations. If there is a non-linear relation, that is, the relation between the two stations is more curve-like, the Pearson correlation coefficient will underestimate the correlation. If the data set contains some outliers, these large values will cause the Pearson correlation coefficient to overestimate the correlation. The rank correlation tests the monotonic relationship between the two variables. This means that large data points will not overestimate the correlation, and that non-linear relationships will not cause the correlation to be underestimated. Finally, a normal distribution of the data is required for the Pearson correlation to be meaningful. In the case of the K index, the distribution is

not normal. The rank correlation works on any type of distribution since it is a nonparametric test.

The value of the rank correlation coefficient will lie in the interval $-1 \leq r \leq 1$. If $r = -1$, then there is a perfect negative correlation between station X and station Y. That means that as station X's K-index values increase, station Y's values decrease. If $r = 1$, then it is a perfect positive correlation between the two stations. As the value of the K-index of one station goes up, so does the other. If $r = 0$, then there is no correlation between the two stations.

As with any test, the results have to be compared to some standard. In the previous paragraphs, the standard for a perfect correlation, no correlation, and perfect-match RMSE values were stated. But, what are other significant values? To answer this question, some benchmark values needed to be established for the rank correlation coefficient and the RMSE. These values were determined by computing the correlation (and RMSE) between an observatory's K index time series and the same time series with a ± 1 K index error. The ± 1 criterion was established since this is the accepted margin of error when comparing different methods of computing the K index at a station [Della-Rose, 1999]. The ± 1 error time series was created by adding a value between -1 and 1 to the actual time series. The value added was created using a random number generator. Obviously, these values will not indicate a perfect correlation (or match). However, the values generated with this method should help establish what correlation values between different station can be considered as good correlations.

Next, the RMSE and the rank correlation coefficient values were plotted as a function of distance. From the physical processes involved with the K-index, one would

expect the correlation to drop off as a function of distance. The distance between two stations are actually Great Circle distances computed using

$$D = \left(\frac{2\pi r}{360}\right) ACos(Cos(\phi_1)Cos(\lambda_1)Cos(\phi_2)Cos(\lambda_2) +$$

$$Cos(\phi_1)Sin(\lambda_1)Cos(\phi_2)Sin(\lambda_2) + Sin(\phi_1)Sin(\phi_2))$$

where (ϕ_1, λ_1) is the latitude and longitude of the first station,

(ϕ_2, λ_2) is the latitude and longitude of the second station, and

r is the radius of the Earth.

The distances were computed using the geographical and geomagnetic coordinates of the observatories.

The last step was to do a least-squares fit to the RMSE and rank correlation coefficient as a function of distance. This was done in two different ways: straight line least-squares fits and polynomial least-squares fits. The goodness of fit for the straight line and the various m^{th} degree polynomials were establish using the following criteria described by Wilks (1995):

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}(x_i) - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

where y_i is the observed value,

$\overline{y}$ is the arithmetic average of the observed values, and

$\hat{y}(x_i)$ is the predicted value from the least-squares fit of the y_i.

R^2 is the coefficient of determination and has a value of one in case of a perfect fit between the data and the least-squares fit values. The goodness of fit was also determined using the RMSE value between the data and the least-squares fits. The

polynomial least squares fit with the largest R^2 (and lowest RMSE) value was selected as the best-fit equation.

A clarification needs to be made at this point. The RMSE and rank correlation coefficients between the observatories were computed using hand-scaled, old Wilson, and new USGS K index time series. That means that for each pair of observatories (78 total), three RMSE and three rank correlation coefficients were computed. Furthermore, since both linear and polynomial least-squares fits were used, there were six data fitting equations for both the RMSE and rank correlation coefficients. The resulting equations were all a function of distance. All these equations were used in section 3.5 to determine the validity of the BLO K index.

The RMSE and rank correlation computations were done using a minimum of five years' worth of data for all of the USGS geomagnetic observatories. Additionally, the computations were carried out using all three types of K index time series. Unfortunately, there was only three months of BLO data available due to initial calibration problems with the BLO magnetometer. Since there was such a large discrepancy with the data set sizes (14600 versus 720 K index values), the validity of using the RMSE and rank correlation values computed from the large data sets was in question. To resolve this possible problem, the RMSE and correlation coefficient computations were re-run using only three months worth of data. The computations were run over 1000 times using a random number generator to select the starting point on the K index time series of both observatories. Once the starting point was identified, three months' worth of data were used to compute the RMSE and the rank correlation coefficient for that station pair. The results from all test runs were averaged and the

standard deviation was computed for each pair of stations. These new "average" RMSE and rank correlation coefficients were used to do the least-squares fits and generate the twelve data fitting equations (six for the RMSE values and six for the rank correlation values).

3.5 Computing and Validating the K-index for BLO

The last step in this project was to compute the K-index for BLO and to determine whether the values were similar to surrounding stations. As previously mentioned, the BLO data had to be converted from X, Y, and Z to H, D, and Z coordinates. Additionally, the BLO data are averaged every 30 seconds while the USGS K-index program uses one-minute average data. Two different methods were used to create BLO one-minute averaged geomagnetic data from the 30 second averaged BLO data. The first method was to average two consecutive 30-second averaged data points. The second method was to disregard the data points that were identified as the half-minute recordings (i.e. 1.5, 2.5, …58.5, 59.5). These data were then written to a formatted binary file compatible with the USGS code so that the K index could be calculated.

Prior to computing the NK for BLO, the data were plotted and checked by comparing it to different stations. To do the qualitative test, two sets of three stations in line were selected. The first set included the Fresno, Tucson, and Del Rio observatories, while the second set included the Newport, Bear Lake, and Boulder observatories. It is easy to see from Figure 18 that geographically, both Bear Lake and Tucson are the middle stations in either set. The idea behind this test was that if similar data patterns

48

were observed in the H, D, and Z fields among the first set of stations, then similar patterns should also be evident among the second set of stations. Also, if the data from Tucson fell in line with the two "bracketing" stations of Fresno and Del Rio, then the same result could be expected when comparing the Bear Lake data with data from Boulder and Newport.

The USGS program was then used to generate NK values. These K index values were then used to compute RMSE and correlation values between BLO, BOU, and NEW for the same time period. Finally, using the line-fitting methods discussed in section 3.4 and the distance between BLO and these two other observatories, the actual correlation and RMSE values were compared against the predicted correlation and RMSE values.

4. Results and Analysis

4.1 Testing the USGS K Index Program

The USGS K-index program was tested to establish its accuracy computing the K-index for different stations and to determine if there were any systematic trends. The tests included comparing values according to computing method, geographic location of the observatory, K-index categories, hour of calculation, day of calculation, and month of calculation.

4.1.1 Rounding, Truncating or Real Value?

Before accomplishing any comparison test, a decision had to be made on how to handle non-integer K indices. The K-index is an integer whose value ranges between 0 and 9. USGS computer-generated K-indices are not integer values, but rather, interpolated values between the integer categories of the K index. The current practice when comparing computer-generated K indices to hand-scales K indices (or when using the K indices to compute the planetary K index, K_P) is to truncate the computer-generated K indices. This first test was designed to determine whether it was more accurate to compare K indices by truncating, rounding, or using the actual interpolated values.

The results of this test (Table 9) show the best agreement between "TK" and "QK" values are achieved by truncating the computer-generated K-index than either one of the other two methods. Based on these results, all other comparison tests were conducted by truncating the computer-generated K-index.

Table 9 Computer-generated K-index accuracy by method expressed as a percentage.

Station ID	K-index compared	Truncating		Rounding		Actual values	
		Exact	Within ±1	Exact	Within ±1	Exact	Within ±1
CMO	OK-NK	86.6	99.9	84.0	99.9	22.1	99.7
BRW	OK-HK	86.3	99.9	83.4	99.9	18.9	99.8
SIT	OK-NK	85.4	99.7	83.7	99.9	21.6	98.6
NEW	OK-NK	84.6	99.8	82.5	99.9	18.8	99.7
FRD	OK-NK	78.5	99.7	73.8	99.8	12.6	99.6
BOU	OK-NK	78.0	99.8	72.6	99.8	12.0	99.4
BOU	OK-HK	78.0	99.8	72.6	99.8	12.0	99.4
CMO	HK-NK	72.3	95.7	63.5	95.7	9.8	93.9
CMO	HK-OK	72.2	95.6	56.7	95.8	7.4	93.7
FRD	HK-OK	71.3	99.1	57.8	97.9	8.9	95.8
SIT	HK-NK	70.3	91.4	54.0	90.7	8.9	88.5
BSL	OK-NK	65.4	99.2	58.4	99.2	6.9	97.6
TUC	HK-NK	63.4	93.5	58.5	92.7	8.9	90.6
TUC	OK-NK	62.7	98.8	56.7	98.5	6.8	96.4
DLR	OK-NK	62.2	98.7	56.0	98.5	7.1	96.4
FRD	HK-NK	60.8	98.4	66.8	98.4	10.5	95.9
GUA	OK-NK	54.3	97.2	57.1	97.0	5.7	93.5
FRN	OK-NK	52.4	86.1	45.2	85.1	5.1	81.4
SJG	OK-NK	52.3	96.8	57.1	96.9	6.2	92.6
HON	OK-NK	43.8	86.9	44.0	83.6	4.2	78.3
GUA	HK-NK	27.2	64.0	25.6	67.9	4.1	55.0

4.1.2 A Complete Comparison Test

One difficulty encountered while attempting to run the comparison tests was the lack of concurrent data (data availability table from chapter 3). There was concurrent data for two stations, Fredericksburg and College, from January 1990 to June 1991. This allowed for a three-way comparison of each station (HK-OK, HK-NK, and OK-NK). These results (Table 10) show that neither the OK nor the NK is an exact match to the HK values. This is not surprising since similar results have already been shown for different K-index algorithms [Wilson, 1987]. Furthermore, this type of difference can also be seen when comparing hand-scaled K-indices computed by different observers [Riddick and Stuart, 1984]. One advantage that the computer-generated K-indices have

51

over the HK values is that the subjectivity among observers is removed. If there is a

systematic bias by the program, it will be applied to all K-index computations in the same

manner. The results for College show that the new USGS algorithm performed as well as

the old Wilson code when compared to the hand-scaled method. The comparison

between the old Wilson code and the new USGS algorithm show that they are exact 86%

of the time, and 99.8% when the results are looked at with a ±1 tolerance. The results for

Fredericksburg were not as good when looking at exact matches. The new algorithm was

almost 10% "less exact" than the old Wilson code. Additionally, the two codes only

agreed 79% of the time. The possible reason for this reduced accuracy is discussed in the

next section. The comparison of the methods improved to over 98% when the results are

looked at with a ±1 tolerance.

Table 10 - Direct Comparison of HK, OK, and NK. All three K index time series were for the same time period.

Station ID	Type Comparison	Difference of					Within ±1
		-2	-1	0	1	2	
CMO	HK-OK	0.9	4.2	72.2	19.3	1.5	95.7
CMO	HK-NK	0.9	3.8	72.3	19.6	1.4	95.7
CMO	OK-NK	0.1	6.9	86.3	6.7	0.1	99.8
FRD	HK-OK	0.5	7.4	71.3	20.4	0.3	98.9
FRD	HK-NK	0.6	5.0	60.8	32.6	0.9	98.4
FRD	OK-NK	0.3	4.9	78.5	16.2	0.1	99.6

4.1.3 Comparing K index distributions

Figure 20 shows selected results from the comparison of the distributions

generated from the three different K index time series for each observatory. These results

show that none of the distributions are exact matches. This figure also shows that the

difference increases in stations of lower latitudes.
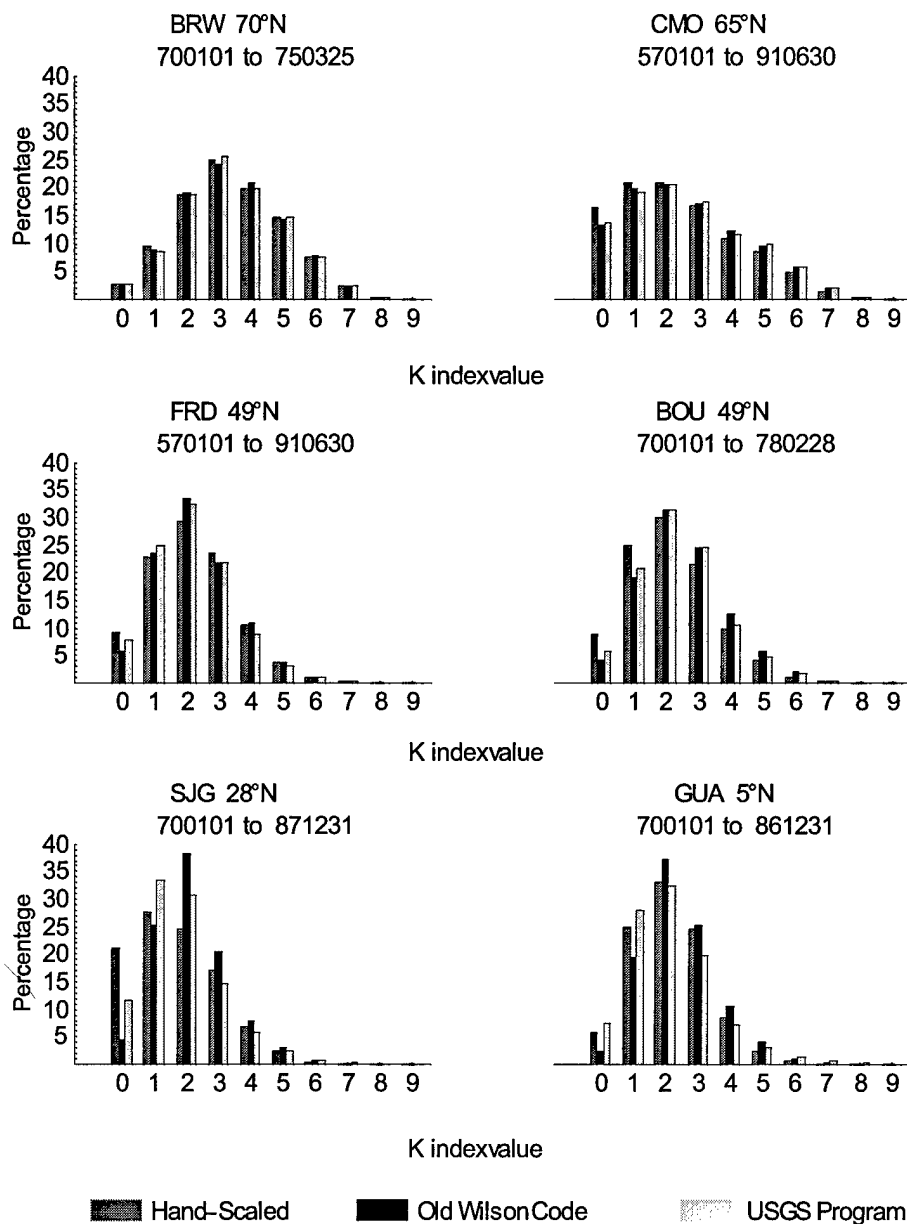
Figure 20 – Comparison of the three types of normalized distributions for different stations. The geomagnetic latitude of the station is shown next to the observatory's name. The dates (YYMMDD) appearing below the station identifier indicate the time period for which the hand-scaled K index time series was available.

The next test performed on the distributions was the two-sample Kolmogorov-Smirnov (KS) test. This test examined whether two distributions came from the same parent distribution. With the exception of Barrow, all null hypotheses had to be rejected even at a 30% significance level. This means that even accepting the risk of being wrong 30% of the time, the distributions did not seem to be related. This confirms the results shown by just plotting the distributions.

Evaluating the means of the distributions was also consistent with the findings of the plots and the KS test. The means of the distributions for a given observatory were different. This difference increased in magnitude as the observatory's latitude decreased.

All of the previous tests indicate that the distributions generated by the different K index computation methods were different. Is this necessarily a bad finding? To answer this question, a new test was carried out. This test was not described in the methodology chapter. Recall from chapter two that the USAF reports the K_P index as quiet ($K_P = 0$-3), active ($K_P = 4$), and storm-level ($K_P = 5$-9). Using this operational criterion, the K index distributions were re-examined. The results showed that in all cases, the computer-generated K index time series reported more K index values in the active and storm-level categories than the hand-scaled method. Further more, the distributions computed with the old Wilson code and the new USGS code were within 1% of each other in the storm level category. Also, with the exception of Bay St. Louis, Del Rio, and Guam, the two computer algorithms' distributions were within ± 2 % of each other in the other two categories. Finally, in most cases, the new USGS code distributions resembled the hand-scaled distributions better than the old Wilson code distributions. The complete results are shown on Appendix B.

4.1.4 Comparison by Observatory Location

In this case, as in the case of the rest of the comparison tests, the old Wilson K index time series was compared to the new USGS K index time series. Figure 21 shows the different observatories ordered according to their geomagnetic latitudes. The results show that the USGS program's ability to reproduce the old Wilson code results increases towards northern latitudes. Also, towards lower latitudes, the USGS algorithm underestimates the old Wilson K index value (the difference between the two is positive). Both of these factors can be attributed to the K index conversion table for each observatory. As explained in chapter two, the largest irregular range value between H and D for a given three-hour period is converted to the K index using an observatory-specific conversion table. Stations at higher latitudes have a broader range between K index classes than stations at lower latitudes. Thus, a small difference in the computed range value could lead to a different K index category at a lower latitude observatory, while the higher latitude observatory might not experience a difference.
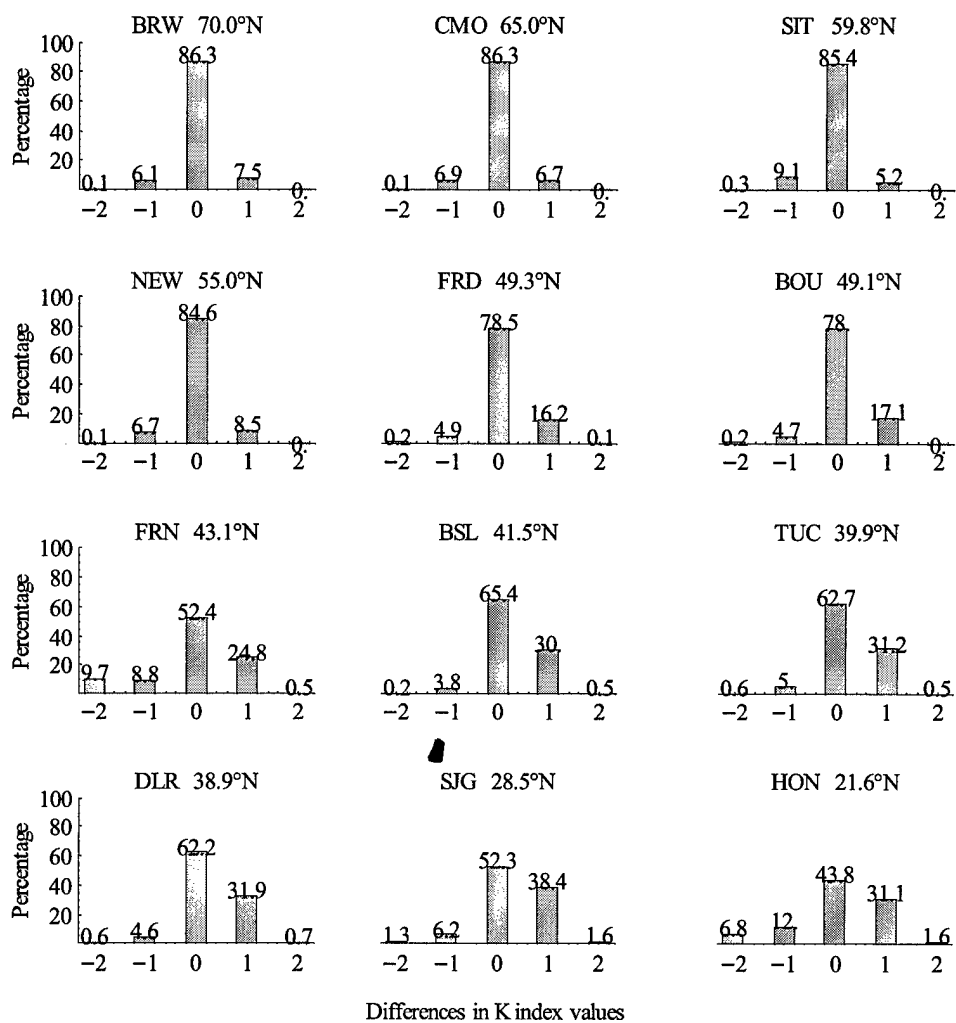
Figure 21- Performance of the new USGS K index algorithm based on observatory location. The geomagnetic latitude of the observatory is shown to the right of the observatory's name.

4.1.5 Comparison by K-index Category

Figure 22 shows the comparisons made at each observatory between K index categories. Once again, the agreement increases towards higher latitudes. These plots could be misleading and need further clarification. There are two features that are a result of the tests conducted and not the performance of the new USGS K index program. One feature is that there is more variability toward the higher K index categories. This is misleading since there are fewer higher value instances of the index than lower value ones. Differences between the two codes in these higher K index categories create a large percentage difference. Another observation from this data is that the zero category differences are always negative (overestimated by the new code). Again, this is misleading and is due to the way the differences were calculated and organized. Recall from chapter three that all comparisons were carried out by subtracting the new code K index from the old Wilson K index value. The result of the difference was assigned to the K index category of the old Wilson code. Therefore, in the zero category, all differences have to be negative. There are two features that highlight limitations of where the USGS code can be applied. Both of these features involve the variability in the mid-range K index values. First, the variability increases with decreasing latitude. Second, the USGS code gives a lower K index value than the old Wilson code in these categories.
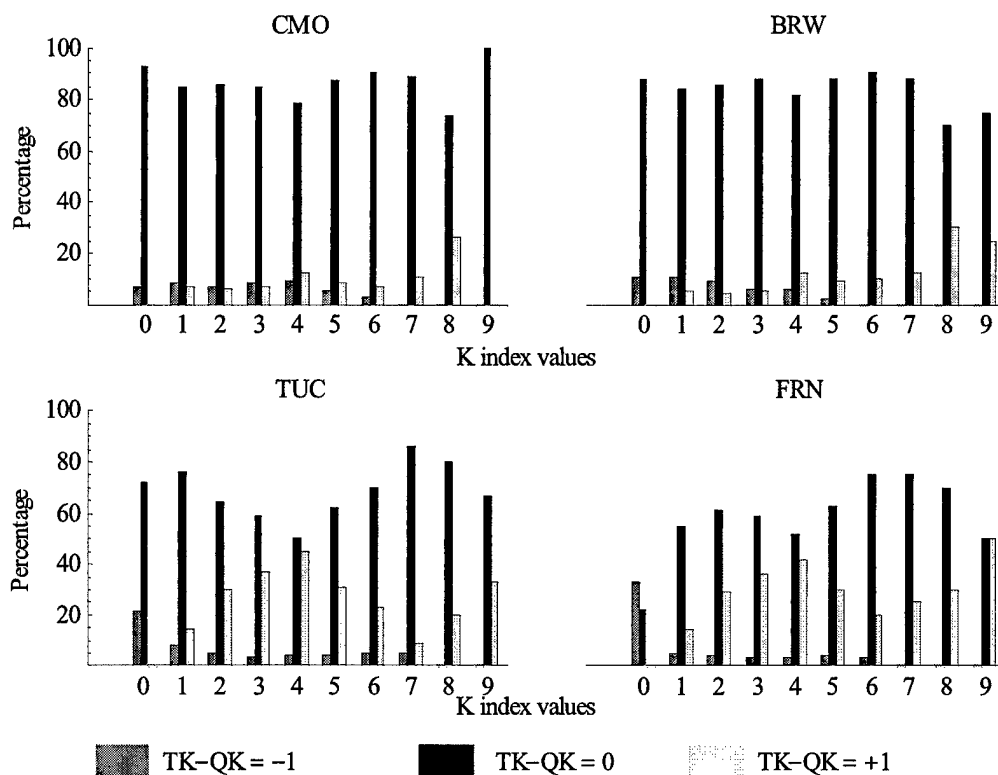
Figure 22 - Performance of the new USGS K index algorithm based on K index category.

4.1.6 Comparison by Hour of Calculation

This test was designed to see if the K index program was removing diurnal effects correctly (recall from chapter two, the Sq has to be removed to compute the K index). If the diurnal effects are removed correctly from the observations, one would expect to see the same proportion of differences for the eight daily computations. On the other hand, if the diurnal variations are not properly removed, then one would expect to see a higher number of differences at a particular time of the K index computation. Figure 23 shows representative results from this test. The asterisk identifies the observatory's local noon K-index computation. In all observatories, the hourly variation is almost equally distributed among the eight daily observations.

58
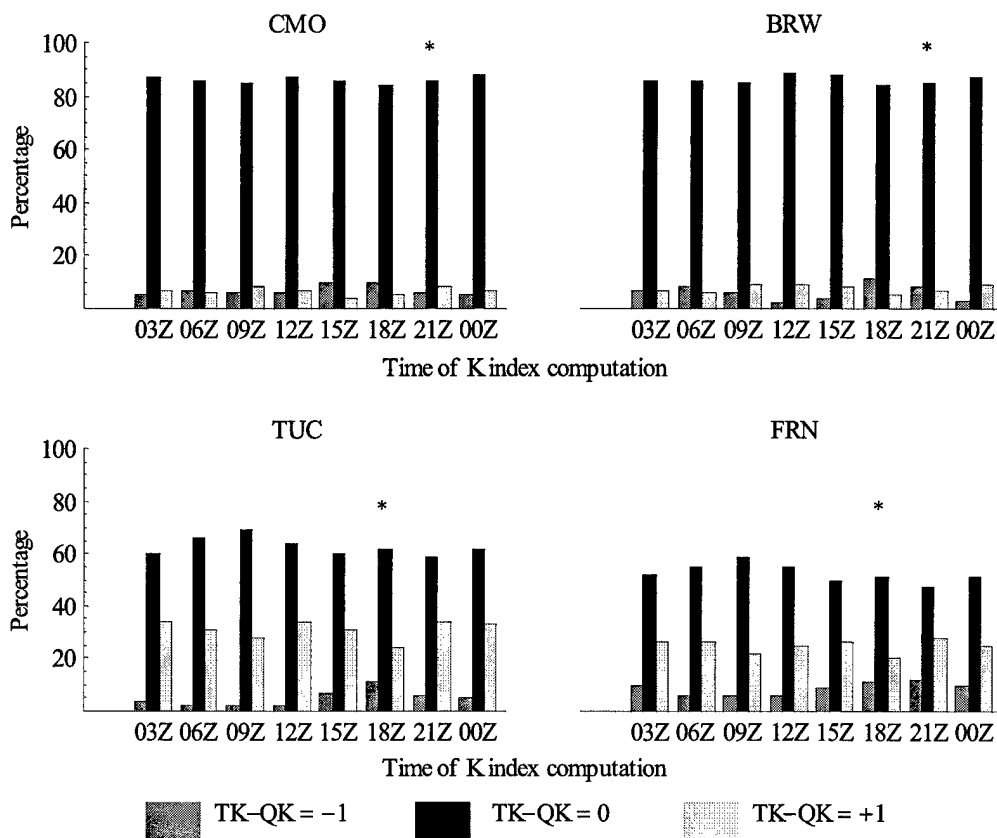
Figure 23 - Performance of the new USGS K index algorithm based on hour of K index computation.

4.1.7 Comparison by Day of Calculation

The amount of data that the Windows-based K index program can work with at any one time is as little as one day or as much as one month. To calculate the K index from 15 Aug to 15 Sep, the August data would be processed first, followed by the September data. On some numerical analysis routines, for example spline fitting, there is usually a problem fitting the endpoints of the curve due to lack of data on one side of the endpoint. This test was designed to see if the program had problems computing the K

index at either the beginning of the month (no data prior to the 1st day) or at the end of the

month (no data after the end of month). If there was a problem, one would expect to see

a difference on the distribution of errors at either the beginning or at the end of the

month. Figure 24 shows the results of this test for College and serve to show that there
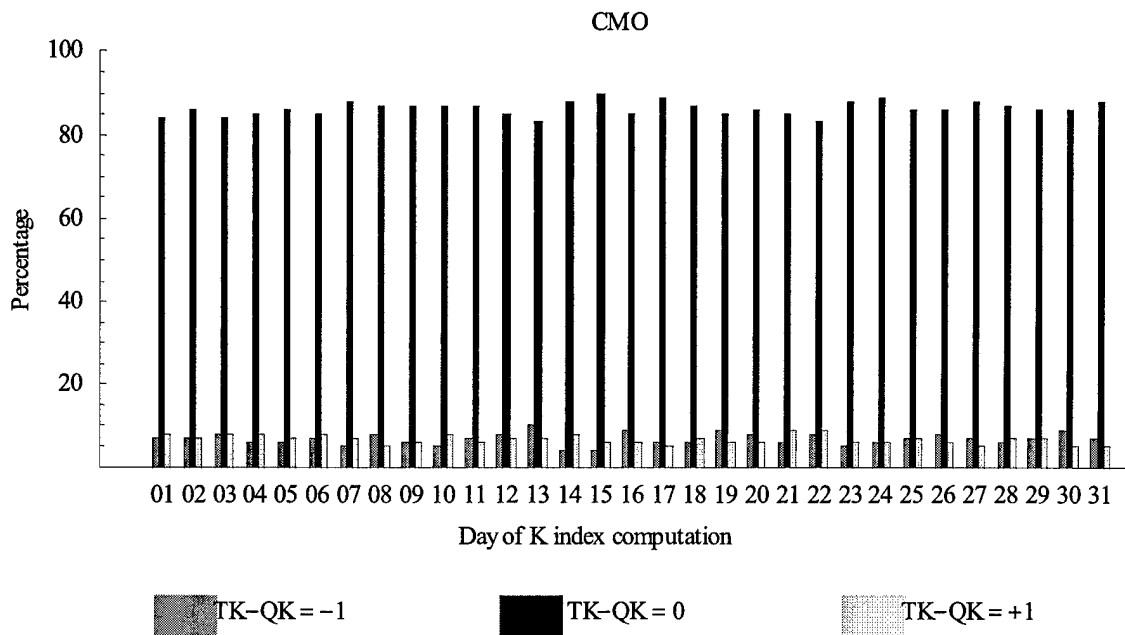
Figure 24 - Performance of the new USGS K index algorithm based on day of K index
computation.

4.1.8 Comparison by Month of Calculation

This test was designed to see if there is any seasonal variability in the

computation of the K index. The results from this test indicate that the error proportion is

distributed equally among all months. Therefore, no seasonal variability was discovered.

Figure 25 shows the results for this test for College. This figure is representative of the

results found at each observatory.
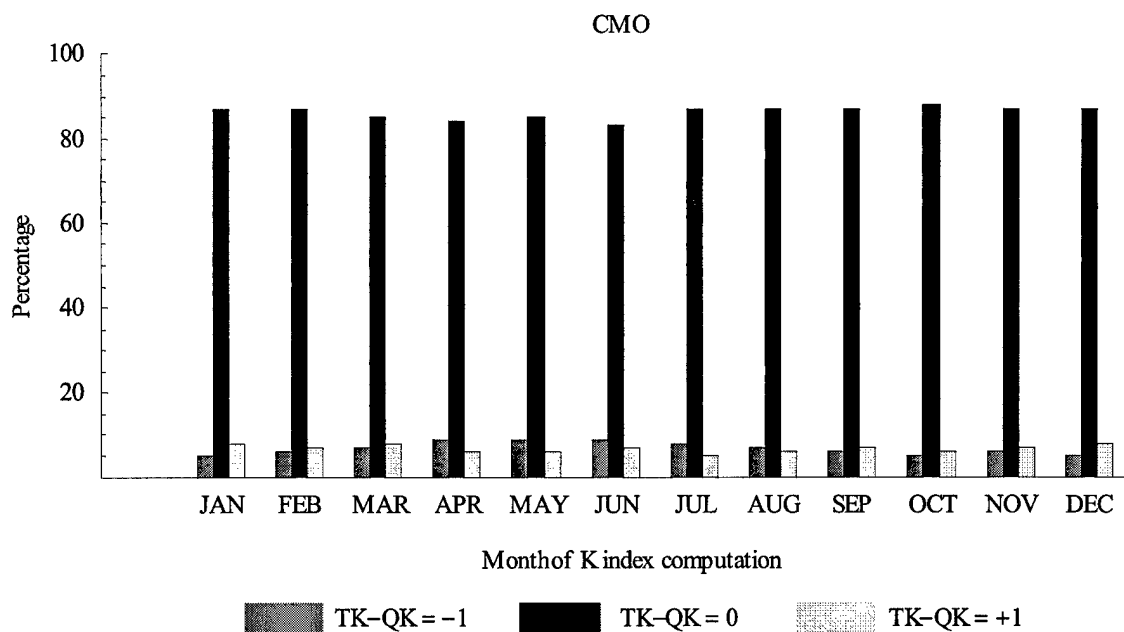
Figure 25 - Performance of the new USGS K index algorithm based on month of K index computation.

4.2 Determining a Correlation Between Stations

The next series of tests were conducted to establish a correlation among the observatories' K index values based on distance. As previously mentioned, the idea behind these tests was that the physical processes causing the surface magnetic deviations will be similar for stations in close proximity. Furthermore, this similarity will decrease as the distance between observatories increases, leading to greater variation on the K index time series. Appendix A lists the Great Circle distances between observatories. The distances were computed using the geographic and the geomagnetic coordinates of the stations. No differences were noted between the geographic and geomagnetic distances separating the observatories. Therefore, all distances used in the project were distance computed using the geographic coordinates of the stations.

As described in chapter 3, four tests were carried out in this phase of the project. The first two tests, time series differences and scatterplots, qualitatively illustrate the index vs. distance relationship. The last two tests were designed to actually compute the RMSE and rank correlation between the observatories.

4.2.1 Difference Plots

This test simply compared the K index time series of two observatories. The time series were generated with the new USGS algorithm and the K index values were truncated prior to computing the difference. For nearby stations, the difference among their K index time series is shown on Figure 26, Figure 27 shows the difference for mid-distance stations, and Figure 28 shows the difference for distant stations. By comparing these three figures, it is easy to see that increasing the distance between the observatories leads to larger differences in the K index time series. These differences are not only in the percentage of the time in which the time series are not an exact match, but also in an increase in the magnitude of K index differences. For nearby stations, the difference (when not exact) is mostly ±1 K index units. As the distance increases, so does the average K index difference between the time series. Appendix C lists all the differences between the observatories.
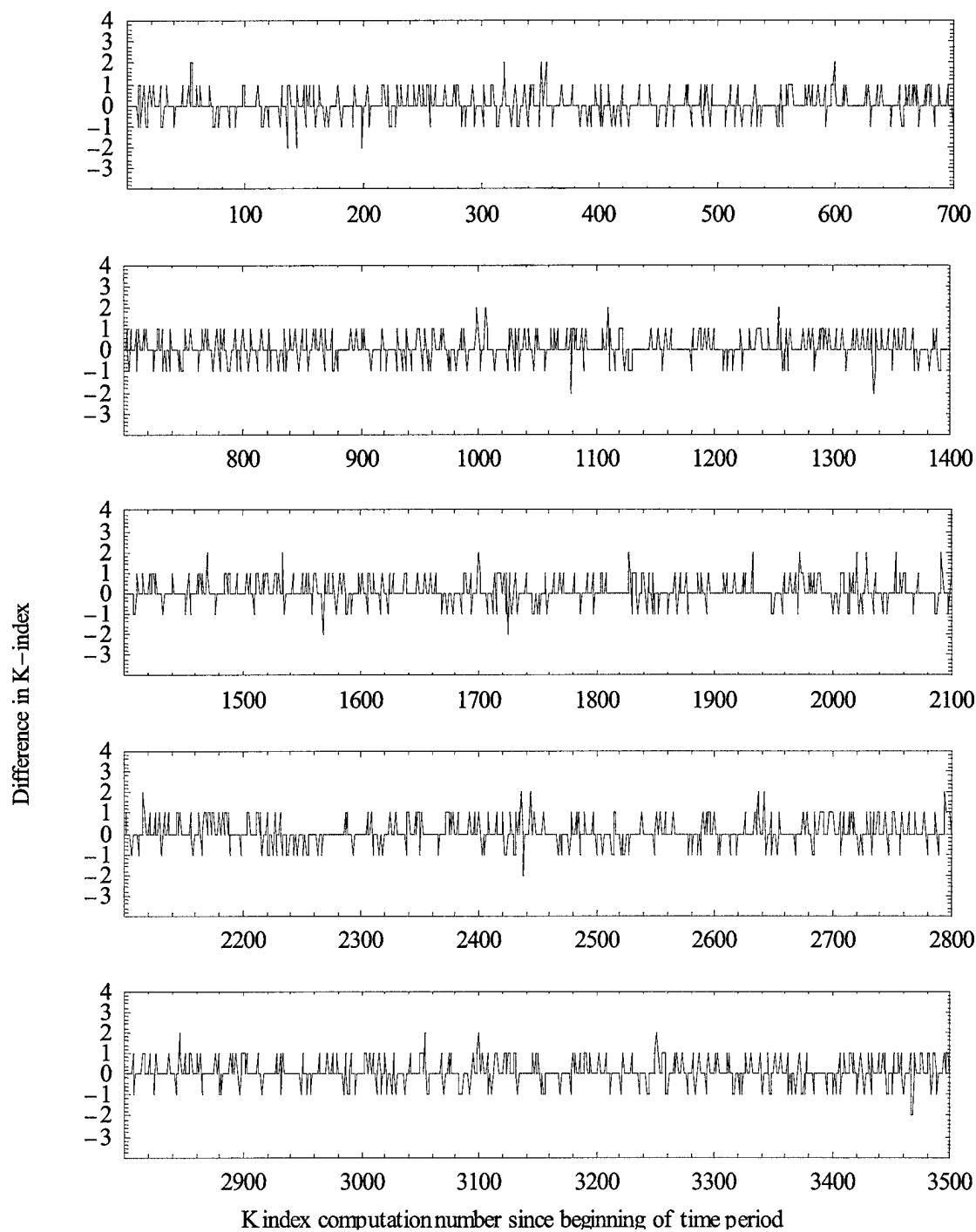
Figure 26 - K index difference between two nearby stations. In this case, Tucson and Fresno were compared from January 1990 to December 1991. The distance between these two stations is 974 Km.
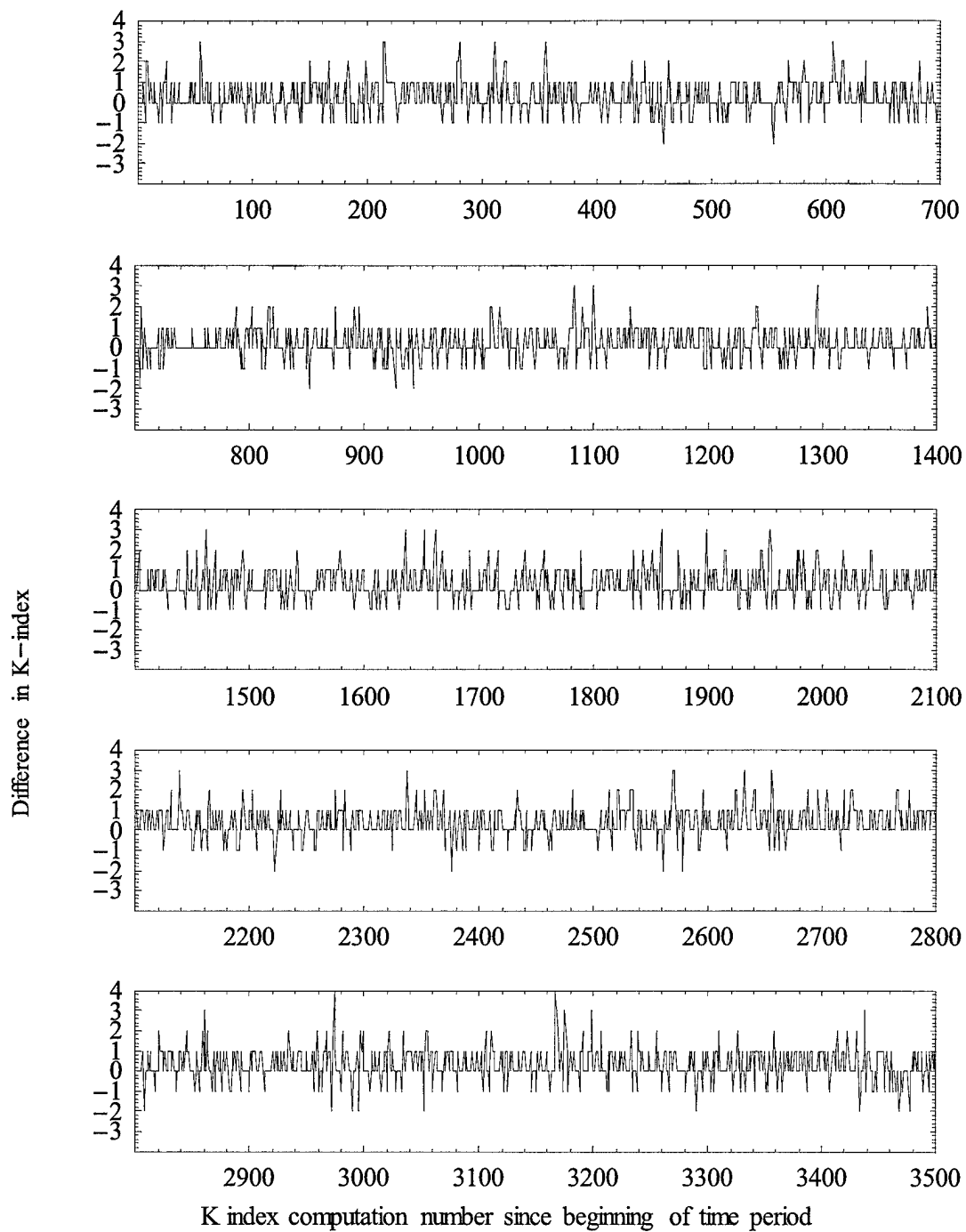
Figure 27 - K index difference between two mid-distance stations. In this case, Tucson and Fredericksburg were compared from January 1990 to December 1991. The distance between these two stations is 3092 Km.
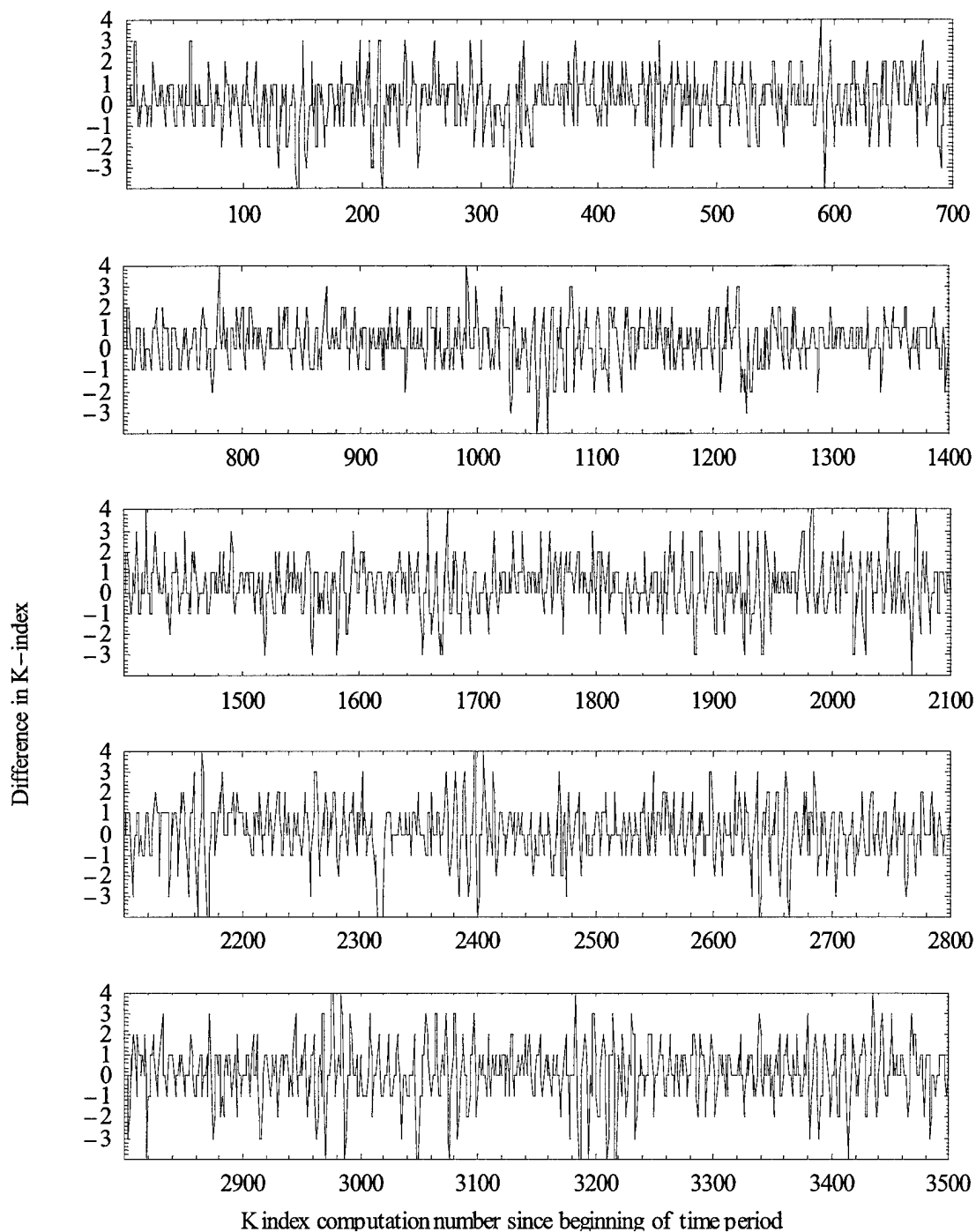
Figure 28 - K index difference between two distant stations. In this case, Tucson and Guam were compared from January 1990 to December 1991. The distance between these two stations is 10500 Km.

4.2.2 Scatterplots

This test used the same stations as the previous one. The scatterplots show a tighter grouping of the data for stations that are close to each other (Figure 29). This grouping spreads as the distance increases (Figure 30 and Figure 31). These findings are consistent with the previous test: distance increases the difference in the K index time series between two stations.

Tucson and Fresno 1990–1994
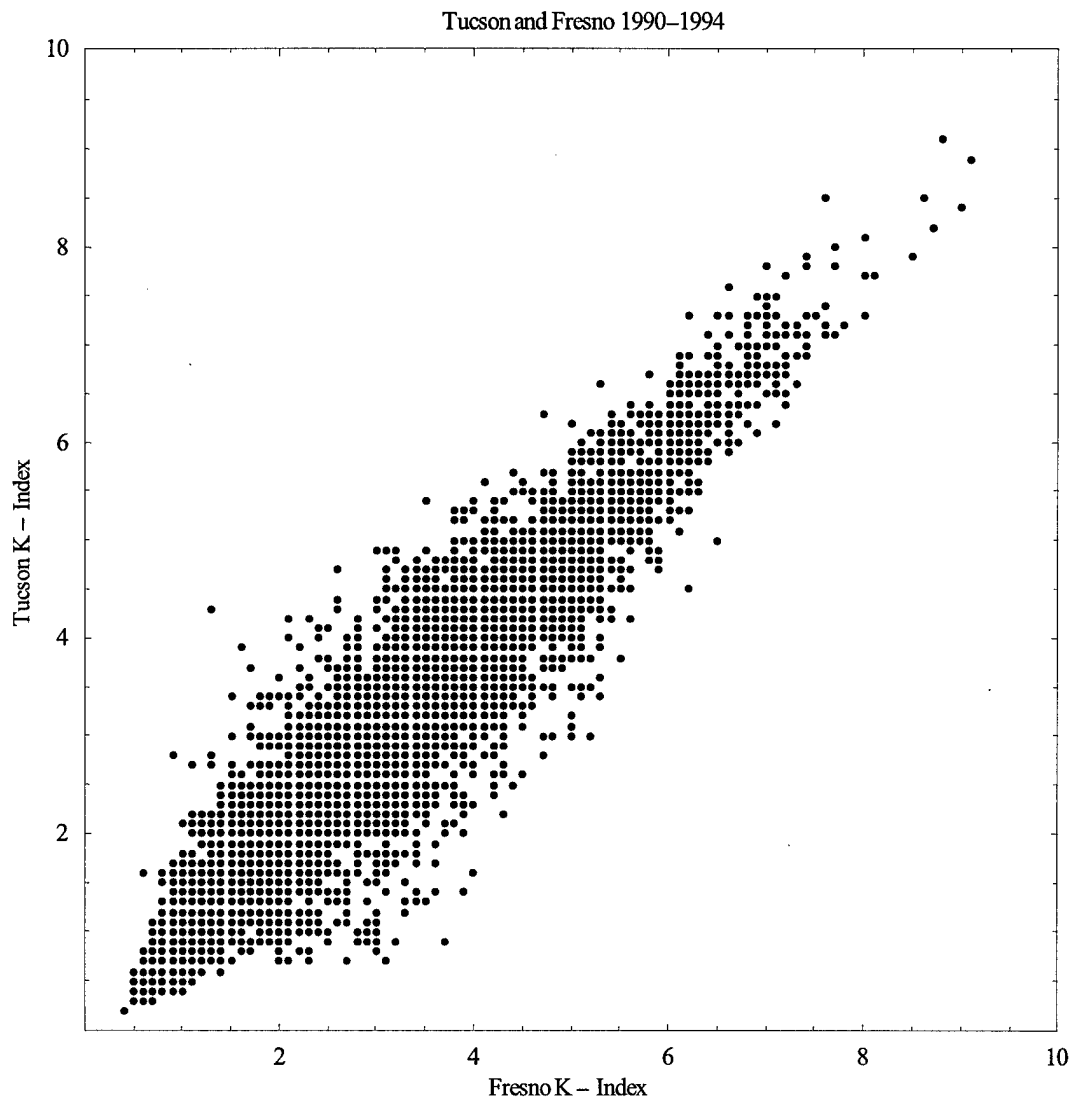
Figure 29 – Scatterplot of two nearby stations. In this case, Tucson and Fresno were compared from January 1990 to December 1994.
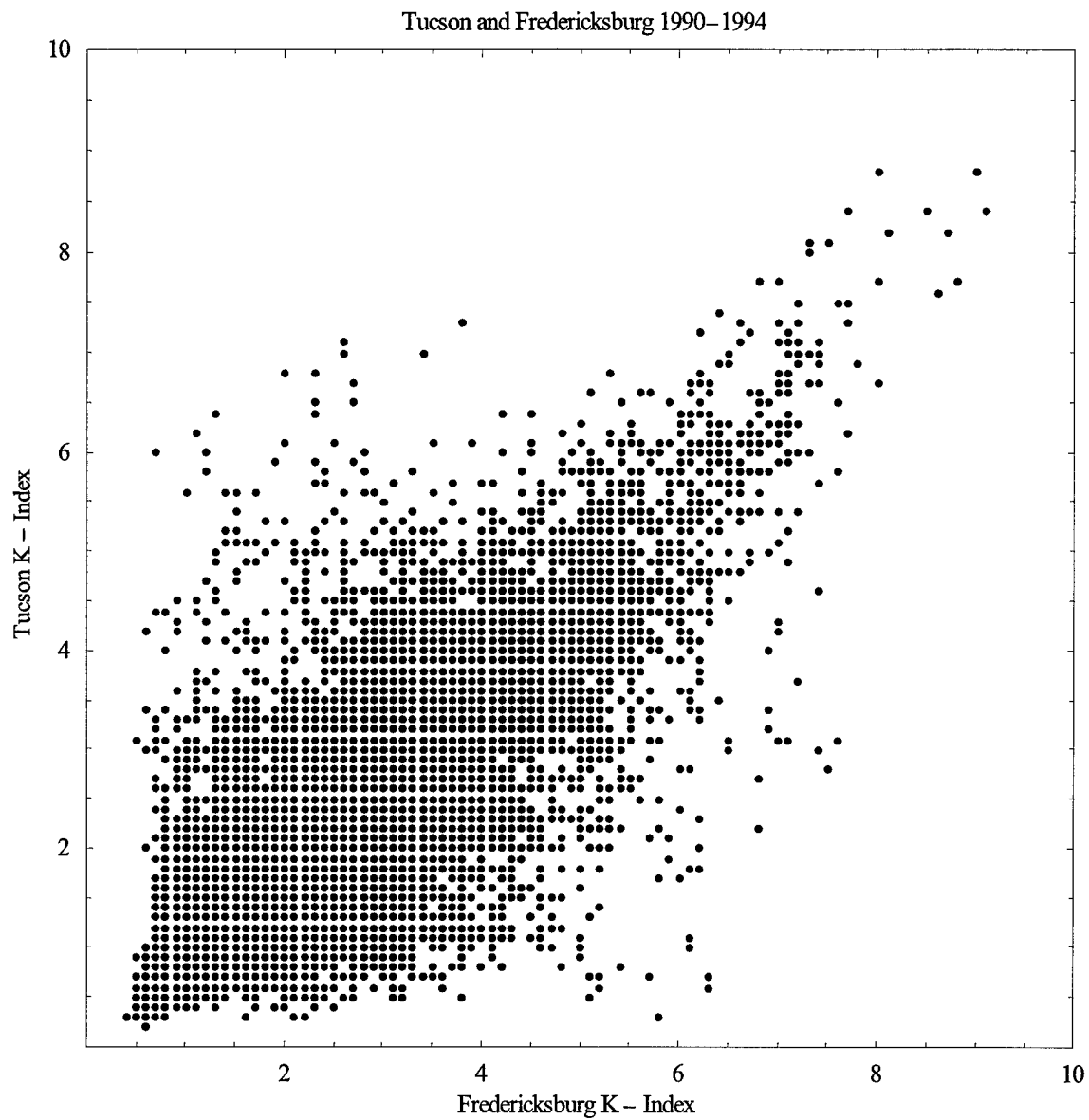
Figure 30 – Scatterplot of two mid-distance stations. In this case, Tucson and Fredericksburg were compared from January 1990 to December 1994.
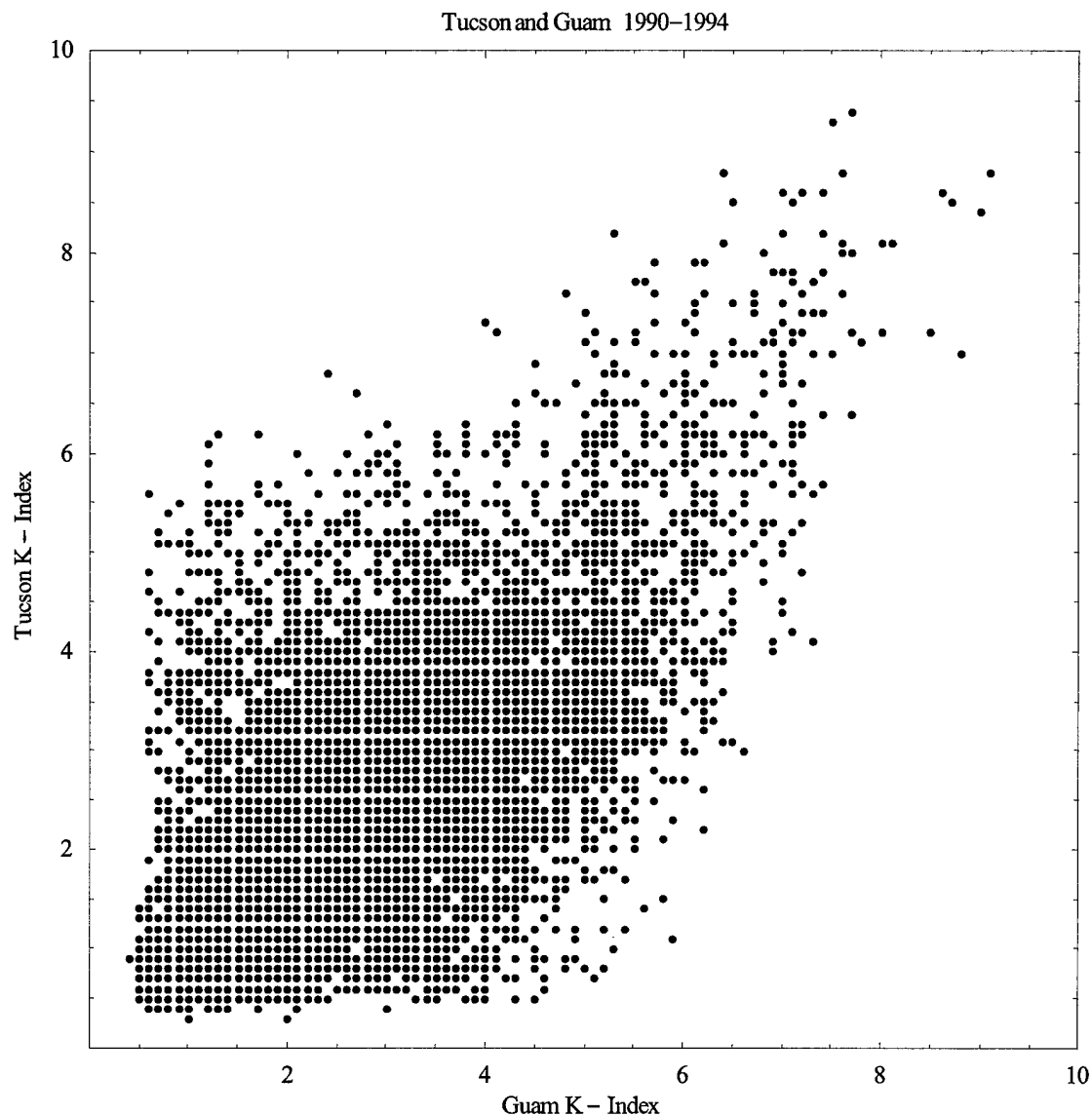
Figure 31 - Scatterplot of two mid-distance stations. In this case, Tucson and Guam were compared from January 1990 to December 1994.

4.2.3 RMSE as function of distance

The previous two tests were qualitative, but helped to illustrate the effects of

distance on the K index time series of two stations. The goal of the next test was to

determine a quantitative relationship between two observatories as a function of distance through the use of the RMSE. As stated before, neighboring stations were expected to have similar trends in their K indices. This similarity was expected to decrease as the distance between the stations increased. Using the RMSE, one would expect a low value for stations that are close to each other and that the RMSE value would increase as a function of distance. The results, shown on figures 32 through 35, confirm this expectation (numerical values are shown in Appendix D). The solid line in these figures is the average value of the RMSE computed by comparing the time series of an observatory with the same time series with ±1 K index error.
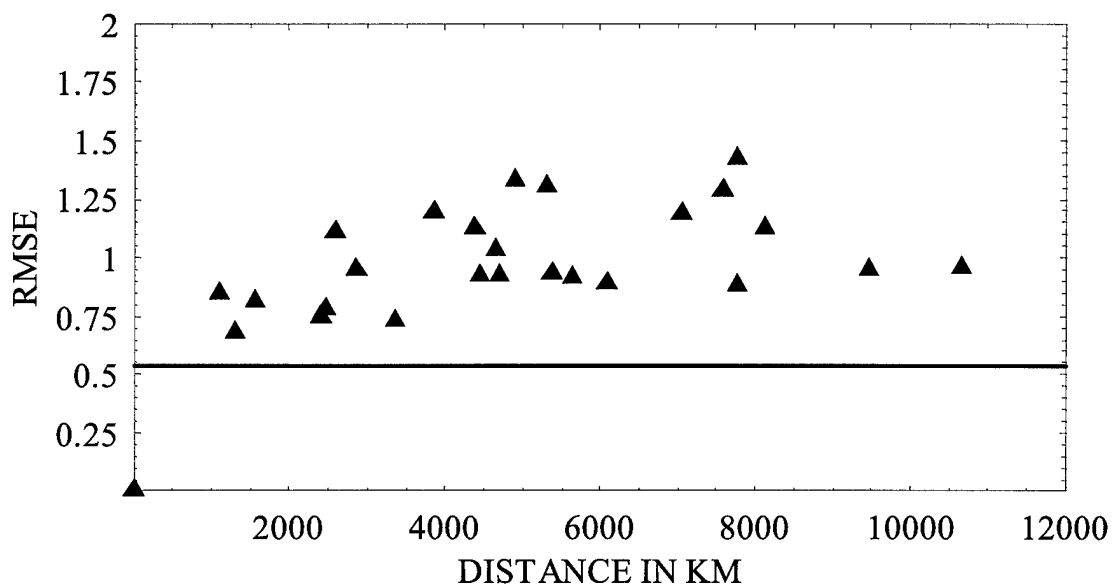
Figure 32 - RMSE using hand-scaled K index time series. The value at the origin is the value when a station is compared to itself.
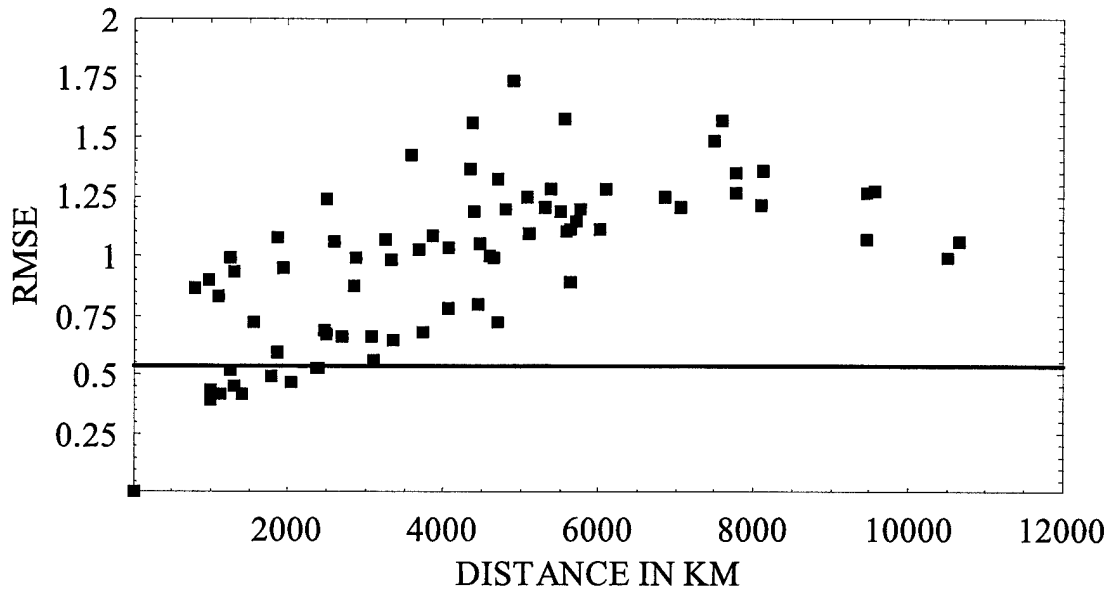
Figure 33 – RMSE using old-Wilson code K index time series. The value at the origin is the value when a station is compared to itself.
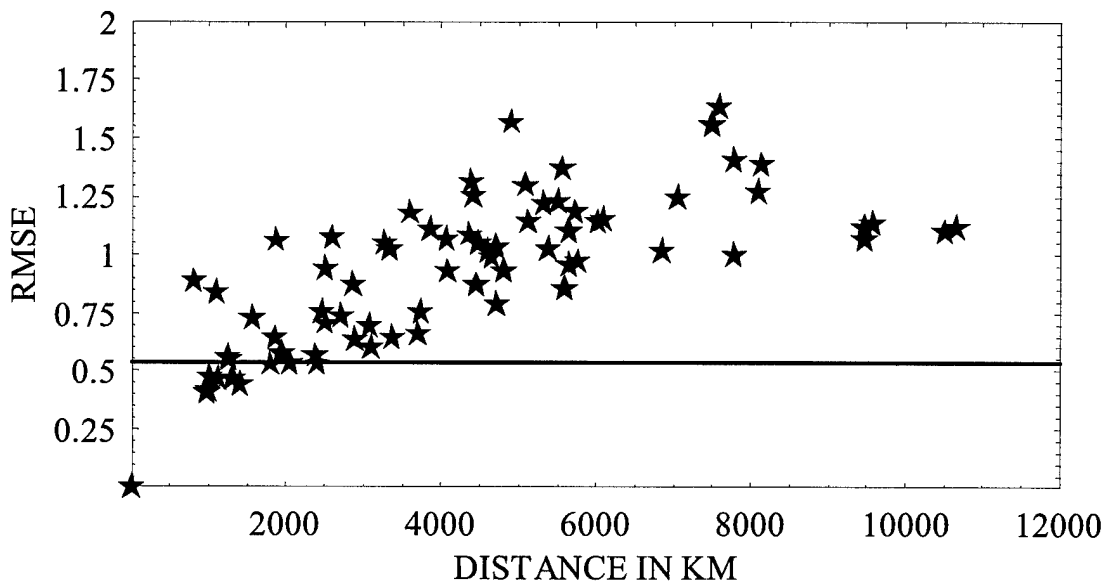
Figure 34 – RMSE using new USGS K index time series. The value at the origin is the value when a station is compared to itself.

70
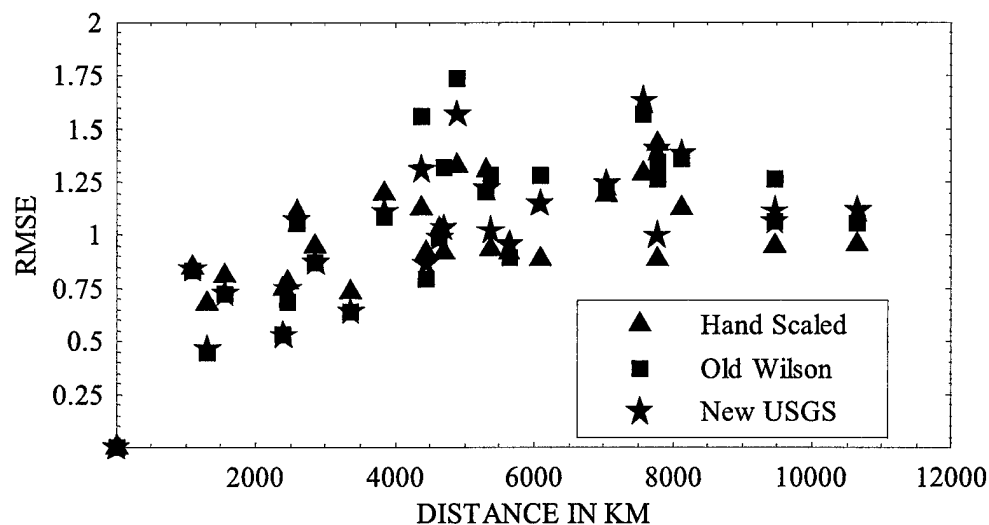
Figure 35 – RMSE using only the observatories that had HK, OK, and NK data for the same time period.

Next, the data generated in this set of tests was fitted using linear and polynomial least-squares fits. Figure 36 shows the linear and polynomial least-squares fits for the new USGS algorithm and Figure 37 shows all three of the polynomial least-squares fits.
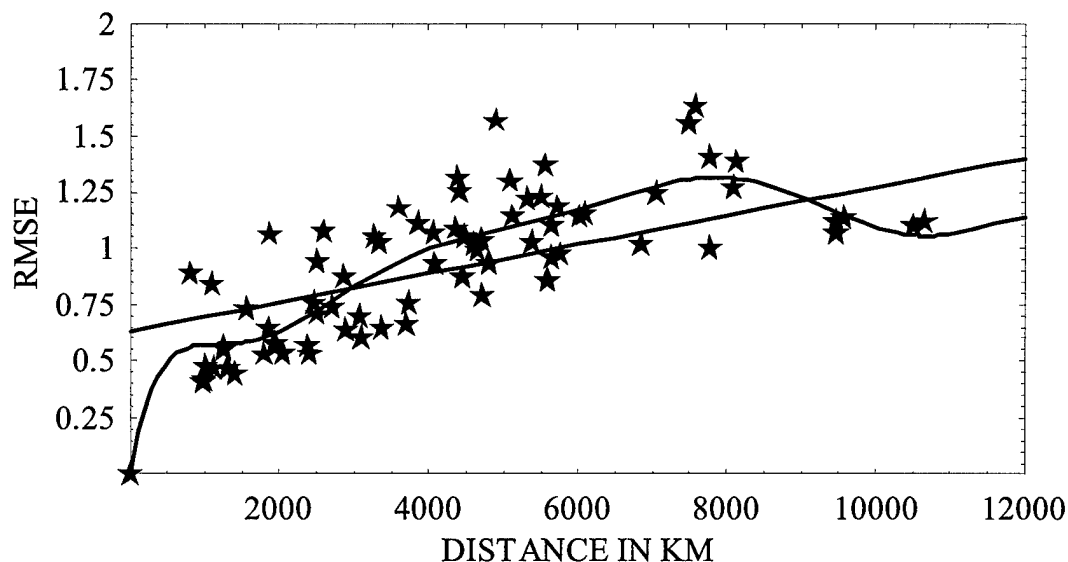
Figure 36 – Polynomial and linear least-square fits of the RMSE values generated with the new USGS algorithm.
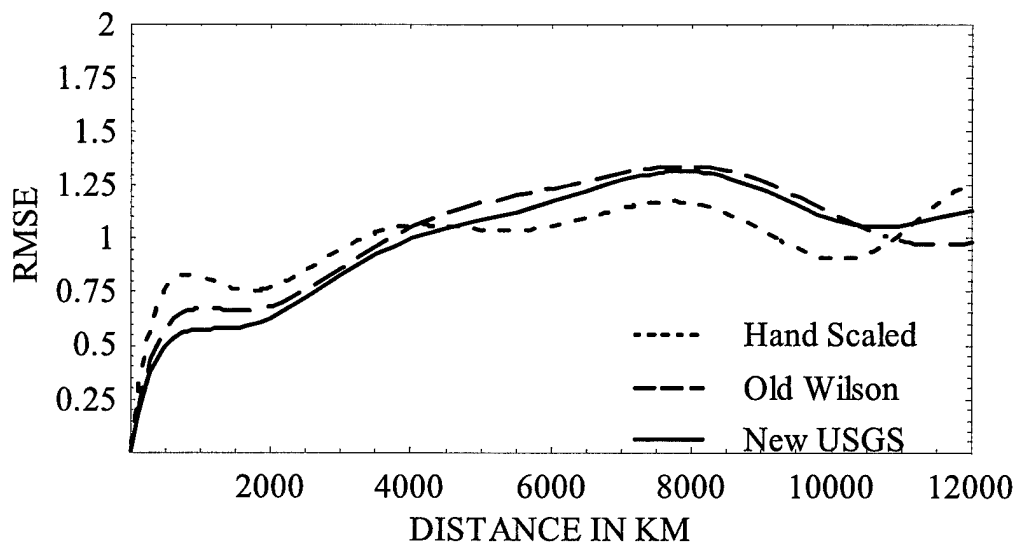
Figure 37 – Polynomial least-squares fits of the hand-scaled, old Wilson, and new USGS algorithm-generated RMSE values.

The same type of tests were run using the average RMSE values computed from 1000 sets of three months' worth of data. The results, shown on Figure 38 and Figure 39, were similar to the ones shown above.
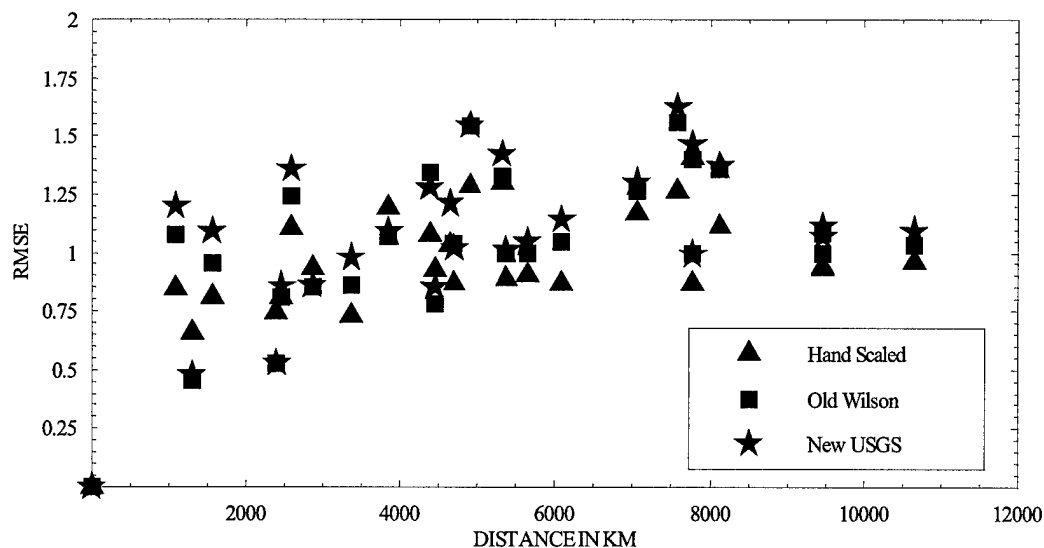
Figure 38 - RMSE using only the observatories that had overlapping data. The values shown are the average values of 1000 runs using 3 months' worth of data.
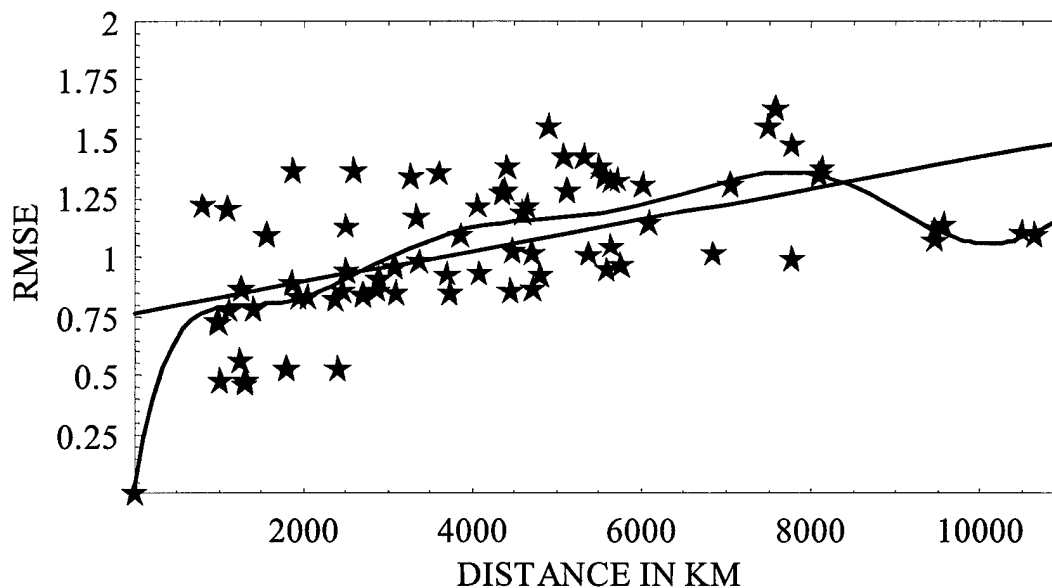
Figure 39 - Polynomial and linear least-square fits of the RMSE values generated with the new USGS algorithm. Data used is the average of 1000 runs using 3 months' worth of data.

Finally, the least-squares fits were re-computed, this time after the removal of some stations. The stations removed were those stations whose exact match between K index computed using the old Wilson method versus the new USGS method fell below 60% during the comparison by observatory location (see section 4.1.4). The stations not considered were Fresno, Guam, Honolulu, and San Juan. One last station, Barrow, was removed solely based on the station's location and possible resulting effects on the K index. This station is close to the polar cap and its K index time series would include effects of the polar cap current. Figures 40 and 41 show the results of the RMSE values increasing with increasing distance after removing the previously mentioned stations.
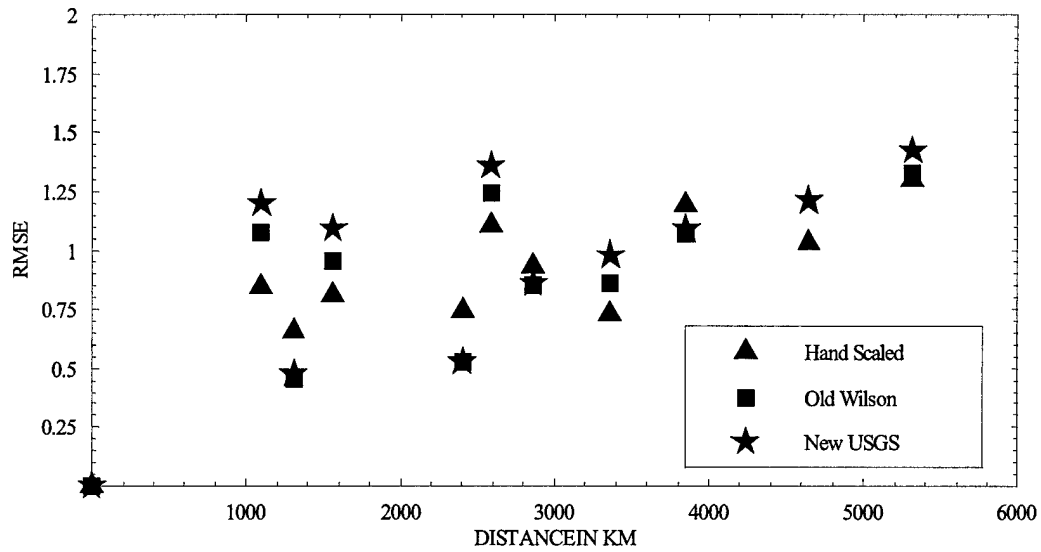
Figure 40 - RMSE using only the observatories that had overlapping data. The data used is the average value of 1000 runs using 3 months' worth of data. Some stations have been removed.
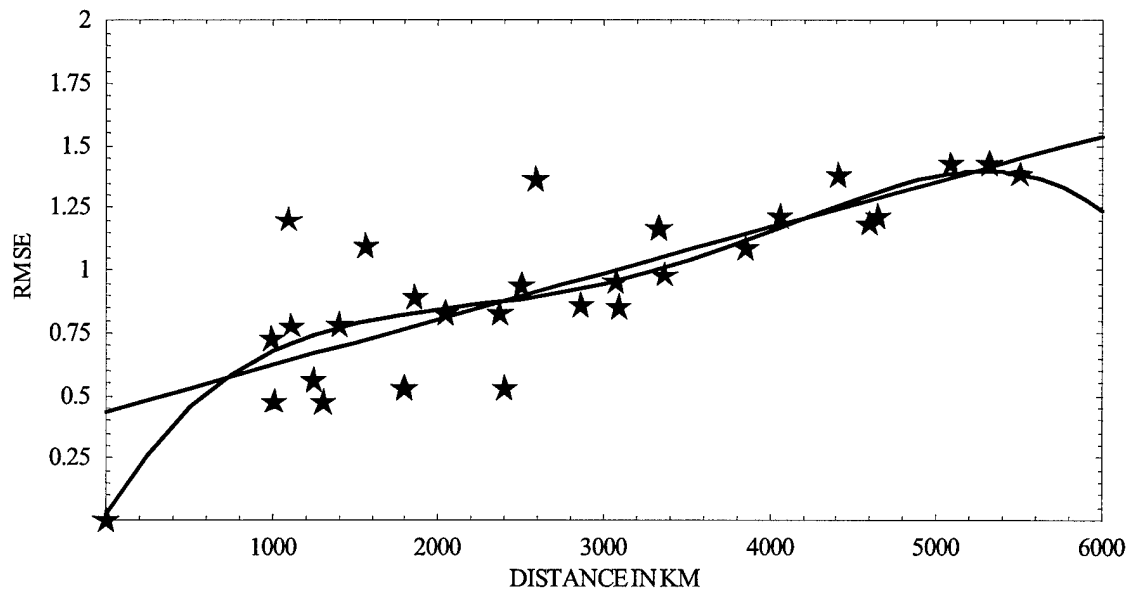
Figure 41 - Polynomial and linear least-square fits of the RMSE values generated with the new USGS algorithm. Data used is the average of 1000 runs using 3 months' worth of data. Some stations have been removed.

All data fits were evaluated for goodness of fit. The results are listed in Table 11.

Overall the polynomial least-squares fits were better than the linear fits. Decreasing the

amount of data by discarding questionable stations improved all data fits (larger R^2

values). All least-squares fit equations were later used to evaluate the K index at BLO.

Table 11 – Goodness of fit data for the RMSE values.

Type	Data	Using all data available		Using 3-month averages		Using 3-month averages and stations removed	
		R^2	RMSE*	R^2	RMSE*	R^2	RMSE*
Linear	Hand-scaled	.263	.230	.259	.227	.646	.197
Linear	Wilson	.413	.258	.338	.232	.657	.184
Linear	USGS	.528	.218	.305	.247	.627	.207
Polynomial	Hand-scaled	.696	.143	.688	.147	.837	.134
Polynomial	Wilson	.594	.215	.547	.192	.691	.174
Polynomial	USGS	.699	.174	.511	.207	.669	.195

* - Please note that the RMSE heading indicates the RMSE between the data and the least square fits and not the RMSE computed by comparing the different time series.

4.2.4 Rank correlation as a function of distance

The same procedures described in Section 4.2.3 were carried out using the rank

correlation values. One main difference needs to be pointed out between the two sets of

tests. While the RMSE values were expected to increase with increasing distance, the

rank correlation values were expected to *decrease* with increasing distance. This proved

to be the case as shown on figures 42 through 44. The solid horizontal line in Figure 42

is the average value of the rank correlation coefficients computed for each observatory by

comparing the time series of an observatory with the same time series ± 1 K index

category error. The "error" was created with a random number generator. The average

correlation coefficient for this "random error" time series and the real time series was

0.9847.
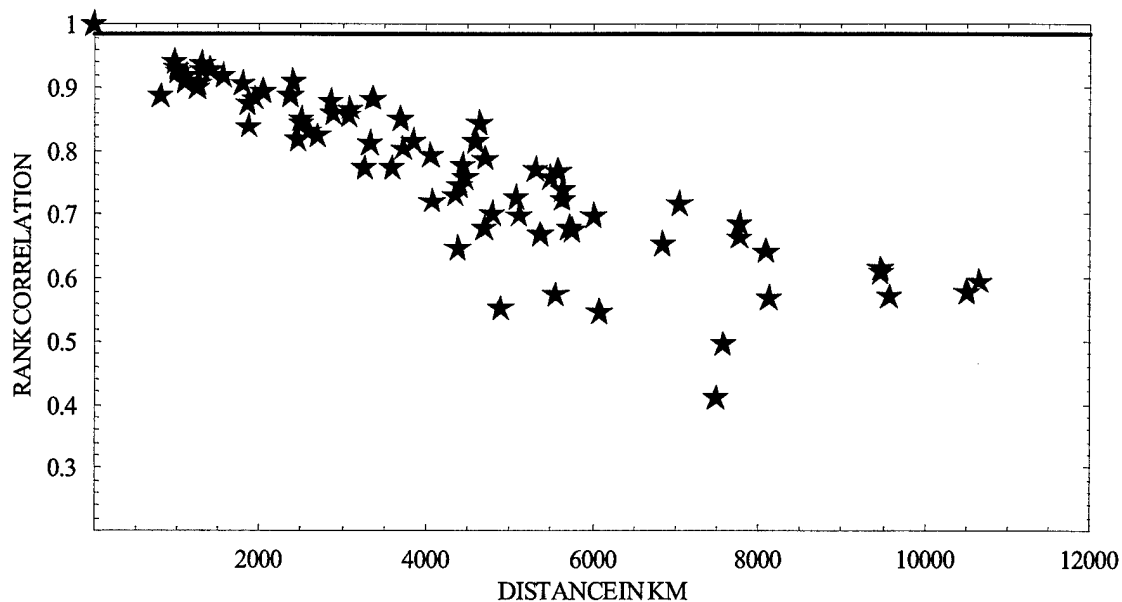
Figure 42 – Rank correlation coefficients using the new USGS K index time series.
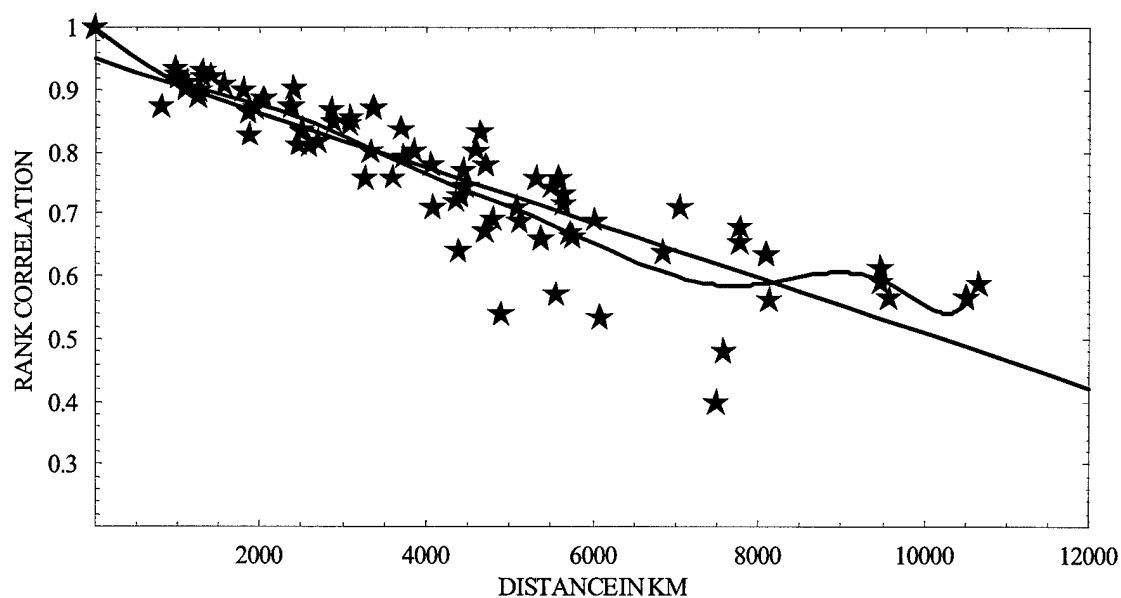
Figure 43 - Polynomial and linear least-square fits of the rank correlation values generated from the new USGS algorithm K index time series. Data used is the average of 1000 runs using 3 months' worth of data.
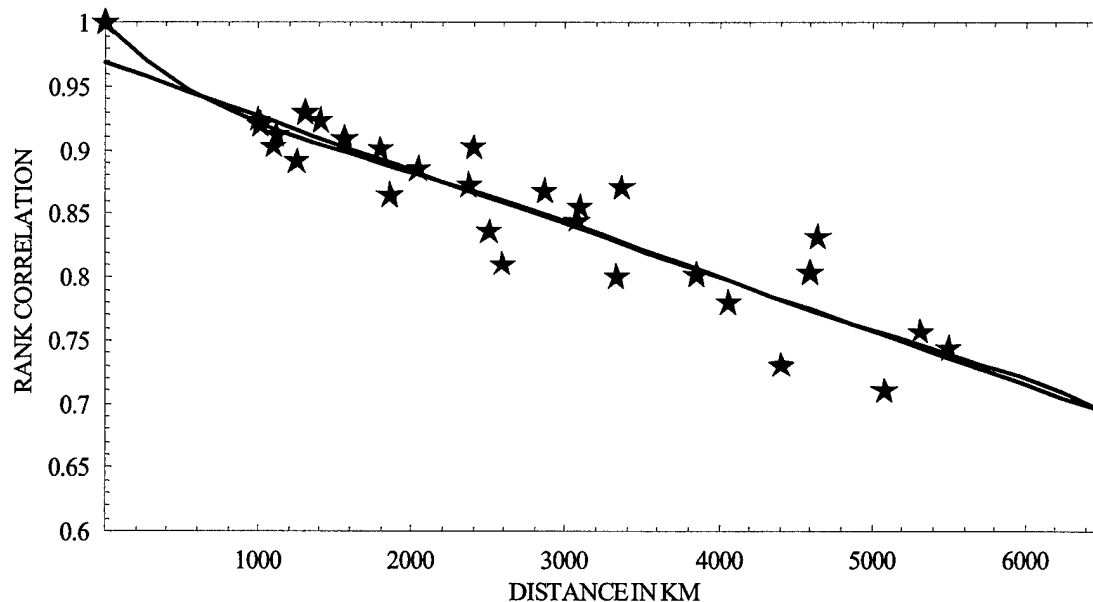
Figure 44 - Polynomial and linear least-square fits of the rank correlation values generated from the new USGS algorithm K index time series. Data used is the average of 1000 runs using 3 months' worth of data. Please refer to section 4.2.3 for a list of stations removed.

The goodness of fit for all data fits are shown on Table 12. Note that the rank correlation values were fitted much better than the RMSE values (larger R^2 values and smaller RMSE values).

Table 12 - Goodness of fit data for the rank correlation values.

Type	Data	Using all data available		Using 3-month averages		Using 3-month averages and stations removed	
		R^2	RMSE[*]	R^2	RMSE[*]	R^2	RMSE[*]
Linear	Hand-scaled	0.657	0.043	0.626	0.045	0.824	0.032
Linear	Wilson	0.533	0.097	0.746	0.060	0.863	0.023
Linear	USGS	0.768	0.060	0.767	0.060	0.843	0.027
Polynomial	Hand-scaled	0.882	0.025	0.855	0.028	0.912	0.022
Polynomial	Wilson	0.599	0.089	0.797	0.054	0.867	0.023
Polynomial	USGS	0.810	0.054	0.810	0.055	0.849	0.026

[*]-Please note that the RMSE heading indicates the RMSE between the data and the least square fits and not the RMSE computed by comparing the different time series.

4.3 Evaluating the K Index for BLO

The last set of tests in this project was to evaluate the K index at BLO. These tests included checking the BLO magnetometer data, computing the K index for BLO, and finally computing the RMSE and rank correlation values between BLO and other stations.

4.3.1 BLO magnetometer data

The BLO magnetometer data were compared to the data from Newport, Boulder, Fresno, Del Rio, and Tucson for a period of sixty days. Figures 45 through 47 show the results of one day's worth of data. As expected, the magnetometer data for Tucson was similar to the geomagnetic trace of the "bracketing" stations (Fresno and Del Rio). However, the BLO data were not similar to the traces for Newport and Boulder.
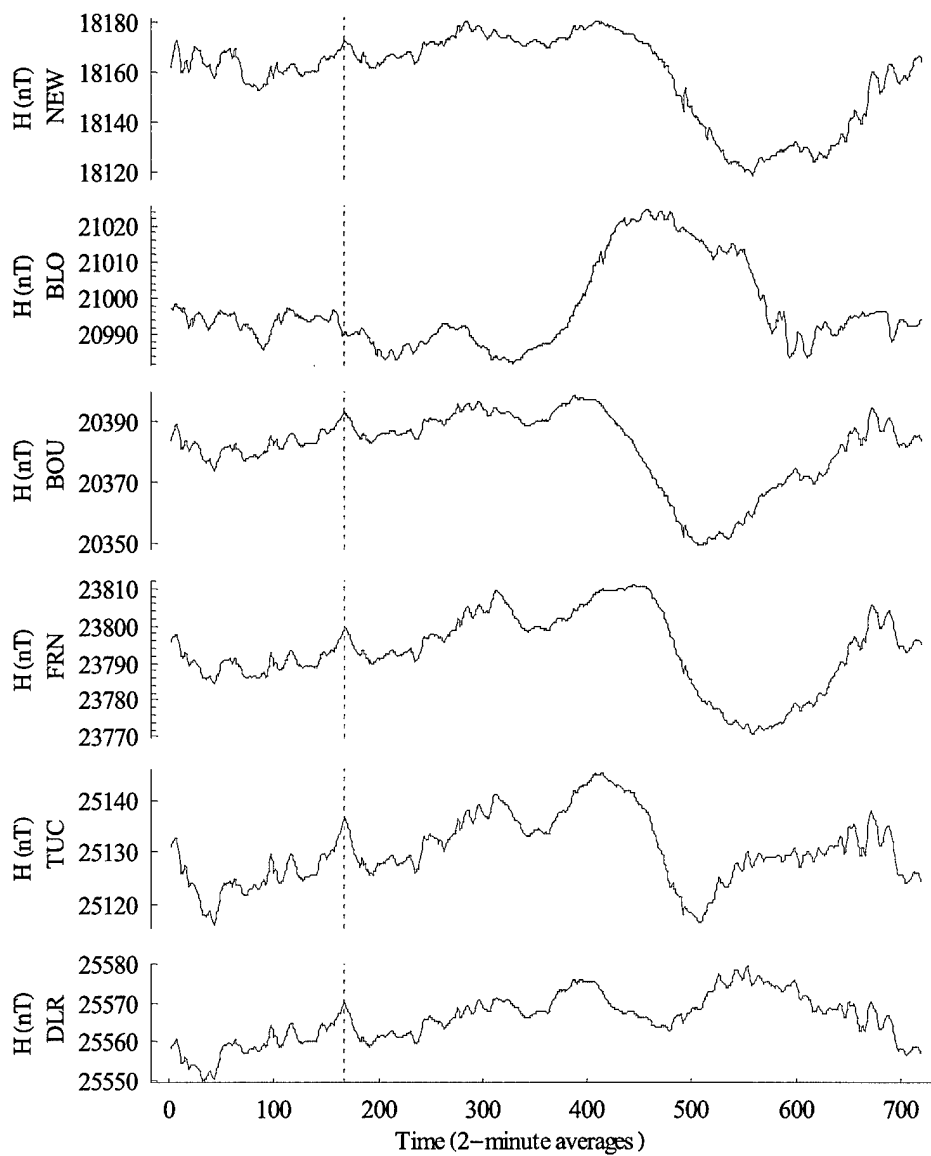
Figure 45 – Magnetometer trace of the H component for 14 Aug 99. The dashed line in all six traces is an arbitrary marker used to line up features among the six traces.
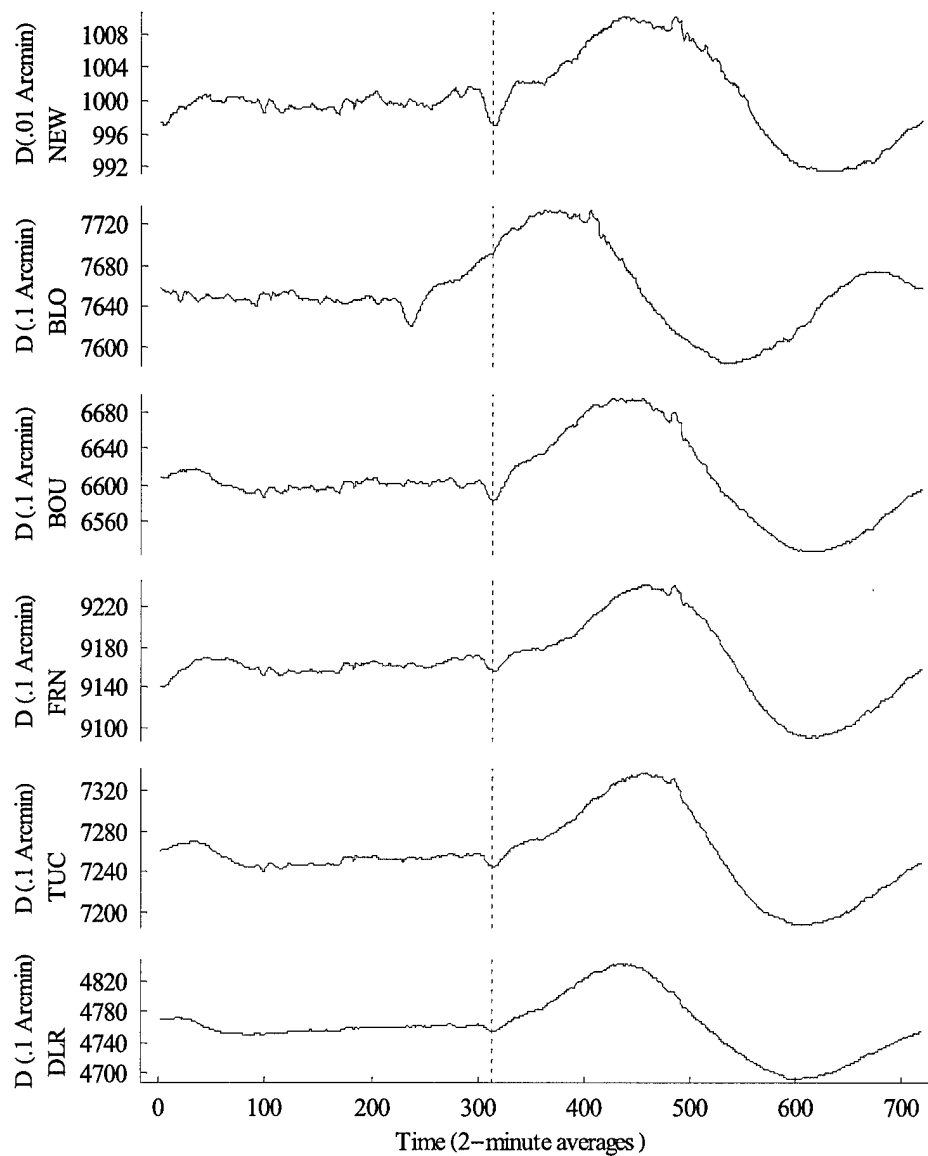
Figure 46 - Magnetometer trace of the D component for 14 Aug 99. The dashed line in all six traces is an arbitrary marker used to line up features among the six traces.
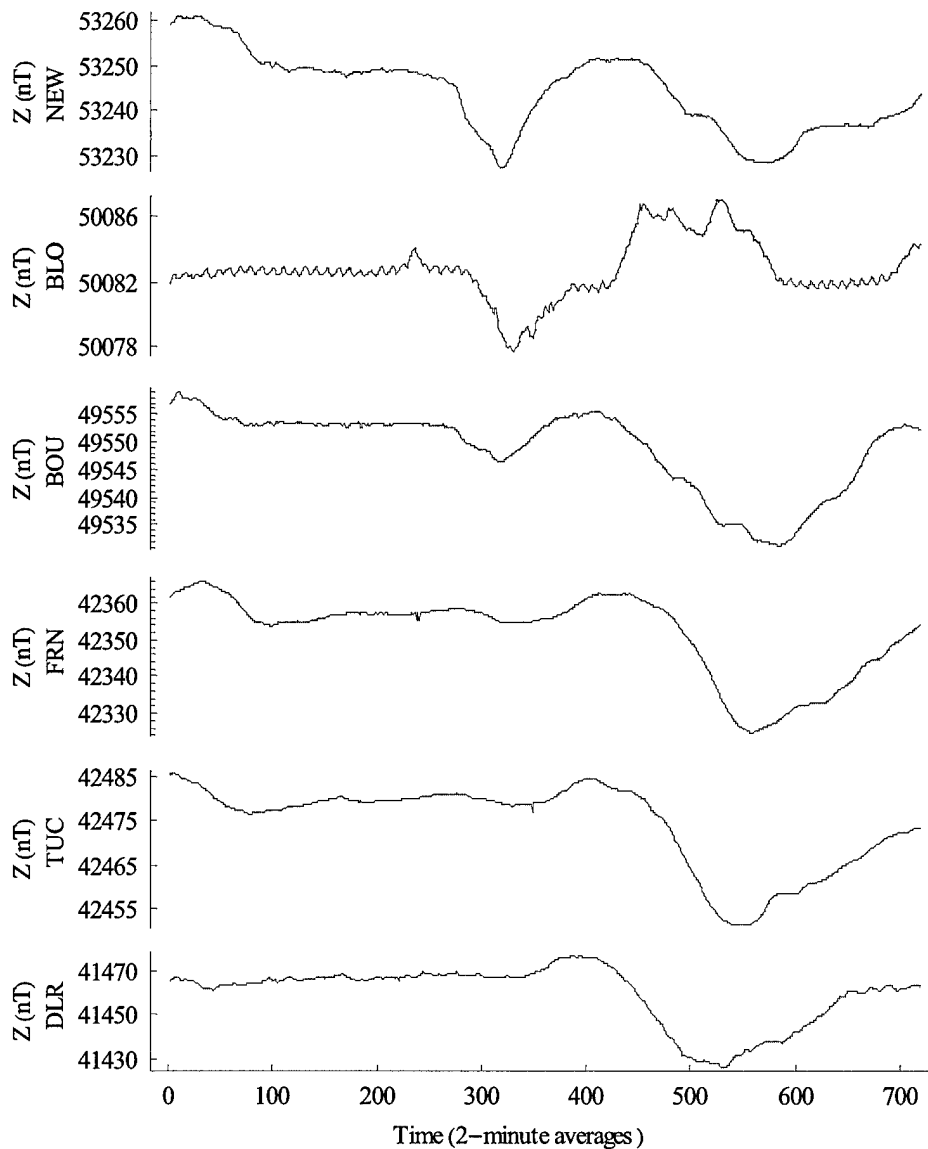
Figure 47 – Magnetometer trace of the Z component for 14 Aug 99. The dashed line in all six traces is an arbitrary marker used to line up features among the six traces.

After observing several days' worth of data, a pattern was noticed in the D component of the BLO magnetometer trace. This component seemed to be out of phase with the other observatories. By lining up geomagnetic features on over 60 days' worth of traces, it was determined that the BLO data were 141 minutes out of phase (slower)

than the other 5 observatories. All data files were reconstructed to account for the time shift needed to line up the features. After this correction, the BLO patterns in the D component of the magnetometer traces were similar to the other 5 observatories (Figure 48).
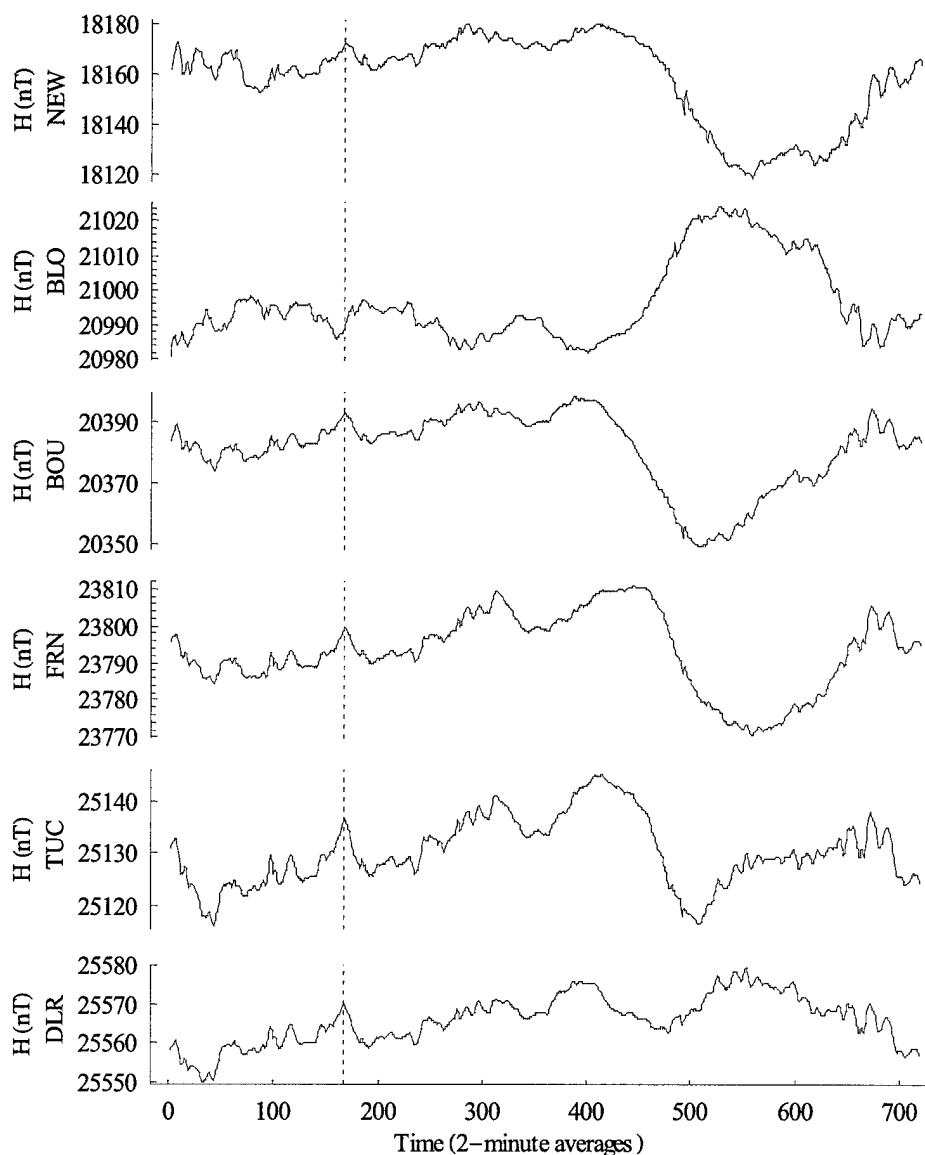
Figure 48 – Traces of the D component for the six observatories on 14 Aug 99. The BLO data were shifted forward 141 minutes.

After accounting for the time error, the traces of the H (Figure 49) and Z

components for BLO were the mirror image of the other five observatories.
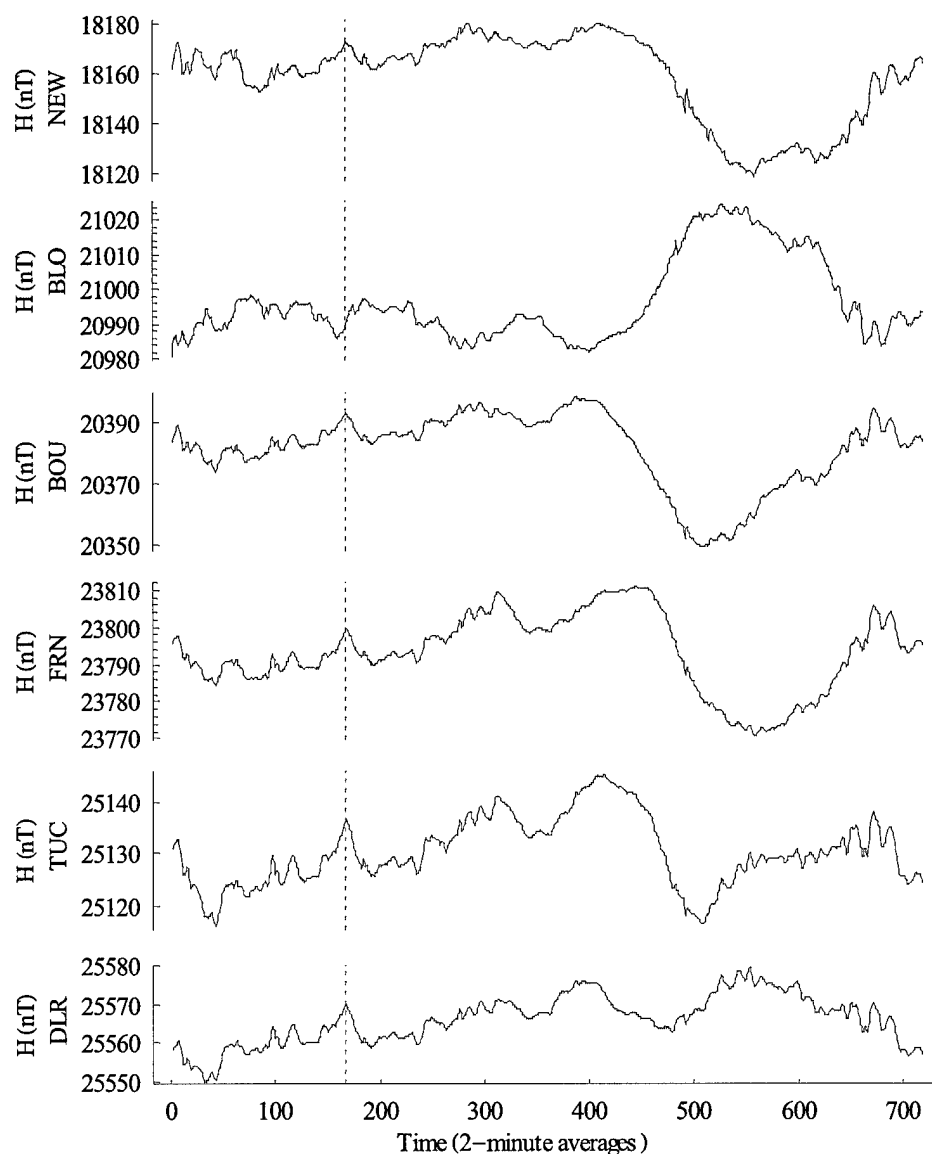
Figure 49 - Traces of the H component for the six observatories on 14 Aug 99. The BLO
data were shifted 141 minutes.

However, there was still a problem with the BLO H component. Either the BLO

trace or the other 5 traces were "upside-down". To verify which was the correct trace,

the Sq H curve was plotted, using a program mentioned in Campbell (1997), for several

83

days and compared to the traces of the six observatories. This proved that the BLO data were incorrect. The BLO data had to be corrected to account for this "mirroring" effect. The procedure to correct the H and Z components of the data were simple:

1. Compute the mean for the component for a given day,

2. Subtract the mean from the 1.0 minute-average values.

3. Subtract the mean from these new values.

The results were plotted in Figure 50. The corrected H component trace of BLO was then similar to that of the other five observatories. The corrections to the BLO data were necessary and satisfactory for this project. The BLO was notified about the data problem and are currently working on a solution to the problem. The source of the error could be from either the magnetometer or from the algorithm used to calibrate the data.
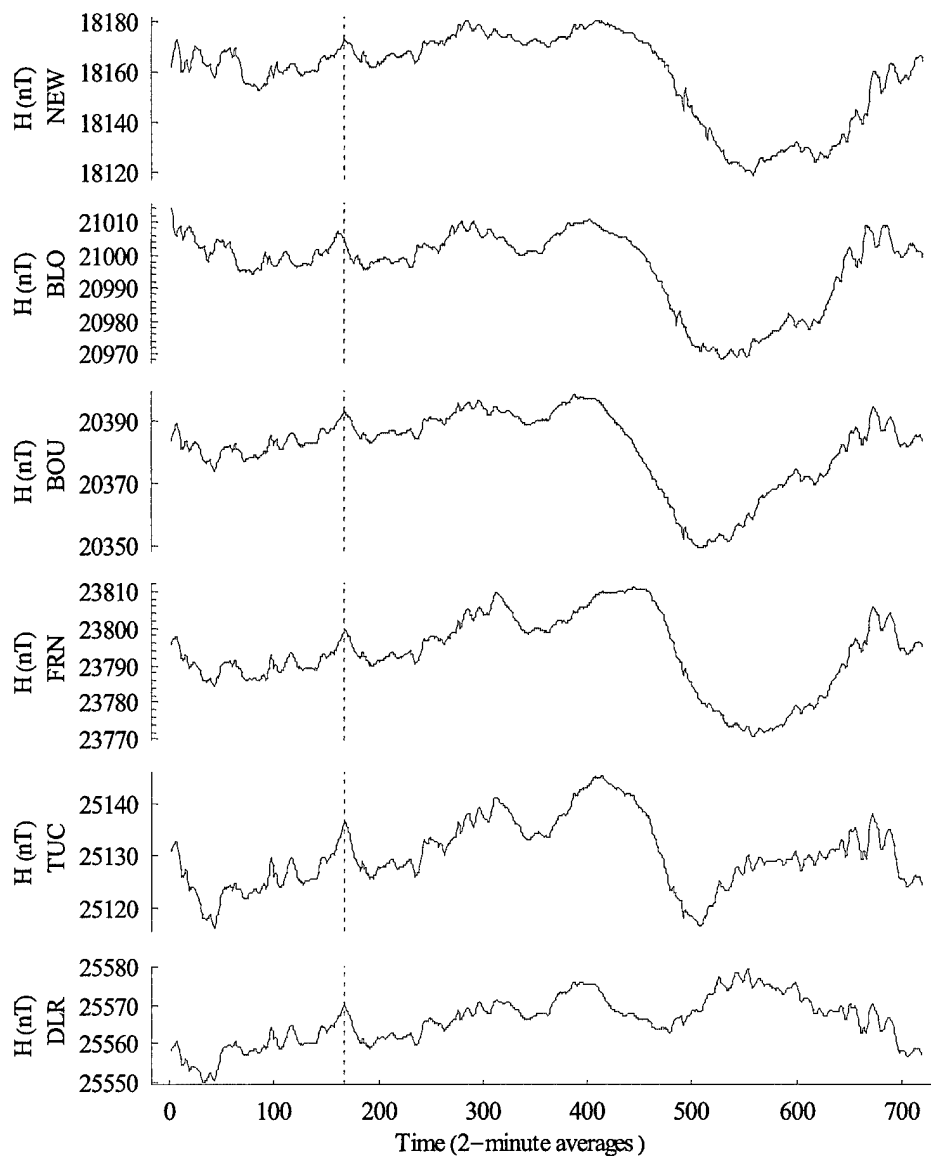
Figure 50 - Traces of the H component for the six observatories on 14 Aug 99. The corrected BLO H trace was shifted forward 141 minutes and "flipped" about the mean.

4.3.2 Computing and evaluating the K index at BLO

Once the BLO data were corrected, the K index was computed using the new

USGS program. But before the actual RMSE and rank correlation values between BLO,

BOU, and NEW were computed, the expected values based on distance between the

85

stations were calculated. This was done using all the available equations from the least-squares fits done in the sections 4.2.3 and 4.2.4. Once all the expected values were computed, the average and standard deviation for each station pair was computed. The results of the expected values are shown on tables 13 and 14.

Table 13 – Expected RMSE values based on distance between the observatories and the various least squares fits.

Type Fit	Data	Amount	BLO-BOU 792 Km	BLO-NEW 738 Km
Linear	Hand	All	0.8270	0.8252
		3 Month	0.5277	0.5182
		Select	0.5276	0.5182
	Old Wilson	All	0.7655	0.7626
		3 Month	0.5256	0.5163
		Select	0.5256	0.5163
	New USGS	All	0.6797	0.6762
		3 Month	0.5831	0.5732
		Select	0.5831	0.5731
Polynomial	Hand	All	0.8319	0.8288
		3 Month	0.5414	0.5196
		Select	0.5414	0.5196
	Old Wilson	All	0.6562	0.6476
		3 Month	0.4890	0.4104
		Select	0.4896	0.4704
	New USGS	All	0.5603	0.5537
		3 Month	0.5414	0.5196
		Select	0.5414	0.5196
Average			0.5965	0.5816
Standard Deviation			0.1095	0.1190
±1 Standard Deviations			0.4870-0.7060	0.4626-0.7006
±2 Standard Deviations			0.3775-0.8155	0.3436-0.8196
±3 Standard Deviations			0.2680-0.9250	0.2246-0.9386

Table 14 – Expected rank correlation values based on distance between the observatories and the various least squares fits.

Type Fit	Data	Amount	BLO-BOU 792 Km	BLO-NEW 738 Km
Linear	Hand	All	0.8681	0.8693
		3 Month	0.8618	0.86229
		Select	0.9114	0.9139
	Old Wilson	All	0.8975	0.8996
		3 Month	0.9224	0.9247
		Select	0.9479	0.9501
	New USGS	All	0.9234	0.9258
		3 Month	0.9146	0.9170
		Select	0.9349	0.9372
Polynomial	Hand	All	0.9387	0.9477
		3 Month	0.9258	0.9340
		Select	0.8988	0.9031
	Old Wilson	All	0.8992	0.9023
		3 Month	0.9258	0.9282
		Select	0.9423	0.9449
	New USGS	All	0.9331	0.9366
		3 Month	0.9263	0.9300
		Select	0.9320	0.9352
Average			0.9168	0.9201
Standard Deviation			0.0239	0.02482
±1 Standard Deviations			0.8929-0.9407	0.8953-0.9449
±2 Standard Deviations			0.8690-0.9646	0.8705-0.9697
±3 Standard Deviations			0.8451-0.9885	0.8456-0.9946

The K index time series for BLO, BOU, and NEW were used to compute an RMSE and rank correlation between BLO-BOU and BLO-NEW. Data availability was again the driving factor in deciding which station to use for this comparison. The actual computed RMSE and rank correlation values for the three pairs of stations are listed in Table 15.

Table 15 – Actual RMSE and rank correlation values using 3 months' worth of data (August through October 1999).

Station Pair	RMSE	Rank
BLO-BOU	0.4651	0.9151
BLO-NEW	0.5239	0.9128

The actual values were within ±1 standard deviations of the expected values with the exception of the BLO-BOU RMSE value, which was within 2 standard deviations.

5. Conclusions and Recommendations

5.1 Introduction

This project had three main goals: evaluate the performance of the USGS Windows-based program, establish the correlation between the K index time series of different observatories as a function of distance between the observatories, and compute the BLO K index using the USGS program. These goals were achieved. Additional comments and recommendations are presented in the following sections.

5.2 The new USGS program

All the comparison tests involved measuring the performance of the new USGS code to that of the hand-scaled method and old Wilson code. While performing these comparisons, the hand-scaled or the old Wilson code values were considered to be the true values. The results from the comparison tests showed that the new USGS program effectively removed the Sq's seasonal and diurnal trends. However, the tests also revealed that the program's performance decreased as the observatories got closer to the equator when compared to the hand-scaled K index time series or the old Wilson code. This is likely due to the fact that stations closer to the equator deal with additional current systems. While mid-latitude stations respond to auroral electrojets and the mid-latitude Sq currents, low latitude magnetometers additionally respond to the equatorial electrojet and the ring current. The regular signature of these added currents would also have to be removed from the geomagnetic traces along with the mid-latitude Sq signature before

computing the K index. It is therefore possible that the algorithm used by the USGS program to compute the Sq signature curve might be inadequate for low latitude observatories. Two possible solutions are offered to solve this problem. Both solutions assume that the hand-scaled method of computing the K index is the more accurate method (i.e. the hand-scaled distributions are the true distributions). The first solution for this problem would be to use different Sq signature curve computing algorithms based on the latitude of the observatory. Another possible solution lies with the K index conversion tables. Using the large amount of archived data for the hand-scaled K index time series, each station's table could be adjusted so that the K index distributions obtained using the USGS program are similar to the distributions obtained with the hand-scaled method. This will provide consistency between the hand-scaled and new USGS K index values. One must keep in mind that the K index is a mid-latitude index, and not an equatorial index. Perhaps a basic issue is to decide if the K index should even be computed at these latitudes where the problems were identified.

In addition, when the two computer methods are compared to the hand-scaled method, the new USGS program does a better job than the old Wilson code at reproducing the hand-scaled distributions. Furthermore, on average, both computer algorithms calculated higher value K indices than the hand-scaled method. These facts should be of interest to researchers and operations personnel who might consider using the new USGS program. Since the USGS Windows based program is a beta version of the program, it is recommended that similar comparison tests involving the three types of K index time series be repeated when the final version is released.

5.3 Correlation as a function of distance

One of the main accomplishments of this research was to calculate the correlation between the K index time series of existing observatories as a function of distance, and then applying these results to assess the integrity of data from a new site (BLO in this case). To our knowledge, such a technique has never been used before in the field of geomagnetism. The difference plots and scatterplots were effective in qualitatively showing how the K index changes from one observatory to another. These tests showed that the idea of increasing differences (both frequency and magnitude of differences) between time series with increasing distances was sound. These types of tests do not need to be repeated should the K index of a new station need to be evaluated. On a more quantitative front, the RMSE and rank correlation coefficient were useful in establishing a correlation as a function of distance between the observatories. Of the two, the rank correlation coefficients showed less variability and is perhaps a better tool to predict the correlation between two stations based on the distance separating the stations. The rank correlation coefficients computations and least squares fitting should be repeated when the final version of the USGS program is released. If there are future needs to evaluate the K index of a new station, BLO data should be used in the study. This addition could improve the least-squares fits and give better results.

This correlation technique is not limited to validating new observatories. This technique needs to be expanded. Future development of global ionospheric forecast models will depend on a better knowledge of the horizontal "correlation" lengths for ionospheric phenomena. These horizontal correlation lengths are the distances over which a given ionospheric phenomenon can influence the state of the ionosphere in other

locations. These distances are different during storm periods than during quiet periods. Quantitative estimates of these distances can be determined using the USGS algorithm and techniques similar to the ones applied in this research.

5.4 The BLO K index

Even though only three months of data were available for the BLO K index computation, the actual correlation values between the Boulder, Newport, and Bear Lake were within one standard deviation of the predicted values. Based on this information and the magnetometer data plots comparisons, the K index for BLO is considered reliable and accurate. Although no significantly different results are expected, this test should be repeated when the new version of the USGS program is released.

Appendix A: Distance between observatories

The distances between observatories were computed using both geographic and geomagnetic coordinates. There were slight differences in the actual distances of some observatory pairs do to rounding errors. Since these differences were very small, the geographic distances were used throughout the research.

Table 16 – Distance between observatories

Station	Distance	Station	Distance	Station	Distance
BOU-BOU	0	BOU-SIT	2862.4	TUC-BRW	5116.6
BRW-CMO	803.4	BSL-FRN	2886.4	FRD-CMO	5315.8
TUC-FRN	974.1	BSL-NEW	3072.6	BOU-HON	5373.9
DLR-TUC	993.7	TUC-FRD	3092.3	BSL-CMO	5499.7
BOU-TUC	1009.9	NEW-BRW	3266.2	HON-BRW	5559.7
SIT-CMO	1095.9	TUC-SIT	3331.1	SJG-FRN	5587.2
DLR-BSL	1113.9	FRD-NEW	3364.4	FRD-BRW	5640.6
BOU-DLR	1247.5	FRN-CMO	3596.2	SJG-NEW	5640.8
FRN-NEW	1259.7	FRN-FRD	3697.1	DLR-BRW	5718.1
BOU-FRN	1301.1	SJG-DLR	3728.6	HON-DLR	5754.6
BOU-NEW	1305.5	BOU-CMO	3852.1	BSL-BRW	6015
BSL-FRD	1403.1	DLR-SIT	4059.8	HON-GUA	6083.9
NEW-SIT	1561.1	HON-FRN	4076.7	HON-BSL	6843.1
BOU-BSL	1795.3	FRN-BRW	4351.5	SJG-SIT	7048.3
TUC-NEW	1856.6	HON-SIT	4383.5	GUA-BRW	7490.2
SIT-BRW	1869.2	TUC-CMO	4408	GUA-CMO	7580.1
DLR-FRN	1935.9	BOU-SJG	4446.4	SJG-CMO	7772.3
TUC-BSL	2042.2	BOU-BRW	4472.1	HON-FRD	7773.7
DLR-FRD	2371.9	BSL-SIT	4594.8	SJG-BRW	8091.4
BOU-FRD	2401.7	FRD-SIT	4644.3	GUA-SIT	8129.2
SJG-FRD	2459.6	HON-NEW	4701.4	GUA-NEW	9465.4
FRN-SIT	2500.7	SJG-TUC	4712.2	HON-SJG	9465.7
DLR-NEW	2502.7	HON-TUC	4802.5	GUA-FRN	9570.6
NEW-CMO	2587.4	HON-CMO	4898.5	GUA-TUC	10503.1
SJG-BSL	2701.1	DLR-CMO	5082.6	BOU-GUA	10654.8

Appendix B: Distribution Comparison Results

These are the results of the different comparison tests ran on the distributions.

The stations are ordered in descending geomagnetic latitude.

Table 17 – Distribution means for each observatory.

Station	K Index computing method			Absolute Difference in Distribution Means		
	WILSON	USGS	HAND	W-U	W-H	H-U
BRW	3.402	3.387	3.364	0.015	0.038	0.023
CMO	2.600	2.602	2.385	0.001	0.216	0.217
SIT	2.439	2.481	2.050	0.042	0.389	0.431
NEW	2.360	2.347	2.330	0.013	0.031	0.017
BOU	2.498	2.374	2.161	0.124	0.337	0.213
FRD	2.255	2.146	2.194	0.109	0.061	0.049
FRN	2.196	2.334		0.137		
TUC	2.678	2.418		0.260		
FRN	2.196	2.334		0.137		
BSL	2.727	2.457		0.269		
DLR	2.588	2.314		0.275		
SJG	2.145	1.823	1.698	0.322	0.447	0.124
HON	1.837	1.908	1.916	0.071	0.079	0.008
GUA	2.413	2.086	1.984	0.327	0.429	0.102

Table 18 – Distributions of the K index for the different methods used to compute the index. The binned values are percentages of the overall distribution.

Station	Method	K Index Category										TOTAL
		0	1	2	3	4	5	6	7	8	9	
BRW	Wilson	2.51	8.72	18.94	24.33	20.65	14.40	7.69	2.32	0.43	0.03	14246
	USGS	2.61	8.52	18.63	25.73	19.77	14.72	7.51	2.18	0.31	0.02	14296
	Hand	2.53	9.62	18.70	24.79	19.72	14.46	7.63	2.35	0.17	0.02	14804
CMO	Wilson	13.01	19.60	20.43	17.16	12.08	9.43	5.81	2.12	0.32	0.04	14493
	USGS	13.56	18.88	20.31	17.49	11.72	9.67	5.95	2.14	0.25	0.04	14526
	Hand	16.25	20.72	20.61	16.55	11.01	8.36	4.82	1.45	0.20	0.04	100762
SIT	Wilson	10.70	22.15	26.58	18.82	9.42	5.38	3.28	1.97	1.22	0.47	14381
	USGS	9.35	21.51	27.58	20.18	8.99	5.37	3.20	2.12	1.21	0.48	14382
	Hand	17.48	24.78	24.35	16.95	8.01	4.34	2.25	1.38	0.35	0.11	67204
NEW	Wilson	6.65	22.39	30.32	21.78	11.04	5.17	1.52	0.65	0.32	0.15	13618
	USGS	6.84	21.94	30.27	23.38	10.16	4.79	1.53	0.64	0.28	0.18	13663
	Hand	7.81	21.17	28.12	25.01	11.60	4.49	1.25	0.35	0.12	0.08	36021
BOU	Wilson	4.19	19.01	31.39	24.72	12.71	5.57	1.89	0.40	0.11	0.02	14426
	USGS	5.88	20.68	31.24	24.57	10.58	4.75	1.76	0.39	0.12	0.01	14409
	Hand	8.65	25.02	29.85	21.27	9.92	4.05	0.99	0.19	0.03	0.01	23745
FRD	Wilson	5.66	23.37	33.41	21.72	10.77	3.68	1.12	0.21	0.06	0.00	14349
	USGS	7.73	24.86	32.23	21.70	8.90	3.25	1.04	0.22	0.07	0.00	14354
	Hand	9.09	22.95	29.22	23.44	10.47	3.64	0.86	0.22	0.07	0.04	99567
FRN	Wilson	16.99	14.76	27.35	23.03	10.83	5.07	1.51	0.38	0.07	0.01	14527
	USGS	5.77	21.95	31.82	23.76	9.74	4.80	1.70	0.41	0.05	0.01	14552
TUC	Wilson	1.96	13.06	33.65	29.38	13.39	6.21	1.90	0.41	0.04	0.02	14262
	USGS	3.32	20.55	33.83	25.51	9.71	4.88	1.63	0.49	0.06	0.01	14291
FRN	Wilson	16.99	14.76	27.35	23.03	10.83	5.07	1.51	0.38	0.07	0.01	14527
	USGS	5.77	21.95	31.82	23.76	9.74	4.80	1.70	0.41	0.05	0.01	14552
BSL	Wilson	2.24	11.97	32.55	29.60	14.39	6.82	1.96	0.39	0.07	0.01	13466
	USGS	3.82	19.28	32.66	26.17	10.50	5.35	1.72	0.40	0.07	0.01	13598
DLR	Wilson	2.66	14.55	34.26	28.41	12.46	5.69	1.55	0.35	0.05	0.01	12821
	USGS	4.52	22.68	33.69	23.76	9.27	4.12	1.49	0.40	0.05	0.02	12983
SJG	Wilson	4.35	25.45	38.08	20.53	7.69	3.08	0.68	0.14	0.00	0.01	14150
	USGS	11.69	33.42	30.55	14.82	5.89	2.50	0.83	0.23	0.05	0.02	14206
	Hand	21.12	27.68	24.59	17.02	6.73	2.29	0.45	0.08	0.02	0.01	52430
HON	Wilson	21.41	17.69	31.88	18.54	6.79	2.79	0.76	0.14	0.00	0.00	14163
	USGS	10.36	31.80	31.23	15.97	6.28	3.06	0.95	0.31	0.04	0.00	14273
	Hand	11.95	26.41	33.25	19.08	6.16	2.38	0.59	0.15	0.03	0.01	127209
GUA	Wilson	2.13	19.38	37.30	25.11	10.45	4.07	1.14	0.33	0.09	0.00	13797
	USGS	7.48	28.05	32.54	19.90	7.16	2.98	1.19	0.47	0.17	0.05	13902
	Hand	7.32	29.50	32.70	21.36	6.77	1.88	0.36	0.09	0.01	0.01	23575

Table 19 – Distributions of the K index binned according to operational requirements.
Values shown are percentages of the overall distributions.

Station	Method	K index categories		
		0-3	4	5-9
BRW	Wilson	54.49	20.65	24.86
	USGS	55.49	19.77	24.74
	Hand	55.65	19.72	24.63
CMO	Wilson	70.20	12.08	17.72
	USGS	70.24	11.72	18.04
	Hand	74.12	11.01	14.87
SIT	Wilson	78.25	9.42	12.33
	USGS	78.63	8.99	12.38
	Hand	83.56	8.01	8.43
NEW	Wilson	81.14	11.04	7.81
	USGS	82.43	10.16	7.41
	Hand	82.10	11.60	6.30
BOU	Wilson	79.31	12.71	7.99
	USGS	82.38	10.58	7.04
	Hand	84.81	9.92	5.27
FRD	Wilson	84.16	10.77	5.07
	USGS	86.53	8.90	4.58
	Hand	84.69	10.47	4.84
FRN	Wilson	82.12	10.83	7.04
	USGS	83.29	9.74	6.96
TUC	Wilson	78.04	13.39	8.58
	USGS	83.21	9.71	7.07
FRN	Wilson	82.12	10.83	7.04
	USGS	83.29	9.74	6.96
BSL	Wilson	76.36	14.39	9.25
	USGS	81.94	10.50	7.56
DLR	Wilson	79.89	12.46	7.64
	USGS	84.65	9.27	6.08
SJG	Wilson	88.40	7.69	3.91
	USGS	90.48	5.89	3.63
	Hand	90.42	6.73	2.85
HON	Wilson	89.52	6.79	3.69
	USGS	89.36	6.28	4.36
	Hand	90.68	6.16	3.16
GUA	Wilson	83.92	10.45	5.63
	USGS	87.98	7.16	4.86
	Hand	90.88	6.77	2.35

Appendix C: Differences in K Index Class Values

The following table lists the results of the differences between the old Wilson K index time series and the new USGS K index time series. The actual K index values were truncated and then one time series was subtracted from the other. The categories are not cumulative, i.e. the ±2 category only include those instances in which the K indices differed by 2 categories. It does not include the instances in which the K indices differed by 1 category.

Table 20 – Differences between K index time series of two observatories

| Station Pair | Distance (in Km) | Percentage matches | | | | | | |
		±1	±2	±3	±4	±5	±6	Total
BOU-BOU	0	100.00	0.00	0.00	0.00	0.00	0.00	14417
BRW-CMO	803	27.90	51.40	21.20	2.80	0.20	0.00	13738
TUC-FRN	974	72.00	28.50	1.00	0.10	0.00	0.00	14015
DLR-TUC	994	68.00	25.80	1.20	0.10	0.00	0.00	13420
BOU-TUC	1010	65.80	32.80	1.20	0.10	0.00	0.00	14133
SIT-CMO	1096	53.10	46.50	10.60	1.20	0.00	0.00	12835
DLR-BSL	1114	63.70	28.20	1.80	0.10	0.00	0.00	12835
BOU-DLR	1248	56.80	36.20	2.00	0.20	0.00	0.00	13478
FRN-NEW	1260	58.70	34.50	2.40	0.20	0.00	0.00	14207
BOU-FRN	1301	68.10	31.90	1.30	0.10	0.00	0.00	14170
BOU-NEW	1306	66.40	32.30	1.20	0.10	0.00	0.00	13478
BSL-FRD	1403	59.10	37.50	2.00	0.10	0.00	0.00	13527
NEW-SIT	1561	55.20	36.90	5.40	1.20	0.10	0.00	13606
BOU-BSL	1795	57.90	33.70	1.90	0.10	0.00	0.00	14335
TUC-NEW	1857	60.00	42.80	3.90	0.40	0.00	0.00	12458
SIT-BRW	1869	24.90	49.50	28.00	6.50	0.70	0.10	12835
DLR-FRN	1936	58.30	33.50	2.50	0.30	0.00	0.00	13671
TUC-BSL	2042	62.80	30.60	2.50	0.20	0.00	0.00	13846
DLR-FRD	2372	60.20	36.00	2.80	0.40	0.00	0.00	12835
BOU-FRD	2402	63.60	40.60	3.00	0.10	0.00	0.00	13223
SJG-FRD	2460	45.30	49.60	6.90	0.90	0.10	0.00	13596
FRN-SIT	2501	47.20	42.30	8.70	2.60	0.40	0.00	14150
DLR-NEW	2503	46.30	36.70	4.20	0.50	0.10	0.00	13823
NEW-CMO	2587	38.20	41.50	14.20	3.60	0.30	0.00	13890
SJG-BSL	2701	29.10	55.60	10.80	1.00	0.10	0.00	13680

Table 20 - Continued

| Station Pair | Percentage matches | | | | | | | |
	Distance (in Km)	±1	±2	±3	±4	±5	±6	Total
BOU-SIT	2862	54.70	46.60	8.10	2.30	0.30	0.00	12668
BSL-FRN	2886	56.50	38.70	3.90	0.40	0.00	0.00	13596
BSL-NEW	3073	47.30	39.10	4.50	0.50	0.00	0.00	13917
TUC-FRD	3092	55.00	41.30	3.70	0.50	0.00	0.00	13997
NEW-BRW	3266	24.20	41.40	25.50	8.60	1.00	0.00	13283
TUC-SIT	3331	42.30	43.90	9.60	3.10	0.50	0.10	14124
FRD-NEW	3364	53.00	41.30	4.10	0.40	0.00	0.00	13596
FRN-CMO	3596	43.10	49.70	19.40	6.30	0.70	0.00	12134
FRN-FRD	3697	60.10	45.20	5.00	0.50	0.00	0.00	12901
SJG-DLR	3729	37.30	52.70	7.90	1.20	0.20	0.00	12769
BOU-CMO	3852	44.00	51.80	17.60	5.10	0.40	0.00	12049
DLR-SIT	4060	42.50	42.40	10.50	3.50	1.00	0.10	12799
FRN-BRW	4352	26.10	45.40	26.80	10.00	1.60	0.10	12953
TUC-CMO	4408	37.90	47.40	18.70	6.90	0.90	0.00	12717
BOU-SJG	4446	35.50	51.90	13.70	1.80	0.20	0.00	13596
BOU-BRW	4472	25.10	42.40	23.40	8.40	1.10	0.00	14061
BSL-SIT	4595	40.10	41.30	9.50	2.60	0.60	0.10	14215
FRD-SIT	4644	45.30	41.80	9.20	3.30	1.00	0.10	14072
SJG-TUC	4712	35.30	55.50	11.50	1.50	0.20	0.00	13346
DLR-CMO	5083	30.80	37.60	14.40	6.40	1.40	0.10	14235
TUC-BRW	5117	25.90	40.90	22.10	9.20	1.70	0.10	13979
FRD-CMO	5316	35.80	42.10	15.00	7.10	1.50	0.00	14061
BSL-CMO	5500	33.50	44.20	16.00	5.50	0.80	0.00	13516
SJG-FRN	5587	37.30	47.90	11.50	2.00	0.20	0.00	14298
FRD-BRW	5641	22.10	41.40	26.50	12.00	2.90	0.30	13360
SJG-NEW	5641	35.30	50.40	15.40	2.60	0.40	0.10	12731
DLR-BRW	5718	23.10	37.80	21.70	9.80	2.50	0.20	13372
BSL-BRW	6015	27.20	41.70	20.90	8.40	1.60	0.10	13347
SJG-SIT	7048	36.10	47.30	16.70	6.30	2.70	0.90	12731
GUA-BRW	7490	16.30	30.70	24.60	16.30	6.40	1.20	14241
GUA-CMO	7580	24.40	35.80	19.10	11.50	4.00	0.70	14470
SJG-CMO	7772	30.50	37.20	16.60	9.60	4.00	0.70	14328
SJG-BRW	8091	14.10	35.00	30.80	16.20	6.20	1.30	13436
GUA-SIT	8129	30.00	42.90	18.20	6.70	2.30	0.60	13596
GUA-NEW	9465	33.90	44.70	14.30	3.00	0.60	0.20	13444
GUA-FRN	9571	35.80	44.00	14.20	3.70	0.70	0.20	14050
GUA-TUC	10503	37.10	46.30	13.90	3.60	0.70	0.10	13364
BOU-GUA	10655	33.00	44.80	14.60	3.10	0.60	0.20	14280
GUA-DLR	11495	33.50	39.40	10.80	2.70	0.60	0.20	14280
GUA-BSL	12427	31.40	42.10	13.50	2.90	0.50	0.10	14308
GUA-FRD	12770	37.40	46.10	10.60	2.20	0.60	0.20	14077

Appendix D: RMSE and Rank Correlation Coefficient Values

These are the values generated when comparing a pair of stations. Note that for each type of computation (RMSE in Table 21 and rank correlation coefficients in Table 22), the values were computed twice for each type of K index method. This led to twelve least squares fits equations (6 linear and 6 polynomial). An additional six least squares fits equations (3 linear and 3 polynomial) were generated when certain stations were discarded.

Table 21 – RMSE values between K index time series of station pairs

Station Pair	Distance	Hand scaled 7 Years	Old Wilson 5 Years	New USGS 5 Years	Hand Scaled 3 months	Old Wilson 3 months	New USGS 3 months
BOU-BOU	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BRW-CMO	803.4		0.8701	0.8862		1.0896	1.2163
TUC-FRN	974.1		0.8989	0.4074		0.7250	0.7310
DLR-TUC	993.7		0.3952	0.4188		0.6962	0.7228
BOU-TUC	1009.9		0.4347	0.4772		0.4311	0.4752
SIT-CMO	1095.9	0.8469	0.8340	0.8403	0.8468	1.0762	1.2005
DLR-BSL	1113.9		0.4188	0.4661		0.7341	0.7752
BOU-DLR	1247.5		0.5222	0.5604		0.5154	0.5594
FRN-NEW	1259.7		0.9913	0.5517		0.7817	0.8649
BOU-FRN	1301.1		0.9330	0.4644		0.5353	0.4623
BOU-NEW	1305.5	0.6836	0.4489	0.4679	0.6589	0.4530	0.4749
BSL-FRD	1403.1		0.4158	0.4414		0.6344	0.7805
NEW-SIT	1561.1	0.8149	0.7236	0.7298	0.8115	0.9539	1.0924
BOU-BSL	1795.3		0.4958	0.5305		0.4886	0.5275
TUC-NEW	1856.6		0.5932	0.6408		0.7781	0.8880
SIT-BRW	1869.2		1.0764	1.0625		1.2617	1.3630
DLR-FRN	1935.9		0.9488	0.5780		0.8206	0.8316
TUC-BSL	2042.2	0.7513	0.4699	0.5338	0.7441	0.7269	0.8280
DLR-FRD	2371.9		0.5248	0.5649		0.7673	0.8240
BOU-FRD	2401.7		0.5315	0.5306		0.5293	0.5276
SJG-FRD	2459.6	0.7840	0.6843	0.7528	0.8074	0.8064	0.8573
FRN-SIT	2500.7		1.2405	0.9419		1.0939	1.1304
DLR-NEW	2502.7		0.6684	0.7079		0.8768	0.9382
NEW-CMO	2587.4	1.1090	1.0576	1.0720	1.1045	1.2409	1.3605
SJG-BSL	2701.1		0.6607	0.7373		0.7979	0.8397

Table 21 - Continued

Station Pair	Distance	Hand scaled 7 Years	Old Wilson 5 Years	New USGS 5 Years	Hand Scaled 3 months	Old Wilson 3 months	New USGS 3 months
BOU-SIT	2862.4	0.9497	0.8706	0.8737	0.9353	0.8546	0.8594
BSL-FRN	2886.4		0.9891	0.6336		0.8235	0.9080
BSL-NEW	3072.6		0.6620	0.6950		0.8156	0.9549
TUC-FRD	3092.3		0.5579	0.6029		0.7506	0.8497
NEW-BRW	3266.2		1.0650	1.0477		1.2422	1.3374
TUC-SIT	3331.1		0.9836	1.0211		1.0823	1.1676
FRD-NEW	3364.4		0.6447	0.6434	0.7280	0.8606	0.9798
FRN-CMO	3596.2		1.4259	1.1810		1.2934	1.3511
FRN-FRD	3697.1		1.0269	0.6608		0.8317	0.9270
SJG-DLR	3728.6		0.6829	0.7570		0.8010	0.8451
BOU-CMO	3852.1		1.0878	1.1112	1.1898	1.0669	1.0893
DLR-SIT	4059.8		1.0387	1.0698		1.1623	1.2142
HON-FRN	4076.7		0.7810	0.9285		0.7754	0.9279
FRN-BRW	4351.5		1.3605	1.0870		1.2008	1.2686
HON-SIT	4383.5		1.5576	1.3099	1.0778	1.3440	1.2826
TUC-CMO	4408		1.1853	1.2504		1.2556	1.3765
BOU-SJG	4446.4		0.7993	0.8691	0.9237	0.7766	0.8547
BOU-BRW	4472.1		1.0539	1.0476		1.0214	1.0257
BSL-SIT	4594.8		0.9982	1.0160		1.0900	1.1850
FRD-SIT	4644.3	1.0313	0.9920	0.9872	1.0343	1.1191	1.2119
HON-NEW	4701.4	0.9222	1.3197	1.0330	0.8660	1.0423	1.0185
SJG-TUC	4712.2		0.7185	0.7889		0.8272	0.8648
HON-TUC	4802.5		1.1925	0.9320		0.9065	0.9252
HON-CMO	4898.5	1.3309	1.7340	1.5697	1.2868	1.5453	1.5474
DLR-CMO	5082.6		1.2457	1.2993		1.3426	1.4224
TUC-BRW	5116.6		1.0981	1.1398		1.1821	1.2786
FRD-CMO	5315.8	1.3075	1.2073	1.2191	1.3020	1.3273	1.4241
BOU-HON	5373.9	0.9327	1.2822	1.0237	0.8895	0.9988	1.0119
BSL-CMO	5499.7		1.1900	1.2275		1.2559	1.3822
HON-BRW	5559.7		1.5760	1.3681		1.3365	1.3381
SJG-FRN	5587.2		1.0984	0.8583		0.9144	0.9448
FRD-BRW	5640.6		1.1122	1.1011	0.9054	1.2375	1.3173
SJG-NEW	5640.8	0.9178	0.8934	0.9582		0.9940	1.0457
DLR-BRW	5718.1		1.1448	1.1827		1.2471	1.3241
HON-DLR	5754.6		1.1966	0.9710		0.9362	0.9679
BSL-BRW	6015		1.1135	1.1396	0.8685	1.1732	1.3033
HON-GUA	6083.9	0.8877	1.2812	1.1498		1.0444	1.1448
HON-BSL	6843.1		1.2457	1.0174	1.1739	0.9801	1.0133
SJG-SIT	7048.3	1.1884	1.2014	1.2457		1.2611	1.3044

100

Table 21 - Continued

Station Pair	Distance	Hand scaled 7 Years	Old Wilson 5 Years	New USGS 5 Years	Hand Scaled 3 months	Old Wilson 3 months	New USGS 3 months
GUA-BRW	7490.2		1.4865	1.5555	1.2666	1.4928	1.5482
GUA-CMO	7580.1	1.2895	1.5637	1.6334	1.4102	1.5632	1.6267
SJG-CMO	7772.3	1.4242	1.3457	1.4068	0.8641	1.4015	1.4694
HON-FRD	7773.7	0.8869	1.2650	1.0006		0.9996	0.9939
SJG-BRW	8091.4		1.2103	1.2721	1.1135	1.2744	1.3462
GUA-SIT	8129.2	1.1287	1.3551	1.3844	0.9405	1.3568	1.3747
GUA-NEW	9465.4	0.9522	1.0661	1.1163	0.9327	1.0864	1.1142
HON-SJG	9465.7	0.9513	1.2625	1.0683		0.9988	1.0704
GUA-FRN	9570.6		1.2706	1.1328		1.0803	1.1364
GUA-TUC	10503.1		0.9893	1.0979	0.9624	1.0371	1.1044
BOU-GUA	10654.8	0.9584	1.0570	1.1206		1.0349	1.0947

Table 22 – Rank correlation values

Station Pair	Distance	Hand scaled 7 Years	Old Wilson 5 Years	New USGS 5 Years	Hand Scaled 3 months	Old Wilson 3 months	New USGS 3 months
BOU-BOU	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
BRW-CMO	803.4		0.8885	0.8855		0.8769	0.8727
TUC-FRN	974.1		0.7871	0.9399		0.9158	0.9351
DLR-TUC	993.7		0.9428	0.9302		0.9352	0.9236
BOU-TUC	1009.9		0.9443	0.9246		0.9383	0.9194
SIT-CMO	1095.9	0.8839	0.9101	0.9097	0.8762	0.9045	0.9031
DLR-BSL	1113.9		0.9367	0.9186		0.9292	0.9113
BOU-DLR	1247.5		0.9165	0.8989		0.9070	0.8908
FRN-NEW	1259.7		0.7502	0.9055		0.9011	0.8990
BOU-FRN	1301.1		0.7710	0.9277		0.9140	0.9221
BOU-NEW	1305.5	0.8704	0.9470	0.9371	0.8641	0.9411	0.9296
BSL-FRD	1403.1		0.9414	0.9285		0.9346	0.9227
NEW-SIT	1561.1	0.8720	0.9239	0.9190	0.8679	0.9136	0.9083
BOU-BSL	1795.3		0.9241	0.9063		0.9178	0.9003
TUC-NEW	1856.6		0.9054	0.8742		0.8954	0.8644
SIT-BRW	1869.2		0.8329	0.8378		0.8253	0.8281
DLR-FRN	1935.9		0.7573	0.8818		0.8760	0.8706
TUC-BSL	2042.2	0.8319	0.9226	0.8923	0.8277	0.9158	0.8853
DLR-FRD	2371.9		0.9050	0.8856		0.8921	0.8730
BOU-FRD	2401.7		0.9143	0.9090		0.9066	0.9017
SJG-FRD	2459.6	0.8255	0.8399	0.8168	0.8193	0.8332	0.8123
FRN-SIT	2500.7		0.7060	0.8418		0.8412	0.8324
DLR-NEW	2502.7		0.8760	0.8497		0.8630	0.8356

Table 22 - Continued

Station Pair	Distance	Hand scaled 7 Years	Old Wilson 5 Years	New USGS 5 Years	Hand Scaled 3 months	Old Wilson 3 months	New USGS 3 months
SJG-BSL	2701.1		0.8471	0.8228		0.8379	0.8170
BOU-SIT	2862.4	0.8200	0.8867	0.8780	0.8127	0.8760	0.8675
BSL-FRN	2886.4		0.7360	0.8592		0.8603	0.8500
BSL-NEW	3072.6		0.8779	0.8545		0.8693	0.8449
TUC-FRD	3092.3		0.8936	0.8663		0.8818	0.8553
NEW-BRW	3266.2		0.7670	0.7722		0.7524	0.7571
TUC-SIT	3331.1		0.8472	0.8119		0.8360	0.8003
FRD-NEW	3364.4		0.8882	0.8817	0.8500	0.8796	0.8702
FRN-CMO	3596.2		0.6414	0.7721		0.7652	0.7592
FRN-FRD	3697.1		0.7191	0.8485		0.8401	0.8379
SJG-DLR	3728.6		0.8291	0.8033		0.8154	0.7931
BOU-CMO	3852.1		0.8209	0.8148	0.7494	0.8092	0.8018
DLR-SIT	4059.8		0.8226	0.7933		0.8079	0.7794
HON-FRN	4076.7		0.8590	0.7212		0.7612	0.7107
FRN-BRW	4351.5		0.6054	0.7307		0.7235	0.7211
HON-SIT	4383.5		0.5193	0.6471	0.7420	0.6464	0.6397
TUC-CMO	4408		0.7771	0.7450		0.7645	0.7299
BOU-SJG	4446.4		0.8029	0.7771	0.7531	0.7948	0.7708
BOU-BRW	4472.1		0.7570	0.7587		0.7501	0.7445
BSL-SIT	4594.8		0.8376	0.8152		0.8286	0.8031
FRD-SIT	4644.3	0.7909	0.8479	0.8434	0.7795	0.8376	0.8320
HON-NEW	4701.4	0.7670	0.5472	0.6780	0.7754	0.6899	0.6718
SJG-TUC	4712.2		0.8150	0.7874		0.8015	0.7784
HON-TUC	4802.5		0.5961	0.7017		0.7130	0.6919
HON-CMO	4898.5	0.6982	0.4503	0.5514	0.6961	0.5662	0.5391
DLR-CMO	5082.6		0.7487	0.7252		0.7360	0.7097
TUC-BRW	5116.6		0.7232	0.6990		0.7163	0.6893
FRD-CMO	5315.8	0.7147	0.7708	0.7710	0.6969	0.7610	0.7567
BOU-HON	5373.9	0.7374	0.5555	0.6683	0.7445	0.6810	0.6592
BSL-CMO	5499.7		0.7738	0.7573		0.7650	0.7432
HON-BRW	5559.7		0.4558	0.5735		0.5829	0.5704
SJG-FRN	5587.2		0.6784	0.7674		0.7687	0.7579
FRD-BRW	5640.6		0.7195	0.7241	0.7784	0.7134	0.7166
SJG-NEW	5640.8	0.7810	0.7688	0.7384		0.7600	0.7331
DLR-BRW	5718.1		0.6980	0.6790		0.6928	0.6679
HON-DLR	5754.6		0.5871	0.6760		0.6846	0.6627
BSL-BRW	6015		0.7129	0.6972	0.7207	0.7112	0.6898
HON-GUA	6083.9	0.7178	0.5135	0.5472		0.5687	0.5343
HON-BSL	6843.1		0.5549	0.6532	0.7066	0.6615	0.6387
SJG-SIT	7048.3	0.7125	0.7426	0.7171		0.7397	0.7106

Table 22 - Continued

Station Pair	Distance	Hand scaled 7 Years	Old Wilson 5 Years	New USGS 5 Years	Hand Scaled 3 months	Old Wilson 3 months	New USGS 3 months
GUA-BRW	7490.2		0.4461	0.4097	0.7136	0.4320	0.3976
GUA-CMO	7580.1	0.7156	0.5271	0.4950	0.6359	0.5153	0.4794
SJG-CMO	7772.3	0.6502	0.7036	0.6843	0.7628	0.7048	0.6776
HON-FRD	7773.7	0.7560	0.5491	0.6645		0.6589	0.6523
SJG-BRW	8091.4		0.6657	0.6426	0.7219	0.6683	0.6351
GUA-SIT	8129.2	0.7208	0.6046	0.5696	0.7310	0.5951	0.5633
GUA-NEW	9465.4	0.7371	0.6507	0.6148	0.7379	0.6499	0.6132
HON-SJG	9465.7	0.7324	0.5306	0.6088		0.6036	0.5891
GUA-FRN	9570.6		0.5463	0.5703		0.5926	0.5643
GUA-TUC	10503.1		0.6432	0.5775	0.7047	0.6179	0.5649
BOU-GUA	10654.8	0.7119	0.6356	0.5953		0.6185	0.5871

Bibliography

Campbell, Wallace H. <u>Introduction to Geomagnetic Fields</u>. New York: Cambridge University Press, 1997.

Conover, W. J., Practical. <u>Nonparametric Statistics</u>. New York: John Wiley & Sons Inc., 1971

Della-Rose, Devin J. <u>An Investigation of Variable Time Interval "K like" Geomagnetic Indices.</u> Ph.D. dissertation. Utah State University, Logan, UT, 1999.

-----. Faculty, Graduate School of Engineering Management, Air Force Institute of Technology, Wright-Patterson AFB OH. Telephone interview. 15 August 1999.

Hargreaves, J.K. <u>The Solar-Terrestrial Environment</u>. New York: Cambridge University Press, 1992.

Hine, Alfred. <u>Magnetic Compasses and Magnetometers</u>. Toronto, Canada: University of Toronto Press, 1968.

Kelly, Michael C. <u>The Earth's Ionosphere: Plasma Physics and Electrodynamics</u>. New York: Academic Press, 1989.

Mayaud, Pierre Nöel. <u>Derivation, Meaning, and Use of Geomagnetic Indices</u>. Washington DC: American Geophysical Union, 1980.

McPherron, Robert L. "Physical Processes Producing Magnetospheric Substorms and Magnetic Storms," in <u>Geomagnetism</u>. Ed. Jacobs, A. New York: Academic Press, 1991.

Menvielle, M., and Berthelier, A. "The K-Derived Planetary Indices: Description and Availability," <u>Reviews of Geophysics, 29</u>: 415-432 (August 1991).

Radcliffe, J. A. <u>An Introduction to the Ionosphere and Magnetosphere</u>. Great Britain: Cambridge University Press, 1972.

Riddick, J. C. and Suart, W.F. "The Generation of K indices from Digitally Recorded Magnetic Data," <u>Geophysical Surveys, 6</u>:439-456 (1984)

Rishbeth, Henry and Garriott, Owen. <u>Introduction to Ionospheric Physics</u>. New York: Academic Press, 1969.

Shea, M. A., and Smart, D. F., "Space Weather: The Effects on Operations in Space," <u>Advance Space Research,22</u>:29-38 (January, 1998).

Tascione, Thomas F. <u>Introduction to the Space Environment</u>. Malabar, FL: Krieger Publishing Company, 1994.

Wilks, Daniel S. Statistical Methods in the Atmospheric Sciences. New York: Academic
Press, 1995.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503			

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE March 2000	3. REPORT TYPE AND DATES COVERED Master's Thesis	

4. TITLE AND SUBTITLE TESTING THE NEW USGS K INDEX ALGORITHM AT BEAR LAKE OBSERVATORY	5. FUNDING NUMBERS EN - if funded, enter funding number
6. AUTHOR(S) Ariel O. Acebal, Captain, USAF	

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 P Street, Building 640 WPAFB OH 45433-7765	8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/GAP/ENP/00M-01

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFWA/CCD Attn: Major Tom Frooninckx 106 Peacekeeper, Suite 2N3 Offutt AFB, NE 68113-4039 DSN: 271-9769	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES Major Devin J. Della-Rose, ENP, DSN: 785-3636, ext. 4514

12a. DISTRIBUTION / AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.	12b. DISTRIBUTION CODE

ABSTRACT (Maximum 200 Words)

The K index was developed by Bartels in 1939 as an estimate of the level of geomagnetic activity caused by the Sun. This index was computed manually every three hours at geomagnetic observatories using the magnetic traces of the surface planetary magnetic field. In 1991, the International Association of Geomagnetism and Aeronomy approved four additional methods to compute the K index; all of them were computer algorithms. One of the approved methods, the Wilson code, recently underwent some modifications. The new algorithm is now part of a Windows-based computer program being developed by the United States Geological Survey (USGS). After successfully evaluating a beta version of this new program, it was used to compute the K index for a new location. This new location is the Bear Lake Observatory (BLO), where the Utah State University has been collecting geomagnetic data from their magnetometer. Statistical techniques were applied to correlate K indices among existing stations in an effort to develop a test for the validity of the K index of a station. These statistical tests were applied to the BLO K index proving that the technique works and that the BLO K index was computed properly.

14. SUBJECT TERMS K index, Space Weather, Geomagnetic Indices, Magnetometer			15. NUMBER OF PAGES 122
			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL