8-12-2016

Statistical Modeling Of Effective Temperature With Cosmic Ray Flux

Xiaohang Zhang

STATISTICAL MODELING OF EFFECTIVE TEMPERATURE WITH COSMIC RAY FLUX

by

XIAOHANG ZHANG

Under the Direction of Xiaochun He

ABSTRACT

The increasing frequency of sporadic weather patterns in the last decade, especially major winter storms, demands improvements in current weather forecasting techniques. Recently, there are growing interests in stratospheric forecasting because of its potential enhancements of weather forecasts. The dominating factors of northern hemisphere wintertime variation of the general circulation in the stratosphere is a phenomenon called stratospheric sudden warming (SSW) events. It is shown in multiple studies that SSW and cosmic ray

muon flux variations are strongly correlated with the effective atmospheric temperature changes, which suggests that cosmic ray detectors could be potentially used as meteorological applications, especially for monitoring SSW events.

A method for determining the effective temperature with cosmic ray flux measurements is studied in this work by using statistical modeling techniques, such as k-fold cross validation and partial least square regression. This method requires the measurement of the vertical profile of the atmospheric temperature, typically measured by radiosonde, for training the model. In this study, cosmic ray flux measured in Atlanta and Yakutsk are chosen for demonstrating this novel technique.

The results of this study show the possibility of realtime monitoring on effective temperature by simultaneous measurement of cosmic ray muon and neutron flux. This technique can also be used for studying the historical SSW events using the past world wide cosmic ray data.

INDEX WORDS: Cosmic Ray, Effective Temperature, Geant4, Statistical Modeling, SSW, Weather Forecasting

STATISTICAL MODELING OF EFFECTIVE TEMPERATURE WITH COSMIC RAY FLUX

by

XIAOHANG ZHANG

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2016

STATISTICAL MODELING OF EFFECTIVE TEMPERATURE WITH COSMIC RAY FLUX

by

XIAOHANG ZHANG

Committee Chair: Xiaochun He

Committee: Douglas R. Gies

Alexander Kozhanov

Murad Sarsour

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2016

DEDICATION

This dissertation is dedicated to the Nuclear Physics Group at GSU.

ACKNOWLEDGEMENTS

It would not have been possible to finish this dissertation study without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

Above all, I am most grateful to my advisor Dr. Xiaochun He, for his precious guidance and inspiration to me through out the time of my PhD program. Without his ongoing support and patience, I could not finish my research.

Special thanks go to Dr. Murad Sarsour, who has carefully listened to every report I gave and made constructive comments, and who has been so nice and supportive every time I asked questions to him. To Dr. Raphael Tieulent, who has given me valuable suggestions and comments on the hardware test and data analysis. To Carola Butler, who has been always ready to share her knowledge with me, and who has always been a good friend of me. To our former post-doc Dr. Jun Ying and Dr. Liang Xue, both of whom helped me with every single problem I had with simulation program setup, automated DAQ system and online monitoring webpage. I owe my sincere gratitude to Olesya Igorevna Sarajlic, who has spent a lot of time in proof-reading on this dissertation.

I would also like to thank Dr. Douglas R. Gies and Dr. Alexander Kozhanov for being my committee members, for their valuable advice and detailed review, for their time extended to me. I am honored to have you joining the committees of both my MS and PhD defense.

I am grateful to every member in the Nuclear Physics Group at GSU, for all of their help and support. I would like to say thank you to everyone of you, thank you. I am truly proud of being part of you.

I would like to acknowledge the Oulu Cosmic Ray Station, Yu.G. Shafer Institute of Cosmophysical Research and Aeronomy, The Weather Company, LLC, and Department of Atmospheric Science, Wyoming University for their excellent work and wealth of data.

In the final words, I would like to thank my parents, my brother and my uncle. Without their love and support, I would not have come this far.

The final thank you is for Xiaona.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- GSU - Georgia State University
- NSC Natural Science Center
- PMT - Photomultiplier Tube
- DAQ - Data Acquisition
- SSWs - Stratospheric Sudden Warming Events
- OLS - Ordinary Least Squares
- PLS - Partial Least Squares
- PLSR - Partial Least Squares Regression
- PCR - Principle Component Regression
- CV - Cross Validation
- RMSE - Root Mean Square Error
- SD - Standard Deviation

CHAPTER 1

INTRODUCTION

The winter storm in January of 2016 was ranked as one of the greatest blizzards of the past 100 years in terms of amount of snowfall, size of impacted areas and the population affected, according to The Northeast Snowfall Impact Scale, known as "NESIS" [**?**]. It dumped heavy snow from the Mid-Atlantic to southern New England. There were eleven states declared a state of emergency because of the anticipated heavy snowfall and blizzard. Hundreds of millions of people were affected by the storm. Tens of thousands of flights were cancelled. At least 55 people lost their lives in this storm-related incidence [**?**]. Meanwhile, bad weather not only affected North America but also Asia. The city of Guangzhou, located in southern China and known for a humid subtropical climate, received snow on January 25, 2016 for the first time since 1967. January 24, 2016 was the coldest day of Hong Kong in 59 years. Snowstorms across Japan in January of 2016 killed six people [**?**].

Sporadic weather patterns have occurred more frequently in recent years. Shown in Fig. 1.1 are numbers of winter storms that impacted daily life of large populations and economy in northeast U.S. since 1957 with 10-year interval [**?**]. The number of high-impact snowstorms increased sharply in last ten years which seems closely associated with climate change.

While the true reason is still under debate, it is a vital problem for our planet. In order to be better prepared and to minimize the impact of this critical change in climate, more research and development on the current weather forecast techniques are an urgent requirement. In recent years, stratospheric forecast has become one of the most promising research field that could potentially enhance weather forecasting, especially predicting winter storms. As the upper boundary of the troposphere, the stratosphere has possible influence on global modes of the weather system, such as the Arctic Oscillation (AO) [**?**]. AO is a climate

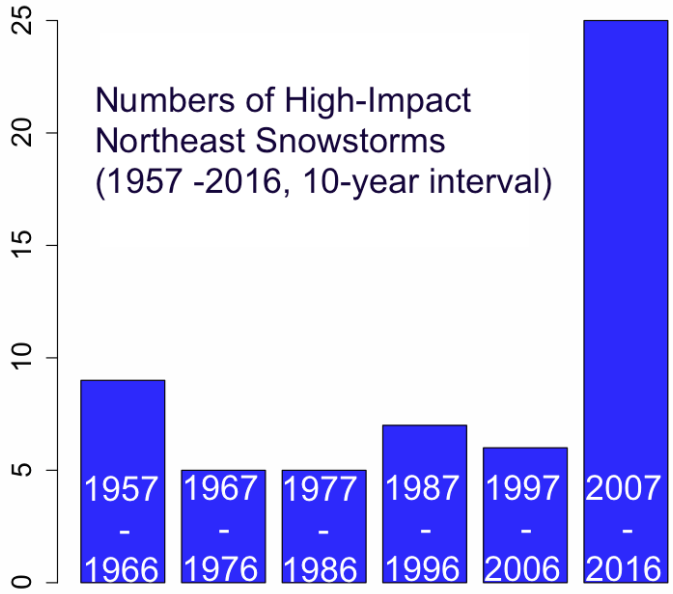

Figure (1.1) Time series of high-impact snowstorms that affected the northeast urban corridor in the past 60 years with 10-year interval. Data is from NESIS [**?**].

pattern characterized by winds circulating counterclockwise around the Arctic at around $55°N$ latitude. AO confines colder air across polar regions in its positive phase and allows increased storminess into the mid-latitudes in its negative phase. A better understanding of the stratosphere can enhance the surface weather forecasting [**?**]. The dominating factors of Northern Hemisphere wintertime variation of the general circulation in the stratosphere is a phenomenon called stratospheric sudden warming events (SSWs) [**?**]. SSWs can dynamically couple the atmosphere all the way from the troposphere to the ionosphere [**?**, **?**, **?**]. It is confirmed that SSWs are also linked with the descent of negative Northern Annular Mode index anomalies which can modify the surface weather [**?**].

The major data sources of SSW observations in past several decades, radiosonde and satellite measurements, have increasingly improved our understanding of the upper atmosphere. However, none of these techniques provides steady and realtime weather data. In 2009, the MINOS collaboration showed that daily variations in secondary cosmic rays measured with its underground detector are associated with the SSWs happened during February

2005 [**?**]. Similar results were shown by the IceCube collaboration with its muon intensity measurement near the south pole [**?**]. These studies demonstrated the possibility of using cosmic ray muon detector for meteorological applications.

It is shown in both of these studies that correlations between SSWs and cosmic ray flux variations are closely associated with the effective atmospheric temperature. In order to monitor SSWs in real time or look for historical SSWs with the recorded cosmic ray measurements, it is necessary to derive the effective temperature from cosmic ray flux variations. The cosmic ray flux is also modulated by other factors such as barometric pressure, solar activities and geomagnetic field. All of these factors need to be taken into account when calculating effective temperature with cosmic ray muon flux variations.

A method for determining the effective temperature with cosmic ray flux measurements is studied in this work by using statistical modeling techniques, such as k-fold cross validation and partial least square regression. This method requires the measurement of the vertical profile of the atmospheric temperature, typically measured by radiosonde, for training the model. In this study, cosmic ray flux measured in Atlanta and Yakutsk are chosen for demonstrating this novel technique.

This dissertation is organized as follows. Chapter 1 presents the introduction, which includes the purpose of this study and an outline of the work. In Chapter 2, topics are discussed such as brief introductions to cosmic ray, meteorological effects on cosmic rays and a review of early studies on potential meteorological applications of cosmic ray detectors. Chapter 3 describes the configurations of four-paddle detector that was developed by the Nuclear Physics Group at Georgia State University (GSU) and the online monitoring system of the four-paddle detector as well as other detectors developed by the group. In Chapter 4, the simulation program as well as method of calculating weighting functions for effective temperature at ground level is described. The extensive discussion about the statistical modeling method, source of data, results of cross validations and discussions is provided in Chapter 5. Finally, Chapter 6 gives the conclusions.

CHAPTER 2

COSMIC RAY AND ATMOSPHERE

2.1 Introduction to Cosmic Ray

Like many other phenomena, cosmic rays were discovered by chance. In 1900, Elster and Geitel noticed an unknown source of ions in the air during their investigation of atmospheric electricity. In the same year, Wilson independently suspected the existence of an ionizing agency that can penetrate a thick layer of earth. The existence of residual ionization after removing all possible causes, scientists were forced to conclude that there should be some as yet unknown radiation that continuously ionizing the air. Moreover, Wilson speculated that the radiation might have a extraterritorial origin. This speculation was proved by Victor Hess' widely accepted discovery with a balloon-borne pressurized ionization chamber. As the balloon went up, Hess observed a slight decrease of ionization and then a rapid increase that persisted up to the highest altitude he reached. This result shows that the source of radiation came from upper level, and therefore the radiation was called cosmic ray. Hess was awarded the Nobel prize in 1936. More detailed studies on the subject were carried out latter. In 1935, Pfotzer showed that cosmic ray intensity reaches a maximum at pressure level about 100 mm Hg, after which it decreases rapidly [**?**]. This maximum was therefore named the "Pfotzer Maximum". Figure 2.1 (a) shows the results of Hess's study [**?**] and (b) shows the results of Pfotzer's measurements [**?**].

Investigations over a century, including balloon flights and satellite experiments, underground detectors and large ground-based arrays, resulted in an overall determination of the composition and energy spectrum of these particles. Cosmic rays can be categorized into two groups according to their source of origin, galactic cosmic rays (GCR) and solar cosmic rays. The former are particles coming from outside the solar system, and the flux varies with solar activities. The majority of the cosmic ray particles come from objects outside

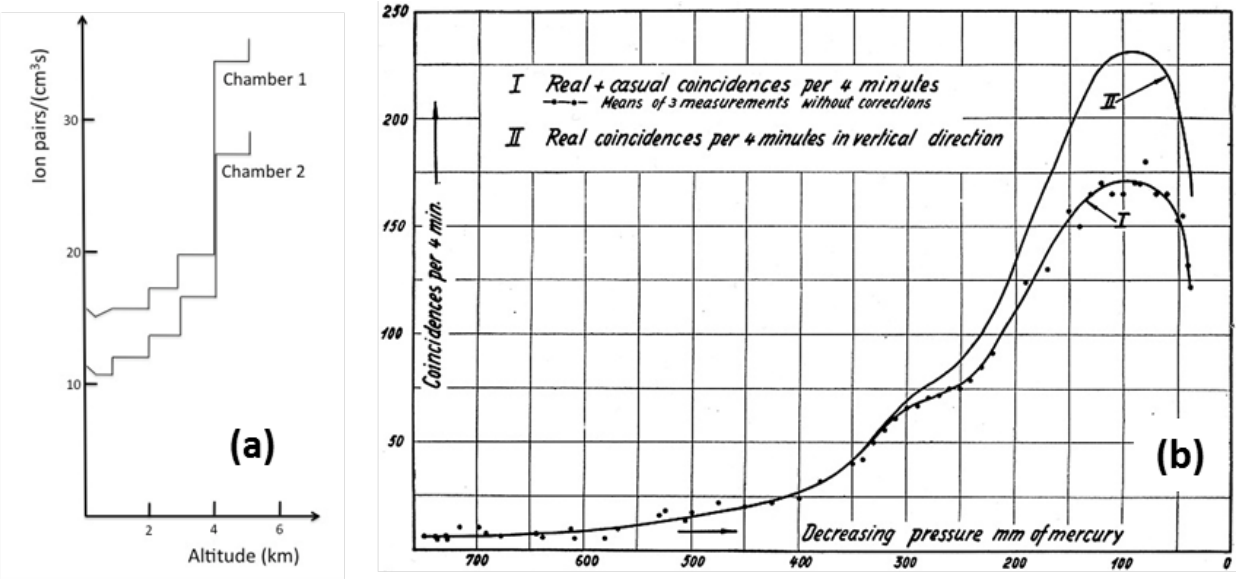

Figure (2.1) (a) Data from ionization rate measurements by Hess (1912), which shows the increase of ionization with altitude. (b) Data from coincidence measurements by Regener and Pfotzer (1935). The coincidence rate per 4 min at a solid angle of 20 degrees about the zenith is shown as function of decreasing pressure (mm Hg) (data points and line I). Data, corrected for dead-time losses, are shown as line II. Note the clear maximum at about 100 mm Hg.

the galaxy [**?**, **?**], such as active galactic nuclei, quasars or gamma ray bursts. Solar comic particles are associated with solar ares and other energetic solar events.

Cosmic rays are also classified in categories of primary cosmic rays and secondary cosmic rays. Technically, "primary" cosmic rays are those particles accelerated at astrophysical sources and "secondary" cosmic rays are those particles produced in interactions of the primary cosmic ray particles with atmospheric molecules. Primary cosmic rays include all stable charged particles and nuclei with lifetimes of order 10^6 years or longer. The intensity of primary cosmic rays in the energy range from several GeV to somewhat beyond 10^5 GeV is given approximately by Eq. 2.1, [**?**].

$$I(E) \approx 1.8 \times 10^4 (E/1\ GeV)^{-2.7} \frac{1}{m^2\ s\ sr\ GeV}, \tag{2.1}$$

where E is the energy-per-nucleon (including rest mass energy) and $\alpha(= \gamma + 1) = 2.7$ is

the differential spectral index of the cosmic-ray flux and γ is the integral spectral index. About 79% of the primary nucleons are free protons and about 15% are nucleons in form of helium nuclei. The fractions of the primary nuclei are nearly constant over the energy ranging from 1 to 10^5 GeV. Figure 2.2 shows the major components for energies greater than 2 GeV/nucleon. Primary cosmic rays are "modulated" by the solar wind, the expanding

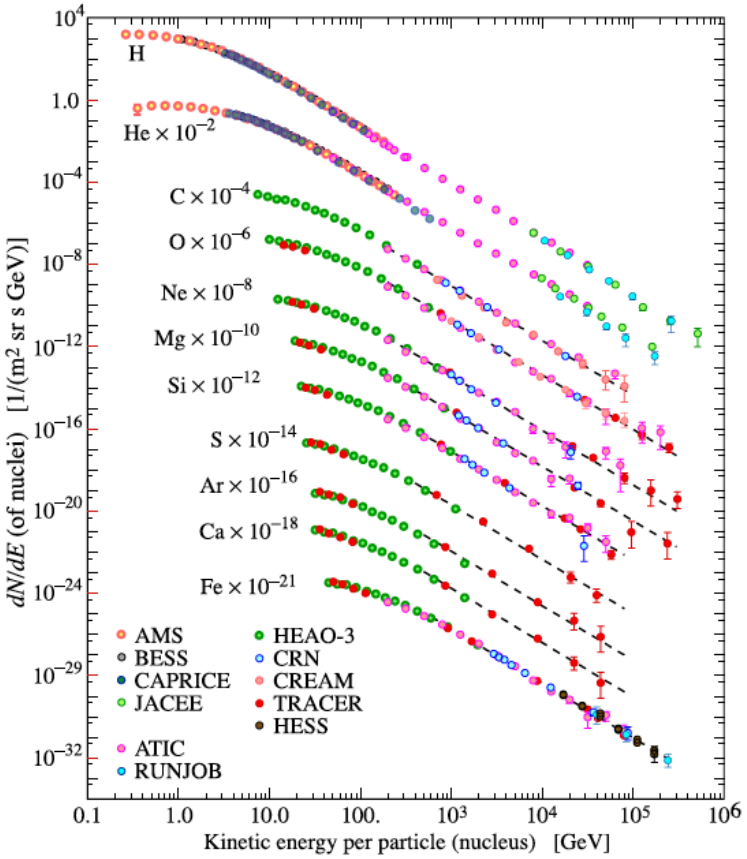

Figure (2.2) Fluxes of nuclei of the primary cosmic radiation in particles per-energy-per-nucleus vs energy-per-nucleus. This figure was created by P. Boyle and D. Muller.

magnetized plasma generated by the Sun, which decelerates and partially dispels the lower energy cosmic rays from the inner solar system. There is a significant anti-correlation between solar activity (which has an alternating eleven-year cycle) and the intensity of the cosmic rays with energies below about 10 GeV. In addition, the lower-energy cosmic rays are affected by the geomagnetic field, which they must penetrate to reach the top of the atmosphere. Thus

the intensity of any component of the cosmic radiation in the GeV range depends both on the location and time [?].

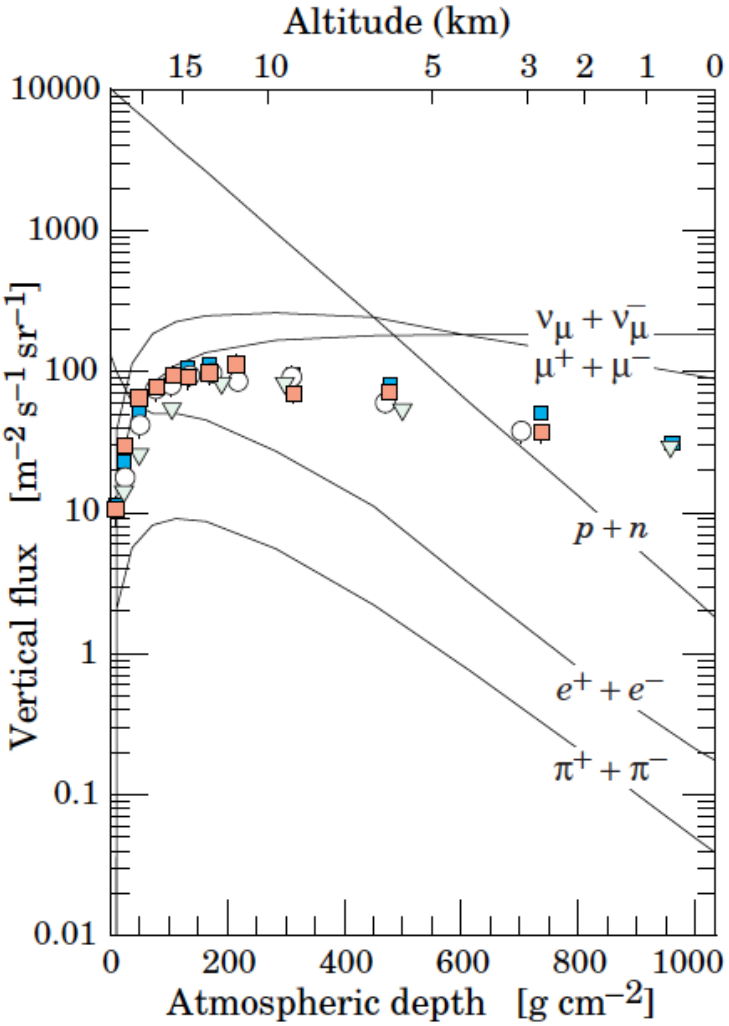

Figure (2.3) Vertical fluxes of cosmic rays in the atmosphere with energy more than 1 GeV estimated from the nucleon flux of Eq. 2.1. The points show measurements of negative muons with energy more than 1 GeV [?, ?, ?, ?, ?, ?].

Figure 2.3 shows the vertical fluxes of the major cosmic ray components in the atmosphere in the energy region where the particles are most numerous. Almost all of the particles in Figure 2.3 are secondary cosmic rays produced in interactions of the primary cosmic rays in the air except for protons and electrons near the top of the atmosphere. Muons and neutrinos are products of the decay chain of charged mesons ($\pi^{\pm}$ and $K^{\pm}$), while electrons

and photons originate in decays of neutral mesons (π^0 and K^0).

Neutrons are produced in the process of cosmic ray spallation when high energy particles collide with atmospheric nuclei. As they propagate through the atmosphere, neutrons collide with atmospheric nuclei, producing additional particles as well as slowing to thermal energies (about 0.025 eV). Thermal neutrons exhibit only random motion due to their energy. These thermal neutrons are typically deposited in the stratosphere for lower energy primaries, with higher energy primaries producing thermal neutrons closer to ground level. Neutron measurements have been taken since 1950s at varies latitudes and altitudes [**?**, **?**, **?**, **?**, **?**]. Those measurements are useful for determination of cosmic ray flux variability.

Muons are the most numerous charged particles at sea level. They are produced from decay of charged mesons, mostly in the region of atmospheric pressure layer 100 hPa to 200 hPa. After traveling through the atmosphere, every muon particle on average loses about 2 GeV before arriving at ground due to ionization. The atmospheric muon intensity variation is the convolution of three major classes, geomagnetic origin, extraterrestrial origin and atmospheric origin (meteorological effects). Variations of atmospheric origins have been extensively studied and quite a few researchers have tried to solve the inverse problem, which is determining atmospheric conditions with cosmic ray measurements [**?**, **?**, **?**]. Meanwhile, theories and experiments have also been being developed and designed to understand the variations of the geomagnetic origin and especially the extraterrestrial origin for more than half century.

2.2 Meteorological Effects on Cosmic Ray

Meteorological effects are interesting topics to study from two aspects. First, careful study of these effects helps in developing a trustworthy method for computing the meteorological corrections which, when applied to the observational data, make it possible to find the cosmic ray variations of extra-atmospheric origin. Second, they yield reliable information about variations in the upper atmosphere of the Earth and about the character of the nuclear-meson cascade of cosmic rays in air. Both aspects are very useful in this work.

The two major meteorological effects on cosmic ray flux measurements are the barometric effect and the temperature effect. The theory of these effects is based on present day notions about the elementary processes and about the nuclear-meson cascade in the atmosphere. For cosmic ray neutron flux, the dominating atmospheric originated variations are caused by barometric effect and the temperature effect is negligible. Meanwhile, most of the cosmic ray muon intensity variations with atmospheric origins are induced by the air temperature fluctuations, and barometric pressure fluctuations have noticeable effect on muon flux variations.

There are other meteorological effects on cosmic ray flux such as humidity effect, wind effect (bernoulli effect) and atmospheric electric field effect. Those effects are all relatively negligible comparing to the barometric effect and temperature effect in most of the cases, therefore are not discussed in this work.

2.2.1 Barometric Effect

The barometric effect on the cosmic ray flux has been extensively studied for many decades. Researchers applied pressure corrections to their measurements to study extraterrestrial factors that influence cosmic ray flux for decades [**?**]. The relationship between the percent variations of secondary cosmic ray flux (neutron or muon) and barometric pressure can be described by the following equation,

$$I = I_0 e^{-\beta \delta P}, \tag{2.2}$$

where I and I_0 are secondary cosmic ray flux from measurement and the reference point of secondary cosmic ray flux (usually the mean of the measurement period). δP is the local barometric pressure, and β is called the barometric coefficient. One can easily write Eq. 2.2 in the following form,

$$\delta I = e^{-\beta \delta P} - 1, \tag{2.3}$$

where $\delta I = \frac{I-I_0}{I_0}$ is the percent variation of muon flux. Usually the barometric coefficient, β, is obtained with linear regression of the following equation,

$$ln(\delta I) = -\beta \delta P. \tag{2.4}$$

Once β is determined from the linear regression, the pressure corrected cosmic ray flux can be calculated as follows

$$I_c = Ie^{\beta \delta P}, \tag{2.5}$$

where I_c is the pressure corrected cosmic ray flux. When applying a pressure correction to the cosmic ray flux, the period of the data for calculating β need to be carefully selected, since the reliability of β strongly depends on the correct choice of data and method of analysis. There are some general rules to follow as pointed out by Dorman [**?**]

- It is better to use 3 to 7 days of data;
- Choose periods of low solar activity, during which cosmic ray flux does not have any anomalous variations.
- Temperature correction should be applied to the data of ionizing components at first, if it is applicable.

In practice, the first order and second order approximations of Eq. 2.3 are often used as well,

$$\delta I \approx -\beta \delta P, \tag{2.6}$$

$$\delta I \approx -\beta \delta P - \frac{\beta^2}{2}(\delta P)^2. \tag{2.7}$$

2.2.2 Temperature Effect

Cosmic ray muon particles are products of the decay chain of charged mesons ($\pi^{\pm}$ and $\kappa^{\pm}$), mostly in the 100 to 200 hPa pressure layers. Muon particles can be detected both at ground level and in deep underground. There have been many studies of temperature

effect on both the neutron and muon flux [**?**] in the past several decades. It was found that the temperature effect is the most pronounced meteorological effect on muon flux variations. Muon flux measured at sea level is mainly contributed by soft muons (below 10 GeV). Muons that survived to the deep underground are hard muons, energy of which are above hundreds of GeV. The hard muon intensity vary with the decay/interaction ratio of the parent mesons ($\pi^{\pm}$ and $\kappa^{\pm}$). When temperature at the production level of muons is higher, the air become less dense, and more mesons tend to decay to muons. The soft muon intensity is mainly modulated by the particles' production levels which rises with higher atmospheric temperature and falls with lower temperature. Because of their short lifetime and relatively low energy, soft muon particles produced at a higher production level caused by the increased temperature, more particles decay to electron/positron and neutrino before they can survive to the ground. Therefore there are negative temperature effect on the low energy muons and positive temperature effect on the high energy muons.

Since the production of secondary muon particles can happen at any level of the atmosphere (mainly in the region of tropopause and lower stratosphere), it is necessary to take the whole atmosphere into account in order to study the temperature effect on cosmic ray muon flux variations. The effective temperature was introduced to simplify this problem.

A few early studies [**?**, **?**] show that the correlations between cosmic ray muon intensity and atmospheric temperature can be described by the effective temperature, T_{eff}, as follows:

$$\frac{\Delta I}{I_0} = \alpha_T \frac{\Delta T_{eff}}{T^0_{eff}}, \tag{2.8}$$

where $\frac{\Delta I}{I_0}$ is the variation of cosmic ray muons, $\frac{\Delta T_{eff}}{T^0_{eff}}$ is the variation of effective temperature, and α_T is called temperature coefficient.

Effective temperature is basically the weighted average of atmospheric temperature from the observation region to the generation level of muons. It is usually calculated with atmospheric temperature vertical profiles and weighting functions. The vertical profiles of the atmospheric temperature are usually measured with balloon-borne radiosondes which have

been used for upper atmosphere measurement since the late 19th century. The balloons are launched multiple times (two to six) everyday and can reach up to 10 hPa pressure layer (about 30 km above sea level) for most of the measurements. There are over a thousand weather stations all over the world. The period of record varies from station to station, with many extending from 1970 to the present. Daily updates of station records are available online at no charge. There are several weighting functions derived theoretically by several groups, such as the one shown by the MACRO collaboration [**?**]. Since all of those weights are for muon flux measurements at far deep underground. In this work, Monte Carlo simulations were used to obtain the weighting functions for muon detectors installed near the surface of the Earth.

Following the work of MACRO collaboration, assuming that the temperature dependence on X (interaction depth) has only negligible effect on the factorization of production spectrum of muons, P_μ,

$$P_\mu(X, E_\mu) = h(X)\Pi(E_\mu), \tag{2.9}$$

where $h(X)$ is the distribution of muon flux along interaction depth, and $\Pi(E_\mu)$ is the muon energy spectrum. Integrate over energy E_μ, one gets

$$\int_0^\infty dE_\mu P_\mu(X, E_\mu) = h(X) \cdot \int_0^\infty dE_\mu \Pi(E_\mu). \tag{2.10}$$

From MACRO's calculation, it is also known that,

$$T_{eff} = \frac{\int_0^\infty dX h(X) T(X)}{\int_0^\infty dX h(X)}. \tag{2.11}$$

Then using eq. 2.10 and eq. 2.11, one gets

$$T_{eff} = \frac{\int_0^\infty dE_\mu \int_0^\infty dX T(X) \cdot P_\mu(X, E_\mu)}{\int_0^\infty dE_\mu \int_0^\infty dX P_\mu(X, E_\mu)}, \tag{2.12}$$

$$T_{eff} = \frac{\int_0^\infty dX T(X) \int_0^\infty dE_\mu P_\mu(X, E_\mu)}{I_\mu}. \tag{2.13}$$

where I_μ is the total number of muon particles arriving at ground, $T(X)$ is the temperature at interaction depth X and $\int_0^\infty dE_\mu P_\mu(X, E_\mu)$ is the total number of muon particles which are produced at interaction depth X and arrive at the ground. The integration can be approximated with a sum over a series of interaction depth ranges,

$$T_{eff} \approx \sum_{i=1}^{k} T_i \cdot \frac{\int_{X_i}^{X_{i+1}} dE_\mu P_\mu(X, E_\mu)}{I_\mu}. \tag{2.14}$$

Because vertical profiles of atmospheric temperature are usually measured along barometric pressure and altitude by radiosondes, it will simplify the calculation if pressure is used instead of interaction depth. Since within 0 to 100 km above the ground, the acceleration of gravity is essentially constant, using $X = P/g$,

$$T_{eff} \approx \sum_{i=1}^{k} T_i \cdot \frac{\int_{p_i}^{p_{i+1}} dE_\mu P_\mu(P/g, E_\mu)}{I_\mu}, \tag{2.15}$$

where T_i is the temperature in the i^th pressure layer, and $\frac{\int_{p_i}^{p_{i+1}} dE_\mu P_\mu(P/g,E_\mu)}{I_\mu}$ is the weight for the i^th pressure layer (labeled as W_i). In W_i, $\int_{p_i}^{p_{i+1}} dE_\mu P_\mu(p/g, E_\mu)$ is the total number of muon particles which are produced in the i^{th} pressure layer and arrive at the observation level, which is very difficult to obtain experimentally. In this work W_i were estimated with Geant4 simulations. With the weighting function (obtained with simulation) and time series of atmospheric temperature vertical profiles T_i (obtained from radiosonde measurements), T_{eff} can be calculated with Eq. 2.15.

2.3 Measuring the Atmospheric Profile with Cosmic Ray Flux

The sensitivity of the cosmic ray muon flux to the temperature change in the atmosphere, especially in the tropopause region (100 - 200 hPa), suggests the possibility of using muon detectors for meteorological applications. As a matter of fact, many methods have been designed and tested in this field since 1970s.

Dorman pointed out in his book [**?**] that using the spectrographic methods one can

get very high accuracy in monitoring atmospheric temperature. Several groups applied the spectrographic methods to obtain information about the upper air with multiple types of cosmic ray detectors[**?**, **?**]. These methods usually need simultaneous measurements of the secondary cosmic ray muon and neutron flux at different altitudes or measuring cosmic ray muons from different zenith angles. Because of the complexity of the setup of the equipments, the potential application of these methods are limited.

In 2005, Borog et al. showed their work of measuring the vertical profile of atmospheric temperature [**?**]. This technique of monitoring the atmospheric temperature at different altitude was suggested by Dorman back in the 1970s [**?**]. Borog applied this technique with a muon detector that measures cosmic ray muon flux from 10 zenith angles. By solving the system of ten linear equations of muon flux variations and temperature fluctuations in each pressure layer, Borog was able to obtain the temperature variations in ten pressure layer ranging from 900 hPa to 100 hPa. The results showed very good consistency with the radiosonde measurements.

In recent years, the possibility of using secondary cosmic ray muon measurements to monitor sudden stratospheric warming (SSW) events in real time was discussed. In 2009, the MINOS collaboration showed that daily variations in secondary cosmic rays from their underground detector are associated with the SSW that happened during February 2005 [**?**]. Similar results were shown by the IceCube collaboration with their muon intensity measurement in the polar region [**?**]. It was shown in both studies that there are close connections between SSW events and effective temperature. The MINOS collaboration suggested that effective temperature can be very useful in the studies of both reconstructing the SSW events with early cosmic ray data and realtime monitoring of the SSWs with current muon measurements.

It is possible to obtain the effective temperature with cosmic ray muon flux measurements. In order to derive the temperature variations with muon flux, it is necessary to unfold the muon flux variations caused by the temperature effect and exclude those caused by other influential factors such as geomagnetic field, extraterrestrial factors and meteorolog-

ical effects except for the temperature effect. Variations of geomagnetic and extraterrestrial origins on muon flux are directly attributable to the fluctuations in the primary cosmic ray fluxes. Conventionally, the secondary neutron flux after pressure correction were used as the first order estimation of primary cosmic ray fluctuations. One of the most practical and sufficient methods to estimate primary cosmic ray intensity fluctuations is using simultaneous measurement of neutron flux at the same location as where muon flux is measured. Other atmospheric effects can be ignored since they are much less significant than barometric effects.

In general, the percent variations of effective temperature can be described with the following equation,

$$\delta T_{eff} = F(\delta I_{\mu}, \delta I_N, \delta P). \tag{2.16}$$

where T_{eff} is the effective temperature percent variations, I_{μ} is the muon intensity percent variations, I_N is the neutron intensity percent variations, and P is the barometric pressure. For those datasets that have muon and neutron flux corrected with pressure, the term P can be ignored as in the case of Yakutsk' datasets. In this work, a series of generalized linear regression models have been evaluated for determining the best approximation of function F.

CHAPTER 3

COSMIC RAY MEASUREMENT

Cosmic ray detectors developed by the Nuclear Group at GSU include the "Mu II" detector, "Pot" detector, double paddle detector and four-paddle detector. These detectors have been operated for years to record the muon flux at the GSU campus located in downtown Atlanta. In this work, for the purpose of studying the muon detector's meteorological application, the four-paddle detector was used. The four-paddle detector was assembled in late 2011. After careful adjustments of the distances between paddles and voltage of each photomultiplier tube (PMT), the four-paddle detector started continuos operation with the updated data acquisition system (DAQ) in the Natural Science Center (NSC), GSU. Measurements by the four-paddle detector covers the whole period from March to September 2014. In October 2014, the detector was moved to a new location, the 26th floor of 25 Park Place on campus of GSU, and has been working since then. A series of scripts and codes were developed to aggregate automatically the hourly and daily counts of the muon flux measured by the four-paddle detector, apply pressure correction, calculate central moving averages of the data, make up-to-date plots of measurements and present plots on the website. In this chapter, a detailed description of four-paddle detector is presented in the first section. The automated data recording procedures, data analysis and data visualization system are given in the second section.

3.1 Cosmic Ray Detector

As shown in Figure 3.1, the four-paddle detector consists of four "paddles", NIM & CAMAC based DAQ system and a computer with a Linux system and PCI GPIB interface card. Separations between paddles are different in order to obtain different acceptance angles ranges. In 2014, the relative heights from bottom paddle to top paddle are 57.0 cm, 99.0 cm

and 145.0 cm. Since October 2014, the relative heights were changed to 32.0 cm, 55.0 cm and 70.5 cm to get larger acceptance angles.

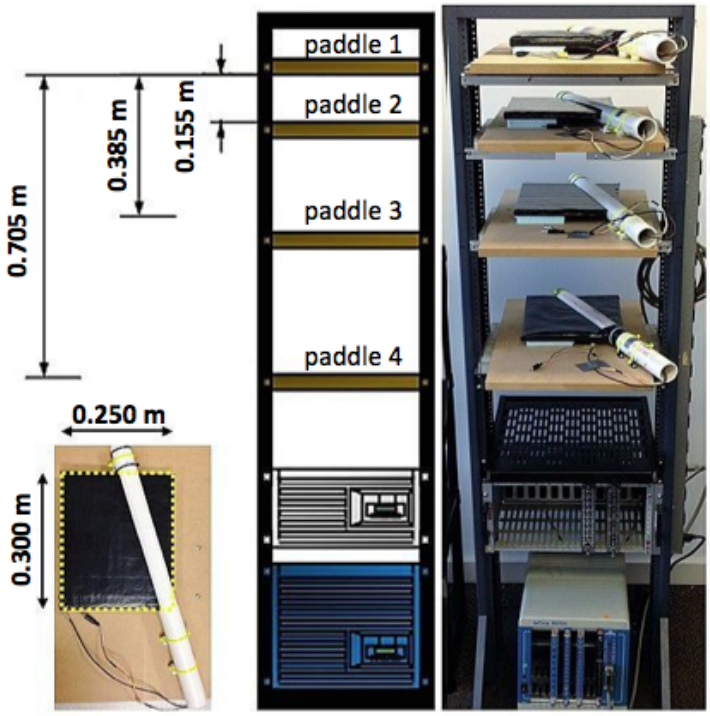

Figure (3.1) Four-paddle detector built and operated by the Nuclear Physics Group at GSU. The detector has four rectangular plastic scintillator plates, four PMTs and NIM & CAMAC based DAQ system. The separations shown in the figure have been used since October 17th 2014 when the detector was moved to 26th floor of 25 Park Place.

Each paddle consists of four components: plastic scintillator plate, light guide "cookie", PMT and a piece of PVC pipe. The "cookie" is basically a connector between the scintillator and the lens of the PMT. On one side of the "cookie" there is a notch that matches with the corner of the scintillator plate and the other side is the same size as the lens of PMT. The scintillator, the "cookie" and the joint between "cookie" and PMT are wrapped with two layers. The inner layer is aluminum foil and the outer layer is electrical tape. The PVC pipe that has its center partially removed works as a holder of the whole combination.

The scintillator is a major component of the four-paddle detector. As a cosmic particle traverses the plastic scintillator, molecules in the vicinity of the trajectory are excited. These

excited molecules will emit visible light through de-excitation process. The emitted visible light is then collected by the PMT through "cookie" and converted to the electric signals. The anode signals from paddles are transmitted to the DAQ system through LEMO cables.

The DAQ system of the four-paddle detector is NIM and CAMAC based. In the NIM crate, there are a Phillips Octal discriminator (model 755) and a Phillips Quad Majority Logic Unit (model 755) installed. Scalers (Lecroy Model 2551 12-Channel 100 MHz Scaler) and CAMAC controller (LeCroy 8901A GPIB To CAMAC Interface Module) on CAMAC are also part of the DAQ.

Shown in Figure 3.2 are analog signals from paddles which are sent to the discriminator with equal length LEMO cables. Four isolated units of the discriminator convert the input signals to digital signals. The digital signals are then fed into different units of the logic module to acquire coincidences of different channels with various folds. As shown in the right bottom of Figure 3.2, when two paddles output signals within a small time window (several nano seconds), the logic module will output one digital signal, which is called a 2-fold coincidence. Then the coincidence signals are fed into the scaler, which count the numbers of input signal in each channel. The DAQ computer reads counts from the scaler through the GPIB controller periodically with a LabView program, which will be discussed latter.

There are several reasons for using the coincidence method. First, muon particles are identified by the coincidence method. The scintillator of the four-paddle detector is not only sensitive to muon particles but also other charged particles. However, out of the possible particles that could generate an output signal in the paddle, the muon particle is the only kind that can penetrate several layers of scintillation plates as well as the air between them. Therefore a coincidence signal could be safely recorded as one count of a muon particle. Second, there are very small chances that the two signals coming in within the time window of the logic module are caused by two different particles. When a cosmic ray shower happens, secondary particles in the shower front usually arrive at ground with very small time differences. It is possible that two particles in the shower front hit two paddles almost at the same

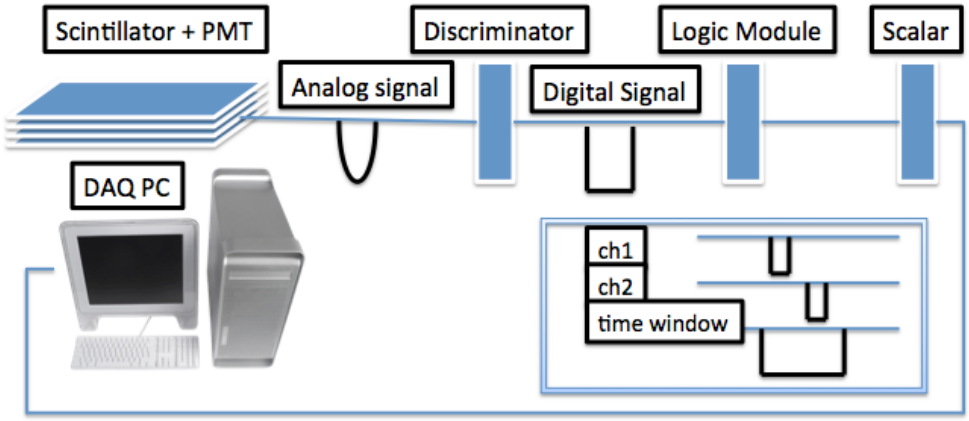

Figure (3.2) A schematic diagram of the signal flow. The signals from each paddle are processed by discriminators, logic module and scaler and then read by computer. Right bottom inset shows an example of a 2-fold coincidence.

time. This type of misidentifications could be avoid by acquiring multi-fold coincidences like 3-fold coincidence and 4-fold coincidence, which ensured the identified particles are those that penetrated all three or four paddles rather than different particles coming from other directions. Third, acquiring the coincidence of multiple paddles confines the incoming directions of the detected muons. The smaller the separation between the two paddles, the larger the acceptance angle. The coincidence method only accepts muon particles that penetrate

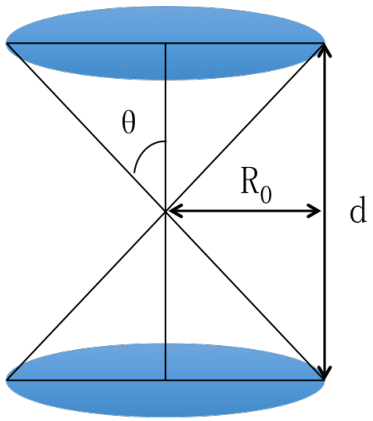

Figure (3.3) Effective area and acceptance angle. R_0 is the radius of the area of the equivalent circle, d is the distance between the two paddles and θ is the approximated acceptance angle.

all paddles considered in the coincidence, which means that only those muons that come from a specific zenith angle range can be detected. These angles are so called acceptance angles of the detector. Since the scintillation plates of the four-paddle detector is square shape, the acceptance angles are different for particles that coming from different side of the detector. As shown in Figure 3.3, the acceptance angles of different coincidences are approximated with a circle that has equivalent area as the scintillation plate. The area of the plates are 30.0 cm × 25.0 cm = 750.0 cm^2, so the radius of a circle that has the same area as the plate is $R_0 \approx 15.5$ cm. Please note that the width and length used here are the measurements of the scintillator itself without the wrap. The approximate acceptance angle for a coincidence between the two scintillators is $\theta = \arctan(\frac{R_0}{d/2})$. The acceptance angle of each coincidence during the data-taking operations in 2014 and 2015 are listed in Table 3.1.

Table (3.1) Approximate angular acceptance of the four-paddle detector at two locations (NSC building and 25 Park Place building).

Location	Coincidence	d (cm)	θ (°)
NSC	2-fold, (paddle 1 and 2)	57.0	28
NSC	3-fold, (paddle 2, 3 and 4)	99.0	17
NSC	4-fold, (all 4 paddles)	145.0	12
25 PP	2-fold, (paddle 1 and 2)	15.5	60
25 PP	3-fold, (paddle 1, 2 and 3)	38.5	40
25 PP	4-fold, (all 4 paddles)	70.5	23

As discussed above, the counts of the muon particles are stored in the scaler. Shown in Figure 3.4, a LabView program was developed to read out the scaler counts and to save data into files for further analysis. The program can communicate with every single channel of the scaler by specifying the slot number (2) of the scaler and channel number (6 - 10) on its control panel. As shown in the sample below, counts of five channel in the scaler are recorded by the program, which are coincidences of paddle 1 & paddle 2 (column 1), paddle1 & paddle 2 & paddle 3 (column 2), paddle1 & paddle 2 & paddle 3 & paddle 4 (column

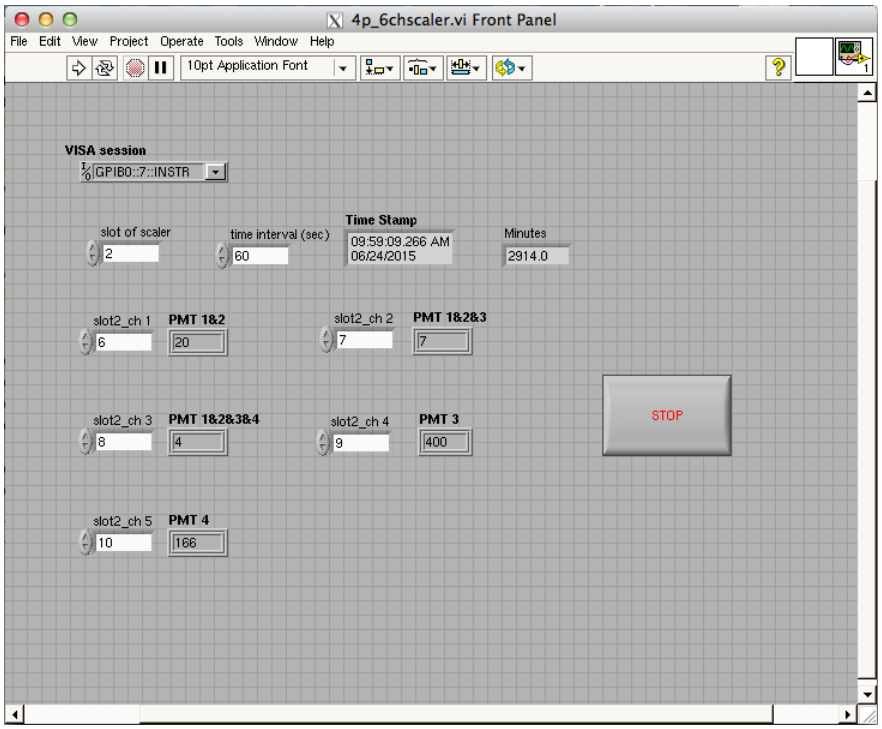

Figure (3.4) Control panel of the LabView program developed for the four-paddle detector.

3), counts of paddle 3 (column 4) and paddle 4 (column 5). The program talks to scaler every 60 seconds, as specified on the control panel (time interval), pulls all five counts back to computer and write to the data file with the timestamp of the current time (column 6), and reset the counts on the scaler. The program controls the accumulating time of scaler, so each group of counts saved to the data file is a sampling of the muon flux during a 60 second period. The data file on the computer is then copied to the local cluster of the Nuclear Physics Group every 24 hours with a bash script for further processing.

410	*158*	*55*	*6879*	*4944*	*03092016200006*
400	*143*	*57*	*6849*	*4799*	*03092016200106*
402	*172*	*65*	*6931*	*4888*	*03092016200206*
385	*153*	*69*	*6886*	*4816*	*03092016200306*
400	*151*	*51*	*6925*	*4928*	*03092016200406*
392	*126*	*59*	*6776*	*4868*	*03092016200506*

3.2 Online Monitoring System

Figure 3.5 shows the layout of the online data monitoring system. The data are transferred from the DAQ computer to a local cluster and stored into a database. On the local cluster the pre-processed data are plotted and uploaded to the monitoring website (`http://phynp6.phy-astr.gsu.edu/~cosmic/`) in every 24 hours. The raw data are aggregated into hourly averaged counts and daily averaged counts with a C++ program, and will then be uploaded to a local database. The local weather data are also updated to the database daily. The plots online include the pressure corrected and uncorrected muon flux (daily and hourly) together with the central moving average. The method of calculating the moving average is discussed in chapter 5.

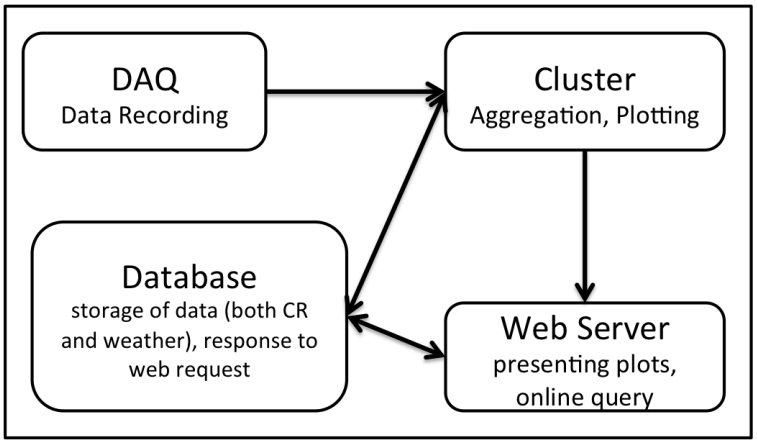

Figure (3.5) Flow chart of the online monitoring system.

The online monitoring website (see Figure 3.6) runs on a local web server. The latest hourly and daily muon flux measured by the four-paddle detector and other detectors are updated to the website every 24 hours. There are also many of other measurements available on the website including the daily updated background radiation measurements (via Geiger counters), local weather data (barometric pressure, humidity and ground temperature) and solar activities. The muon flux data can be downloaded from the website on the online query page as shown in Figure 3.7. One can also request for data plots online.

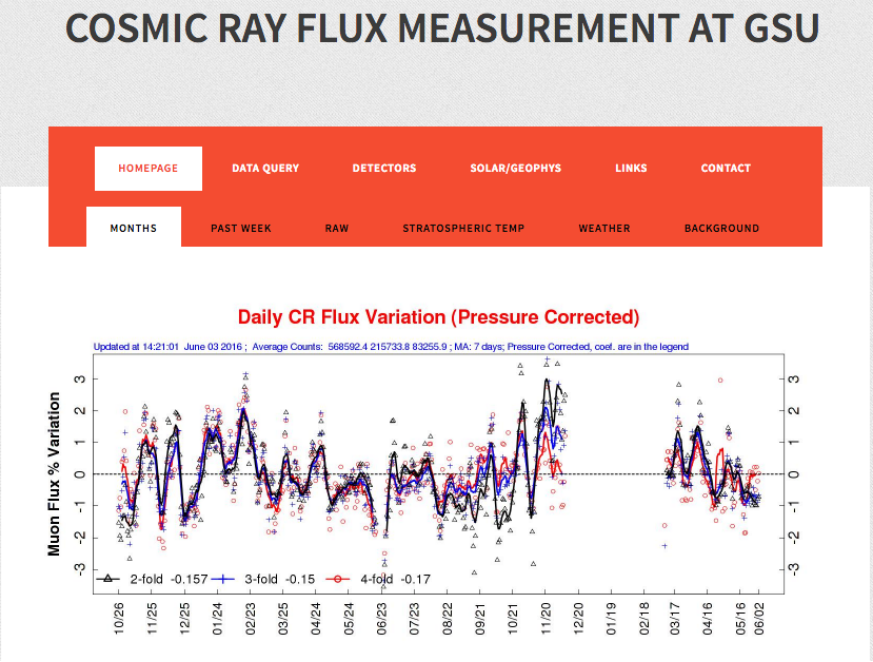

Figure (3.6) A screen capture of the online monitoring website.

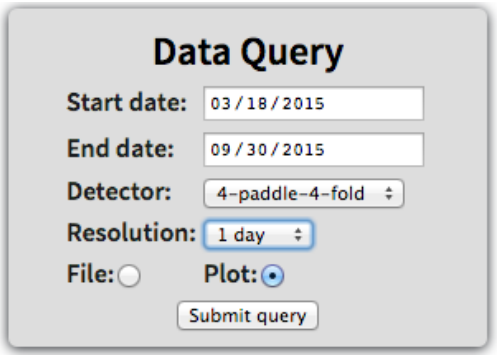

Figure (3.7) The online query page.

CHAPTER 4

ANALYSIS METHODS

4.1 Introduction

This chapter describes the details of the data analysis procedure for determining the effective temperature using the cosmic ray flux measurement. As discussed in Chapter 2, it is necessary to have both the atmospheric temperature vertical profiles (radiosonde measurement) and proper weighting functions for a given location. In this work, the weighting functions are determined with Geant4 simulations. In order to evaluate the predictive power of every candidate model, statistical modeling techniques such as partial least square regression (PRSL) and k-fold cross validation (CV) are used.

4.2 Weighting Function Determination

Because of the geomagnetic effect, the weighting function for characterizing the cosmic ray muon production in each atmospheric layer is geoposition dependent. In this work, the weighting function is calculated with Geant4 simulation. Geant4 has been used in nuclear and particle physics research for decades to simulate radiation and its effect on matter. It has wide applications in space radiation and medical physics.

4.2.1 Geant4 Simulation

In the Geant4 simulation, the atmosphere is constructed as 100-km thick air shell (composed of 70% Nitrogen and 30% Oxygen) which is subdivided into 1000 sub-shells. Every sub-shell is 0.1 km in thicknesses (outer radius minus inner radius). The air density of each sub-shell is defined as the density at its altitude in the atmospheric profiles provided by the MSIS-E-90 Atmosphere Model (`http://omniweb.gsfc.nasa.gov/vitmo/msis_vitmo.html`).

In MSIS-E-90 Atmosphere Model, one can get air density vs altitude at different locations (specifying longitude and latitude) and at different times (specifying time in a day, date, month and year). Air density data from 0 km to 100 km on the 15th day (12:00pm local) of every month is used to calculate the average air density over a year at each location. In this work, the weighting functions both for Atlanta, USA (year 2014) and Yakutsk, Russian (year 1990) are calculated with this simulation procedure. The air densities at each altitude of Atlanta and Yakutsk are shown in Figure 4.1. The atmosphere above 100 km is

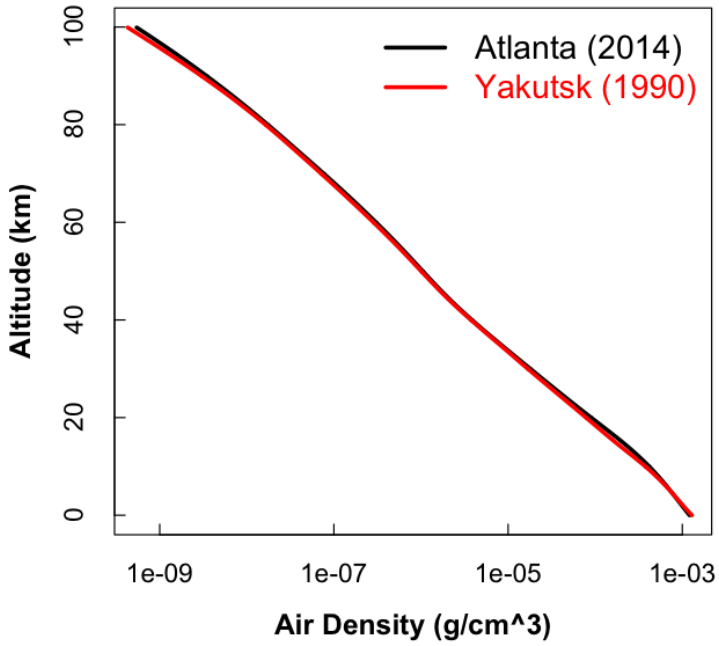

Figure (4.1) Air density at each altitude of Atlanta, GA, USA and Yakutsk, Russia. The black curve is air density vertical distribution at Atlanta, which are average values over the year of 2014. The red curve is air density vertical distribution at Yakutsk, which are average values over the year of 1990.

not considered since barely any interactions between air molecules and primary cosmic rays can happen in this region.

The primary cosmic particles used in the Geant4 consist of 79% protons and a small fraction of alpha particles and heavier nuclei [**?**]. Muon particles detected at ground are

predominantly low energy muons (10 GeV), which are mainly produced in the interactions between primary protons and air molecules. Muon flux data used in this work were all measured at ground, therefore only protons were included in the simulation as primary particles in this work.

The primary protons are launched at100 km from the ground pointing to center of the air shell (center of the Earth). The energy of the protons follows the energy spectrum described in Eq. 2.1 [**?**]. The maximum energy of the primary proton is set to 100 GeV since it is very rare to see muon particles at ground level that are produced in the interactions of protons with energy above 100 GeV. The lower bound of the proton energy spectrum is set to 3.6 GeV and 1.8 GeV for Atlanta and Yakutsk, respectively, which correspond to the rigidity cut-off at Atlanta (3.6 GV) and at Yakutsk (1.8 GeV). Shown in Figure 4.2 is the energy distribution of primary protons at Atlanta.

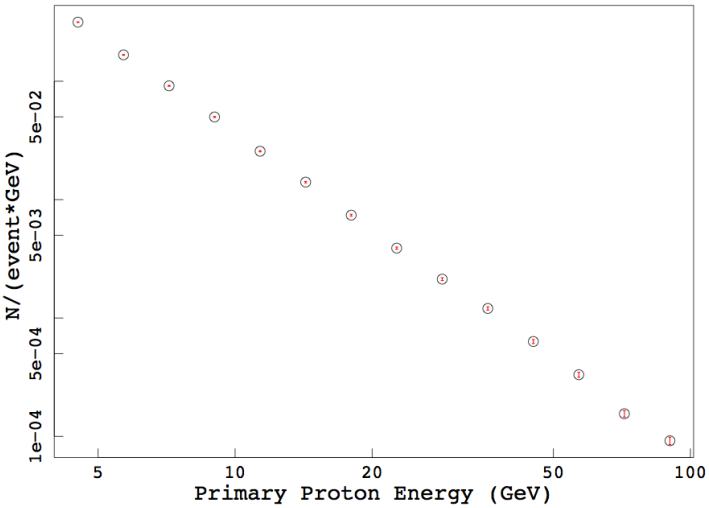

Figure (4.2) Primary cosmic ray proton energy distribution used in simulations for Atlanta.

A screen capture of the simulation display is shown in Figure 4.3. In this figure, the blue shell represents the partial atmosphere. The blue line pointing vertically downward is the trajectory of a primary proton particle with kinetic energy of 10 GeV launched from altitude of 100 km. The green lines are the gamma rays produced in the interactions.

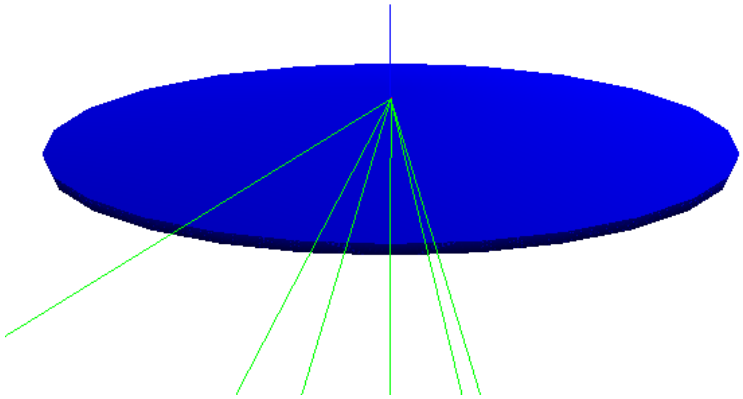

Figure (4.3) A screen capture of the Geant4 simulation program with a 10 GeV proton particle launched toward the surface of the Earth.

The simulation output includes particle ID, track ID, event ID, parent ID, location (x,y,z in geocentric frame), and momentum. One million protons are simulated for Atlanta and 10 million for Yakutsk in order to achieve reasonable statistics for determining the weighting functions.

4.2.2 Calculation of Weighting Function

Two quantities are required for calculating the weighting function from the simulation. One is the number of muons produced in i$^{\text{th}}$ layer which survive to the ground in each event ($\int_{p_i}^{p_{i+1}} dE_\mu P_\mu(p/g, E_\mu)$), and the other is the total number of muons that reach the ground in each event (I_μ). To simplify the notations, the former is labeled as N_i and the latter one is labeled as N_{total}. The weighting function is then

$$W_i = \frac{N_i}{N_{total}}. \tag{4.1}$$

The relative uncertainty of W_i is calculated as follows:

$$\sigma_{W_i} = \sqrt{(\frac{\sigma_{N_i}}{N_i})^2 + (\frac{\sigma_{N_{total}}}{N_{total}})^2 - 2 * \frac{\sigma_{N_i}}{N_i} * \frac{\sigma_{N_{total}}}{N_{total}}}, \tag{4.2}$$

where σ_{N_i} and $\sigma_{N_{total}}$ are standard errors of N_i and N_{total}.

Once the weighting function is determined, it is straightforward to calculate the effective temperature using the temperature profile measurement from radiosonde,

$$T_{eff} \approx \sum_{i=1}^{k} T_i \cdot W_i. \tag{4.3}$$

To keep consistent with the radiosonde measurements, The atmosphere is redivided into 10 isobaric layers, 20, 30, 50, 70, 100, 150, 200, 300, 500 and 700 hPa.

Figure 4.4 shows the weighting functions both for Atlanta and Yakutsk. Both weighting functions peak around the 100 to 200 hPa region (i.e, from tropopause and lower stratosphere) where most of the secondary muon particles are produced and in which most of the SSW events occur. The larger error bars in the upper and lower regions result from the relatively low statistics in muon counts. This indicates that the temperature variations in the tropopause and lower stratosphere region contribute more to the changes of the effective temperature.

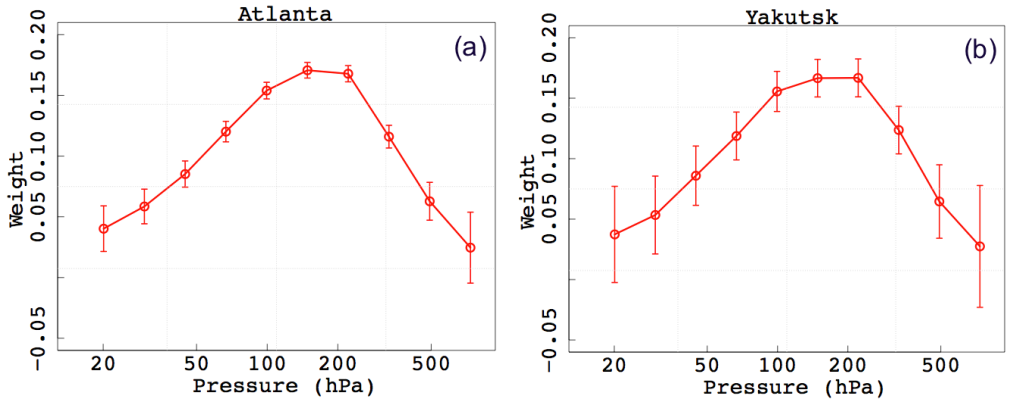

Figure (4.4) Weights for calculating effective temperature of ground level muon detector at (a) Atlanta and (b) Yakutsk.

4.3 Statistical Modeling Techniques

A statistical model contains parameters that need to be estimated based on assumptions about the model and the observed data. When describing observational, experimental, or survey data, statistical models are preferred over purely mathematical models. While the latter can be very important theoretical tools but the former can be more practical. Even if a mathematical model can be formulated for the phenomenon under study, a statistical model can provide a simpler and more straightforward description.

The properties of the model and the properties of quantities derived from it must be studied in a long-run, average sense through expectations, variances, and covariances. The process of estimating the parameters in a statistical model based on data is called fitting the model. Model parameters can be estimated by different regression techniques, such as elastic net regression and partial least squares regression. In this work, partial least squares regression is used and will be introduced below.

With given data, there exist a set of candidate models to describe the relationship among the variables. In the case of this study, a simple linear model can be used to fit the measurement of cosmic ray flux and effective temperature. Meanwhile models with higher orders of the muon flux variations and neutron flux variations as well as interaction terms between them can also be taken into account. There are several techniques can help with selecting a final model such as adjusted R^2, Mallow's C_p, AIC and BIC. In this analysis, k-fold cross validation was used to find the model that has the best prediction performance.

4.3.1 Regression Methodologies

In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables. Parameter estimation obtained by different regression methods typically have different statistical properties: distribution, variance, bias, etc. The choice between competing estimation principles is often made on the basis of properties of the estimators. Distinguishing properties might include (but are not necessarily limited to)

computational ease, interpretive ease, bias, variance, mean squared error, and consistency.

In this work, a series of general linear models are used as candidate models. A linear model either assumes that the regression function is linear, or the linear model is a reasonable approximation. The predictor variables can be any type of the following:

- quantitative inputs,
- transformations of quantitative inputs, such as log, square-root or square,
- higher order of inputs, leading to a polynomial regression,
- numeric or "*dummy*" coding of the levels of qualitative inputs.
- interactions between variables, i.e. production terms of two variables.

No matter the source of the predictors, the model is linear in the parameters.

The most popular estimation method of linear regression is Ordinary Least Squares (OLS) method, in which the coefficients were chosen to minimize the residual sum of squares (RSS). OLS can perform badly in some special cases such as those having outliers or multicollinearity. In this analysis, multicollinearity can be a big problem which is caused by the strong correlations between muon flux variations and neutron flux variations as well as barometric pressure fluctuations.

Methods like Principal Component Regression (PCR) and Partial Least Square Regression (PLSR) can overcome the multicollinearity problem. The PLSR generalizes and combines features from principal component regression (PCR) and multiple regression, and is for constructing predictive models when the predictor variables are many and highly collinear. The goal of PLSR is to predict or to analyze a set of dependent variables or response variables from a set of independent variables or predictor variables. This prediction is achieved by extracting from the predictors a set of orthogonal factors called latent variables which have the best predictive power [**?**]. The general idea of PLSR is trying to extract the latent factors, accounting for as much of the manifest factor variation as possible while modeling the

responses well. For this reason, the acronym PLSR has also been taken to mean "projection to latent structure".

In theory, PLSR has an advantage over PCR. In a situation where a minor component in predictors is highly correlated with response variables, not selecting enough components would then lead to very bad predictions. In PLSR, such a component would be automatically present in the first latent variables (LVs). Therefore, the PLSR was chosen as the regression method in this work. But one need to keep this in mind that there is barely any difference between the use of PLSR and PCR in practice.

4.3.2 Model Selection

Given a set of data, the objective of model selection is to determine the optimal model out of all candidates. An optimal statistical model is characterized by three attributes: 1) Goodness-of-fit, 2) Generalizability and 3) Parsimony (model simplicity). Since the purpose of this work is finding the model with the strongest predicting power, one of the most commonly used techniques for prediction error estimation, k-fold CV with root mean squared error (RMSE), is used for model selection in this analysis.

K-fold CV is a model validation technique for assessing how a statistical model will generalize to a new data set. It is mainly used when prediction is the purpose of the analysis and when the predicting performance of the model in practice needs to be evaluated. As shown in Figure 4.5, in k-fold CV, the whole dataset is randomly, non-repeatedly resampled into k equal sized subsets. Out of the k subsets, a single subset is retained as the validation data for model testing (RMSE in this study), and the remaining $k-1$ subsamples are used as training data (regression in this study). The cross validation process is repeated k times (k folds), with each of the k subsets used exactly once as the validation data. The estimated prediction error is the average value of all k folds. RMSEs of different CV can vary from one test to another, especially in the case of relatively small fold number (less than 10 folds). In order to obtain more stable results, k-fold CV over one dataset can be repeated multiple times, this method is called repeated k-fold CV.

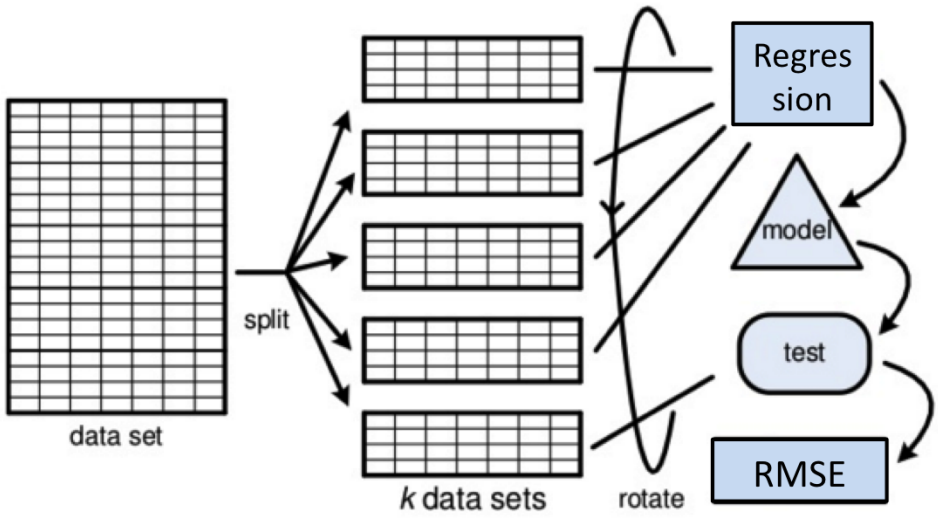

Figure (4.5) A schematic diagram of k-fold cross validation procedures, which includes re-sampling, getting models with regression on the training set (k-1 subsets), testing model performance with the holdout subsets and calculate the average RMSE across all hold-out tests. This cartoon is from Dr. Van Der Aalst's slides `http://www.slideshare.net/wvdaalst/process-mining-chapter03datamining`

The choices of the fold number (k) and repeat number (n) can be tricky. When using a large fold number, RMSE is usually small but its standard deviation is large (large error bar). By increasing the repeat number of the CV, the average RMSE and their standard deviations generally tend to become stable, but it needs much longer time to finish the calculations. Therefore, before the final calculation, values of k and n need to be determined at first.

4.3.3 Statistical Modeling with R

Most of the analysis in this work was carried out with a statistical software named R [**?**]. R is an integrated software for data manipulation, statistical analysis and data visualization. R offers an environment within which many classical and modern statistical techniques have been implemented. A few of these are built into the base R environment, but many are supplied as packages. The two most important packages used in this work are the "pls" [**?**] and the " caret" [**?**] package. The "pls" package implements PLS Regression in R. "plsr()"

is a function inside the pls package, which is commonly used for PLSR. The "caret" package (short for Classification And Regression Training) contains functions to streamline the model training process for complex regression and classification problems. The package utilizes a number of R packages including the "pls" package.

The package "caret" has several functions that attempt to streamline the model building and evaluation process, as well as feature selection and other techniques. One of the primary function in the package is the "train" function which can be used to 1) evaluate the effect of model tuning parameters on performance, 2) choose the "optimal" model across these parameters, 3) estimate model performance from a training set. Another function "trainCrontrol" can be used in the "train" function to specify the type of resampling and the resampling parameters such as "k" and "n" for the repeated k-fold CV.

To use the "train" function, besides specifying the model, dataset and regression method (pls in this work) to be used, there are a set of parameters to choose. For example, if fitting a PLSR model, the number of PLSR components to evaluate must be specified ("tuneGrid" in the function). Other parameters including the type of resampling (repeated k-fold cross-validation in this work) as well as resampling parameters (k and n in this work) should also be specified. All of these parameters can be controlled by the "trainCrontrol" function. At last, the performance metric which will guide the user to which tuning parameter values should be chosen. In this work, it is "RMSE". A example is listed below to show the basic syntax of the two functions.

```
caret.control <- trainControl(method = "repeatedcv", number = 3, repeats=30,
    returnResamp = "all", search = "grid")
caret.pls.out <- train(form=model, data=data, method = "pls", trControl =
    caret.control, metric = "RMSE", maximize = F, tuneGrid = expand.grid(.
    ncomp = 1:10))
```

CHAPTER 5

DATA ANALYSIS AND RESULTS

5.1 Introduction

The Nuclear Physics Group at Georgia State University has been monitoring cosmic ray flux since the year of 2009 with multiple generations of cosmic ray detectors. The modulation mechanisms of atmospheric factors, geomagnetic field as well as extraterrestrial conditions is being studied by the group. Correlations between the cosmic ray flux measurement of the group and extraterrestrial conditions including interplanetary magnetic field, solar proton density, solar plasma speed, plasma pressure and solar wind speed were carefully studied when a Forbush event happenes. Atmospheric effects such as barometric effect and temperature effect were also studied with both experimental measurements as well as Monte Carlo simulations. Currently, the group is developing a new Monte Carlo simulation program which will allow it to better understand the effects of both interplanetary magnetic field and geomagnetic field on different species of secondary cosmic rays.

Most of the researches mentioned above are fundamental studies, however, the purpose of this analysis is improving the development of meteorological applications. The focus of this work is to build predictive models which predict/reconstruct effective temperatures with cosmic ray intensity. The atmospheric temperature variation has very significant effect on muon flux variations and there exist a simple linear correlation between fluctuations of effective temperature and cosmic ray muon flux variations. The effective temperature method has been used to correct muon flux for many years by quite a few groups. It is also possible to solve the inverse problem: monitoring effective temperature with cosmic ray muon flux measurements. It is very challenging and impractical to find the analytical solution of this problem. One of the reasons is that the modulation mechanism of extraterrestrial factors on primary cosmic ray flux is still under studied both theoretically and experimentally.

Another reason is that measuring those factors itself is very expensive and difficult. In this work, statistical modeling techniques were used to develop models that predict/reconstruct effective temperature at two locations (Atlanta, GA, USA and Yakutsk, Russia) with cosmic ray muon flux measured by Four-Paddle detector as well as the muon telescope at Yakutsk.

This chapter has two parts. In the first section, sources of cosmic ray muon data, neutron data, local surface weather data and vertical profiles of atmospheric temperature are listed in details. The second section discusses the analysis procedures and results of the statistical modeling analysis on data recorded at Yakutsk. Inspired by the analysis results of Yakutsk's cosmic ray data, similar analysis was carried out on cosmic ray muon flux measured by Four-Paddle detector at GSU.

5.2 Data Sources

There are four types of data needed for this study: 1) muon intensity, 2) neutron intensity measured in the vicinity of the muon detector, 3) vertical distribution of atmospheric temperature measured in vicinity of muon detector, 4) barometric pressure at a point which close to the muon detector. Cosmic ray intensity measured by the Yu.G. Shafer Institute of Cosmophysical Research and Aeronomy at Yakutsk, Russia is one of the best early cosmic ray datasets that fulfills all requirements of this study. At Yakutsk, cosmic ray muon and neutron fluxes have been measured since 1960s. The barometric pressure data is not necessary for Yakutsk since their muon and neutron data are pressure corrected.

At GSU, Atlanta, Georgia, cosmic ray muon flux has been continuously measured by the four-paddle detector for almost two years. Since there are no recent neutron flux data can be found around Atlanta, neutron data measured by one of the most famous cosmic ray stations, cosmic ray station at Oulu, Finland are used in this analysis. Data of ground level barometric pressure at Atlanta, GA is available from Underground Weather Inc.. As for vertical profiles of atmospheric temperature, there are also weather stations in the city of Yakutsk and Atlanta, data of which are available from the world radiosonde network. More details are discussed bellow.

5.2.1 Data from Yakutsk

There are tens of cosmic ray muon stations all over the world. Some of the significant stations with their coordinates and operation starting time are listed in Table 5.1. More information can be found here: `http://cr0.izmiran.ru/gmdnet/`.

Table (5.1) Worldwide cosmic ray muon stations

Staion	Geoposition	Cut-off Rigidity (GV)	In Operation Since
Nagoya	35.12°N 136.97°E	11.5	1970
Sao Martinho	150.56°N 53.81°E	-	2006
Hobart	-42.90°N 147.33° E	1.88	1992
Yakutsk	62.02°N 129.72° E	1.65	1971
Greifswald	54.08°N 13.38° E	-	2006
YangBaJing	30.11°N 90.53° E	14.1	2007
Moscow	55.47°N 37.32° E	2.43	2007
Putre	18.20°N 69.55° E	14.0	2003
Yerevan	40.50°N 44.17° E	7.58	-
Mawson	-67.60°N 62.88° E	0.20	1982
BEO Moussala	42.18°N 23.59° E	6.31	2006
SWU Blagoevgrad	42.01°N 23.10° E	6.31	2007
Belgrade	44.85°N 20.38° E	5.30	2002
Santyago	-33.45°N 70.60° E	11.0	1960
Hafelekar	47.31°N 11.38° E	4.38	-
Adelaide	-34.93°N 138.58° E	-	2003
Guangzhou	23.60°N 113.18° E	16.0	1987-

The station that has both long term muon and neutron monitoring, as well as a weather station close by is the best choice for this analysis. From Table 5.1, it can be seen that stations at Nagoya, Yakutsk and Santyago have relative long data histories and can potentially be the best candidates. Unfortunately, Santyago's data are not available online and no neutron data and weather data can be found near the Nagoya station. Therefore the cosmic ray station at Yakutsk becomes the only choice for this analysis.

The Yakutsk (62.02°N 129.72° E, cut-off rigidity: 1.65 GV) muon telescope was built by Yu.G. Shafer Institute of Cosmophysical Research and Aeronomy. The telescope measures

the muon flux at ground level from both the vertical direction and 30° from the zenith. The historical data of this telescope cover the period from 1971 to 2003. For most of the time, the telescope was in good running condition except for several sudden counting rate jumps as shown in Figure 5.1. Those sudden jumps are most likely caused by hardware updates or repairs. In order to avoid additional errors related to hardware problems, three periods were chosen during which no sudden changes in counts can be found in the data preprocessing step. More information about the station as well as the telescope can be found here: `http://ikfia.ysn.ru/en/scientific-laboratories/9-uncategorised/846-muon-telescope.html`.

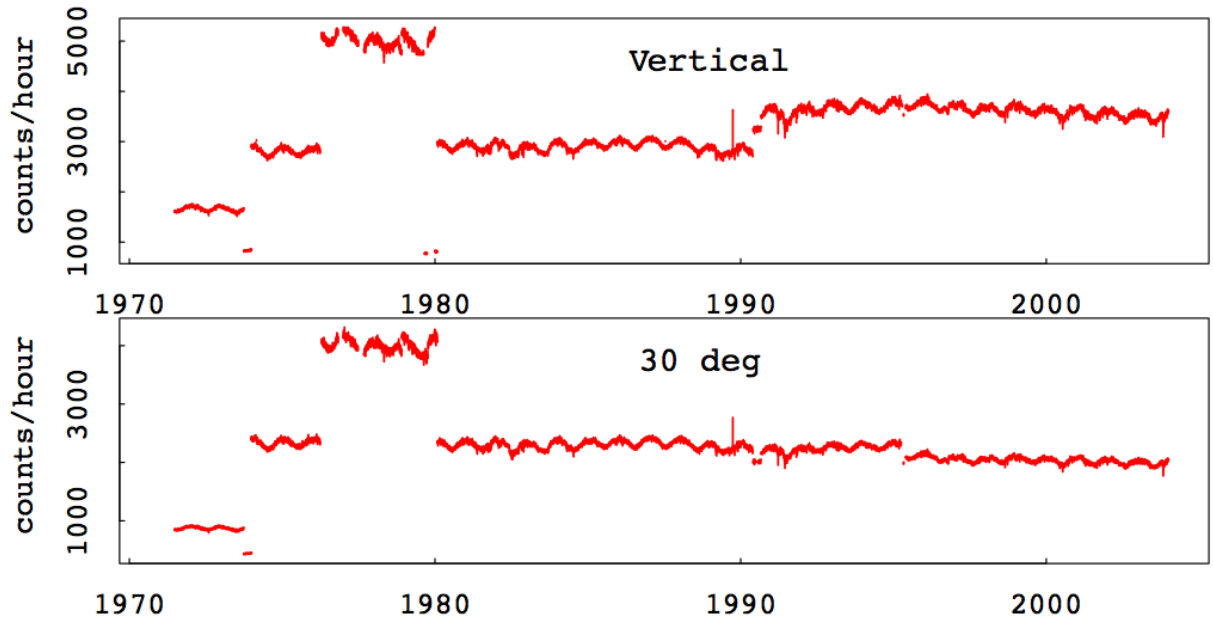

Figure (5.1) Cosmic ray muon flux (hourly counts averaged over a day, pressure corrected) measured by Yakutsk muon telescope on ground level.

Neutron flux has been measured by the same facility at Yakutsk for decades, which is shown in Figure 5.2. Neutron data from this station are available from the World Data Center for Cosmic Rays (WDC-CR) with station code "YAKUTS". Please note that both the muon and neutron flux data recorded in this station are pressure corrected.

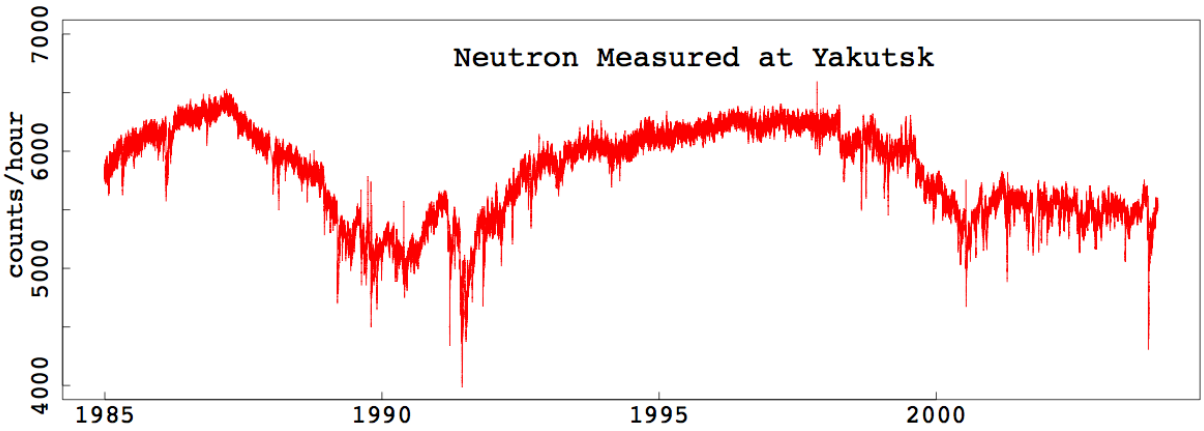

Figure (5.2) Neutron flux data (hourly counts averaged over a day, pressure corrected) measured by the Yu.G. Shafer Institute of Cosmophysical Research and Aeronomy.

5.2.2 Data Measured by the Four-Paddle Detector

The four-paddle detector has been measuring cosmic ray muon intensity at multiple locations. Data recorded from March, 2014 to December, 2015 were analyzed in this work. From 03/11/2014 to 10/04/2014/, Four-Paddle detector was in continuous operation at room 122 of Nature Science Center (NSC), GSU, with acceptance angles of $0°-12°$ (4-fold), $0°-17°$ (3-fold) and $0°-28°$ (2-fold). By subtracting the 4-fold counts from the 3-fold counts, one can get the counting rate of acceptance angles ranging from $12°$ to $17°$. Similarly, subtracting the 3-fold counts from the 2-fold counts results in the counting rate of acceptance angles from $17°$ to $28°$. Since one of the purposes of this analysis is to compare the angular effect with the model performances, muon flux from three non-overlapping acceptance angle ranges were derived with this operation. The time series of the three groups muon flux are shown in Figure 5.3.

In mid October 2014, the detector was moved to a place with much higher altitude, the 26th floor of 25 Park Place, with rearranged acceptance angle ranges of $0° - 23°$, $0° - 40°$ and $0° - 60°$. The four-paddle detector has been collecting data since then. Similarly, muon flux from three non-overlapping acceptance angle ranges, $0° - 23°$, $23° - 40°$ and $40° - 60°$, were derived from the measurements by the four-paddle detector at 25 Park Place, which

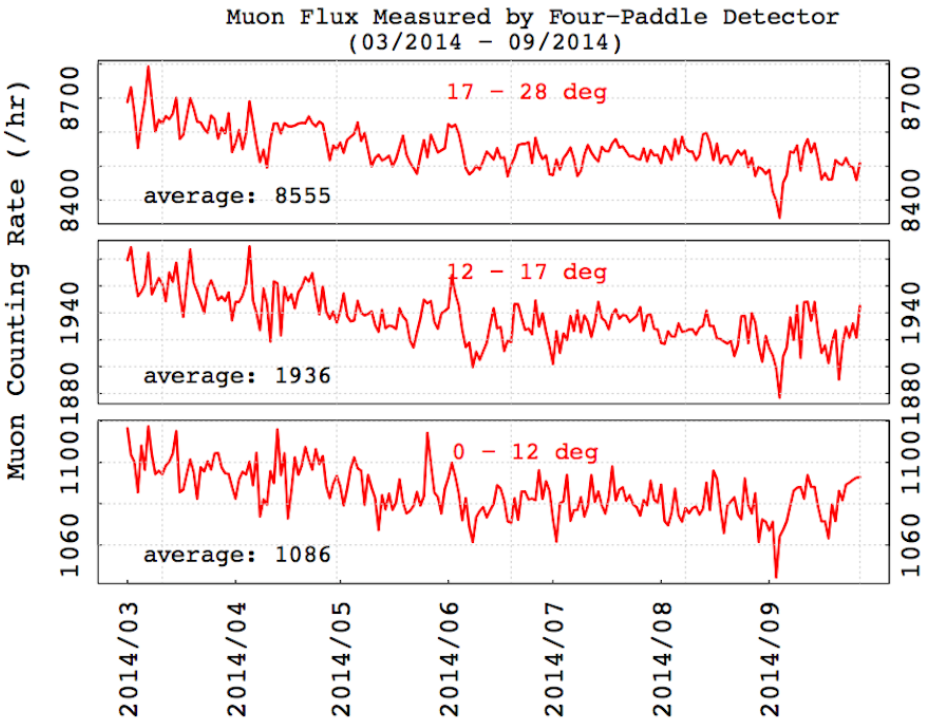

Figure (5.3) Muon fluxes from three acceptance angle ranges: $0°-12°$, $12°-17°$ and $17°-28°$. The three groups of data were derived from the measurements of the four-paddle detector, which was in operation 24/7 at Nature Science Center Room 122, Georgia State University, from 03/11/2014 to 10/07/2014. The muon counting rate are hourly counts averaged within a day, which leads to one data point per day.

are shown in Figure 5.4.

5.2.3 Neutron Data from Oulu

As discussed in Chapter 2, neutron flux variations can be used to estimate the primary effect on secondary muon flux variations, which mainly caused by solar activities. There are tens of neutron detector stations actively measuring neutron intensity time variations on the ground level. One of the most famous station is the Cosmic Ray Station operated by the Sodankyla Geophysical Observatory, which is located 120 km north of the Arctic Circle in Finland, is an independent department of the University of Oulu. Its data is available from `https://cosmicrays.oulu.fi/`. This station has been operating their neutron detectors since 1964, offering high quality neutron data for tens of years. Both the detectors and cosmic

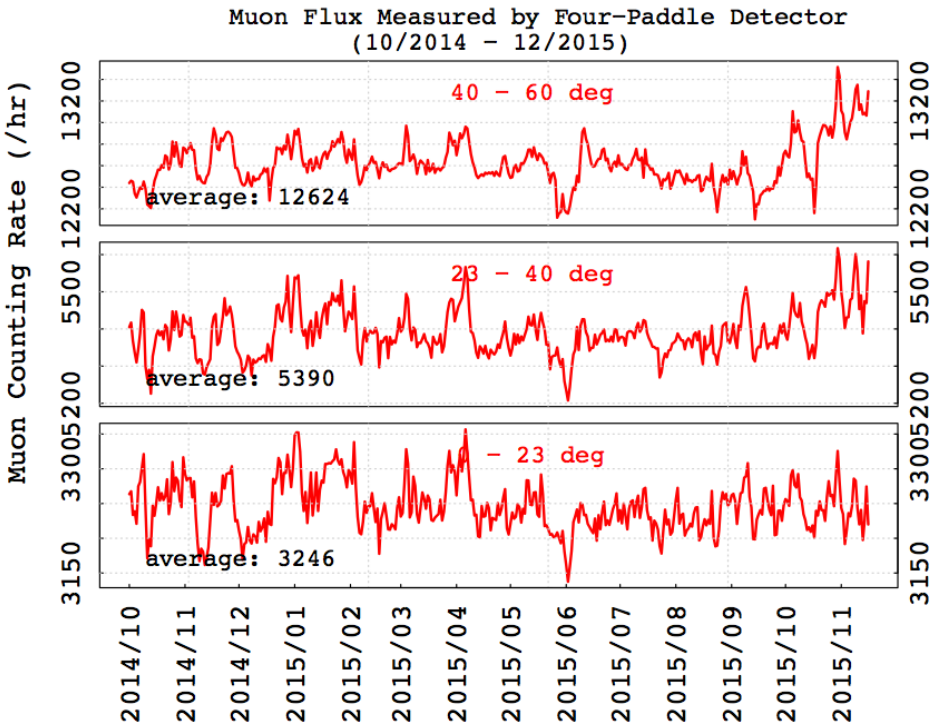

Figure (5.4) Muon fluxes from three acceptance angle ranges: $0°-23°$, $23°-40°$ and $40°-60°$. The three groups of data were derived from the measurements of the four-paddle detector, which was in operation 24/7 at the top floor (26th) of 25 Park Place, Georgia State University, from 10/24/2014 to present. The muon counting rate are hourly counts averaged within a day, which leads to one data point per day.

ray data have been very well maintained. The high data quality and longest overlapped data period with the four-paddle detector's data make the neutron flux measured by the Cosmic Ray Station at Oulu the choice in this analysis. The neutron fluxes measured at Oulu is shown in Figure 5.5. In the same way as the muon flux data, the neutron flux are daily averaged hourly counts, and are shown in two separated plots for the time periods that correspond to the two periods of muon data. Please note that this analysis uses the pressure corrected neutron data from Oulu.

5.2.4 Barometric Pressure Data from Underground Weather

The local barometric pressure variation is also included in the model, considering the significant barometric effect on the variations of muon flux. In this work, barometric pressure

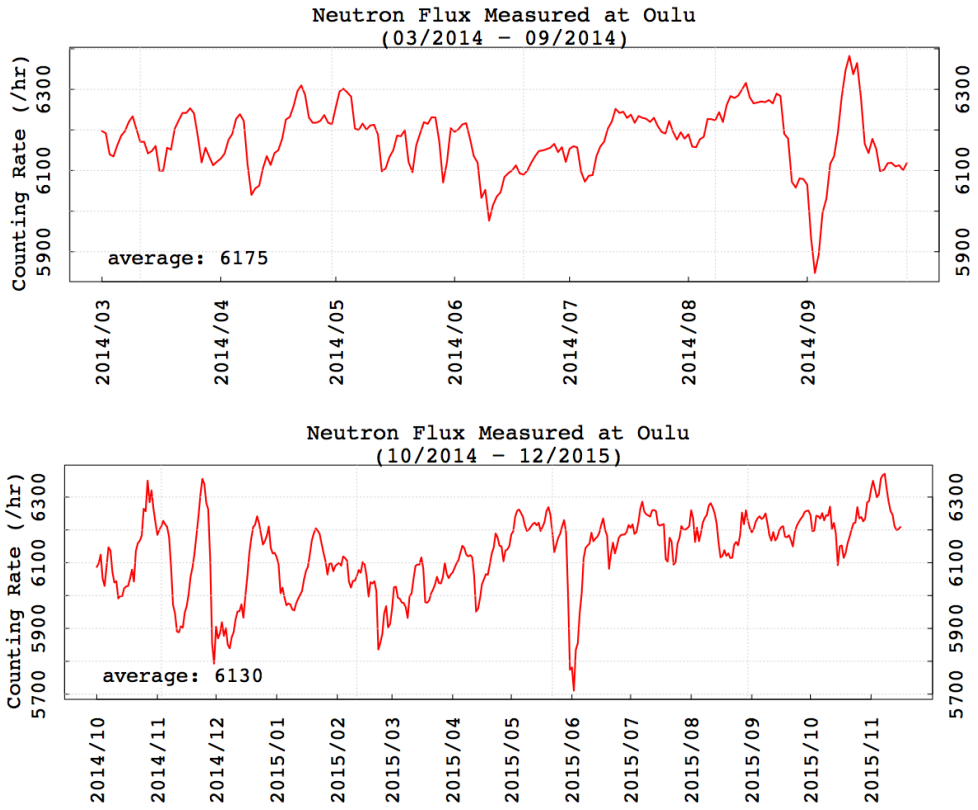

Figure (5.5) The daily averaged hourly neutron counts (pressure corrected) measured by the Cosmic Ray Station at Oulu, operated by the Sodankyla Geophysical Observatory. The top plot shows the neutron flux within the period of March, 2014 to September, 2014, and the bottom plot shows the neutron flux within the period of October, 2014 to December, 2015, which are corresponding to the two periods of the four-paddle detector's data.

data is measured by Fulton County Airport (FAA identifier FTY, Charlie Brown Field). Data are available online on the website "Weather Underground" `http://www.wunderground.com/`, which is operated by The Weather Company, LLC. The barometric pressure data are average values over a day. As shown in Figure 5.6, the data were also split into two time periods, March, 2014 to September, 2014 and October, 2014 to December, 2015, to be consistent with the muon data.

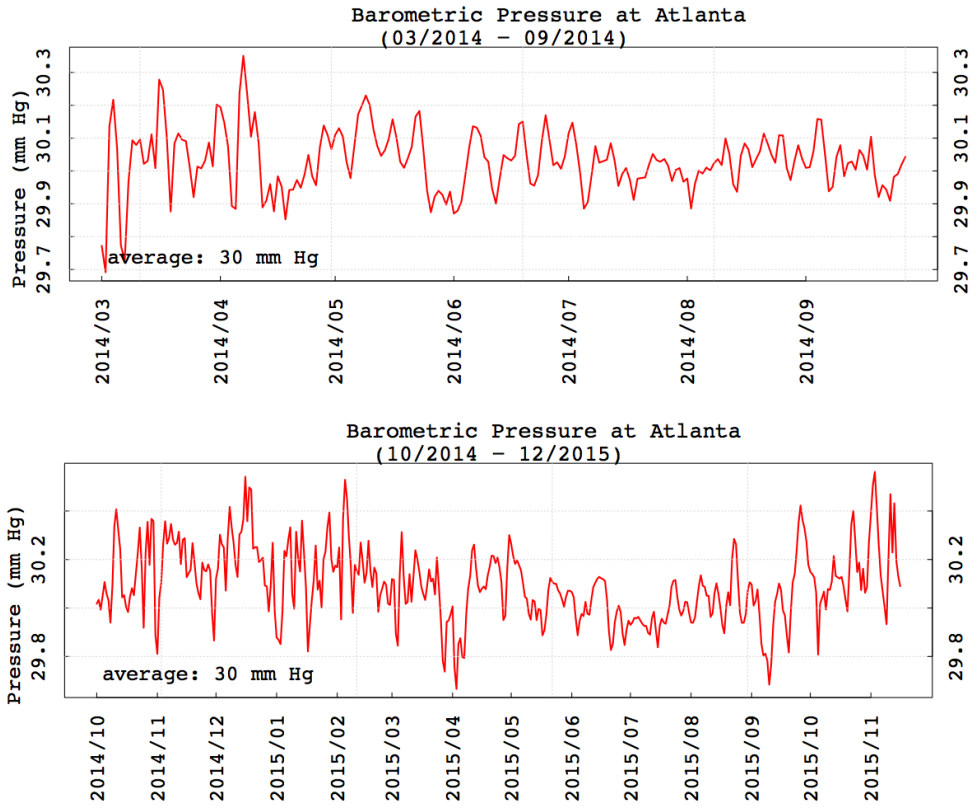

Figure (5.6) The daily barometric pressure data of Atlanta, GA in mm Hg. The top plot shows the data from March, 2014 to September, 2014, and the bottom plot shows the data from October, 2014 to December, 2015, which are corresponding to the two periods of the four-paddle detector's data. Measured by the Fulton County Airport (FTY, Charlie Brown Field) and available online from http://www.wunderground.com/.

5.2.5 Data from Radiosonde Network

The atmospheric profiles used in this study are from the Department of Atmospheric Science, Wyoming University `http://weather.uwyo.edu/upperair/sounding.html`. Wyoming University offers radiosonde data recorded by the world radiosonde network. There are two stations in the network which are in the greater Atlanta area and City of Yakutsk, respectively. Vertical temperature profiles used in the analysis of the four-paddle detecter's data are from Peachtree City Observations (station number 72215), which is 35 miles away from GSU campus. As for the analysis of Yakutsk's cosmic ray data, temperature profiles measured by Yakutsk Observations (station number 24959) are used, which is 7 km

away from campus of Yakutsk University.

Both stations launch air balloon carried radiosonde twice per day, at 00:00 am and 12:00 pm. The air balloons usually reach to more than 30 km (< 10 hPa). Because the data points measured by the radiosonde are not evenly distributed, the average temperature in each layer was calculated in the following way. At first, the atmosphere is divided into 10 isobaric layers, 20, 30, 50, 70, 100, 150, 200, 300, 500 and 700 hPa. Then a generalized linear regression with the temperature data in the pressure layer with the training equation $T = alog(P) + b$ is calculated, where P is the pressure at one data point and T is the temperature at that point, a is the correlation coefficients and b is the intercept. Then the temperature of the pressure layer is calculated by plugging the representing pressure (20, 30, ...,700 hPa) into the trained model in the second step. For every temperature profile, there are ten temperature data points (T_i) at ten isobaric layers. Using Eq. 4.3, $T_{eff} \approx \sum_{i=1}^{k} T_i \cdot W_i$, and weights obtained in chapter 4, the effective temperature at every trip of the balloon was derived. The results of effective temperature at Atlanta and Yakutsk are shown in Figure 5.7 and 5.8. Please note that the effective temperature shown in the two figures are averaged within a day, to be consistent with other data.

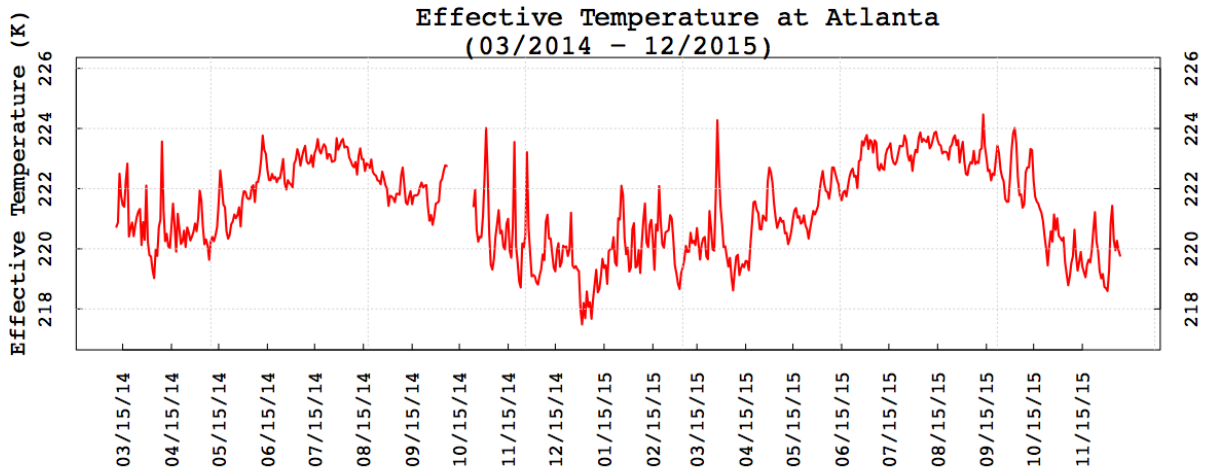

Figure (5.7) Effective temperature of muon flux in Atlanta from 2014 to 2015. Calculated with temperature profiles measure by Peachtree City Observations and weighting functions obtained with Geant4 simulations.

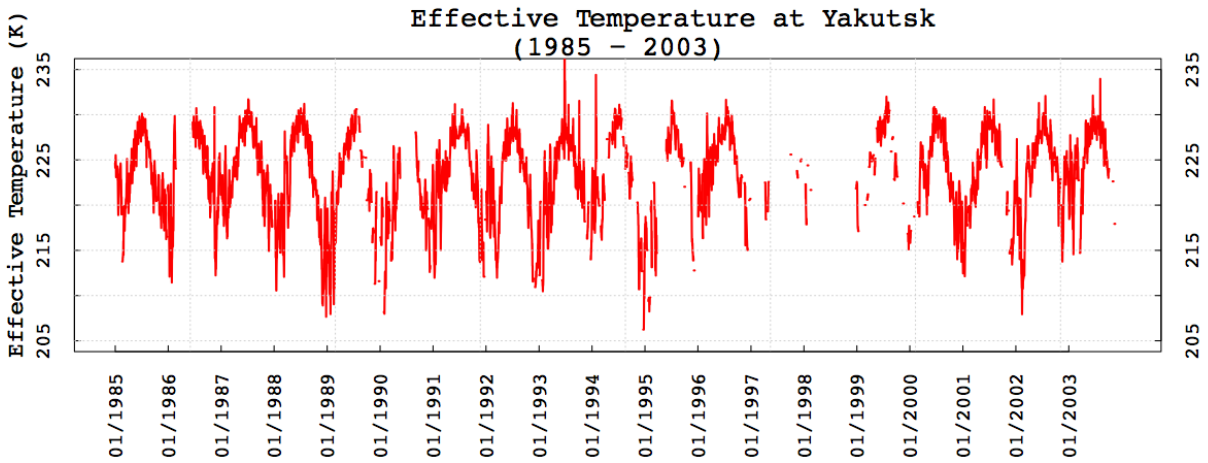

Figure (5.8) Effective temperature of ground level muon flux at Yakutsk from 1985 to 2003. Calculated with temperature profiles measure by Yakutsk Observations and weighting functions obtained with Geant4 simulations. There are some missing values in the time series especially during 1997 to 1999, which can be caused by failed flights of the balloons (did not reach to 15 hPa).

5.3 Statistical Modeling on Effective Temperature

In this section, the detailed analysis are discussed, which include candidate models, parameter selection of the cross validation, predicting performances of all candidate models and predictions with final models. In the candidate models, the response (dependent) variable is the percent variation of effective temperature (δT_{eff}). Predictor (independent) variables are the percent variation of muon flux (δI_μ), neutron flux (δI_N) and the barometric pressure (δP, in the case of the four-paddle detector). Both the higher order of the predictors, such as δI_μ^2 and δI_N^2, and interaction terms of them, such as $\delta I_\mu \cdot \delta I_N$ are also used as predictors. After carefully choosing the tuning parameters of the k-fold CV (number of folds and the repeat time of the CV), the predictive performances (RMSEs) of model candidates are compared. The models that have the least prediction errors as well as follow the rule of Occam's razor are the best models. Both the cosmic ray muon flux measured in Atlanta and Yakutsk are analyzed in this section. The effects of acceptance angles and locations on the predictive power of the statistical models are studied with both datasets. The chosen models are used

to predict the effective temperatures with cosmic ray data, which are compared with those calculated from radiosonde measurements.

5.3.1 Analysis of Data Measured at Yakutsk

Pre-processing of the Data As discussed in section 5.2.1, datasets from Yakutsk fulfills all most all of the requirements of this analysis. At Yakutsk cosmic ray station, muon flux, neutron flux and barometric pressure were measured in the same facility for decades. Atmospheric profiles were measured with radiosonde by a weather station that is less than ten kilometers away from the cosmic ray station. In addition, at Yakutsk, muon fluxes were measured from two zenith angles, 0° and 30° with similar counting rate. This offers a ideal data to compare the predictive power of the two directions. The pre-poccessing of the data is discussed below.

First, muon flux data measured in Yakutsk are divided into three periods: January 1985 to May 1990, September 1990 to March 1995, June 1995 to December 2003. Within the three periods, there are no apparent counting rate changes that might introduce significant systematic errors to the analysis. These three periods cover about two solar cycles (solar cycle 22, September 1986 - May 1996; solar cycle 23, May 1996 - January 2008). The 1985 - 1990 period covers the first half of solar cycle 22, 1990 - 1995 period covers the second half of solar cycle 22 and 1995 - 2003 covers the most time of solar cycle 23. Data from the three periods are analyzed independently.

Second, there are some factors that need to be excluded from the effective temperature data include missing values, noise in the effective temperature data and inconsistent periods. The effective temperatures are calculated with the atmospheric profiles measured twice every day. The balloons are usually launched twice per day at midnight and at noon, but the temperature in different pressure layers can some times change significantly within several hours. Therefore, the average of the two data points is not a good estimation of a whole

day. In this work, two weeks central moving average are used to overcome this problem.

$$I_{mai} = (\sum I_{i-7} + I_{i-6} + ... + I_i + ... + I_{i+6} + I_{i+7})/15 \tag{5.1}$$

As shown in Eq. 5.1, the two weeks central moving average of a time series is the mean of the current data point plus seven data before and after it. The moving average works similarly to a low-pass filter and eliminates the short term fluctuations. One of the disadvantages of using a moving average is that it generates missing values. In the case of Eq. 5.1, there are missing values generated for the moving average of the first seven data points and the last seven data points, because in those cases some of the subscriptions $i - 7$ to $i - 1$ or $i + 1$ to $i + 7$ are out of bound of i. In addition, when there is a missing value in the time series, there are 14 missing values in the moving average. This significantly reduced the sample size when there are some random missing values in the time series. Unfortunately, there are missing values in the calculated effective temperature which can be caused by hardware failure or adverse weather condition. In order to prevent too many missing values from being generated in the moving average of effective temperature, the missing values in the time series of original effective temperature dataset were replace with the average of the three data points before and after it (in total six data points). There are no missing values in muon flux data of the four-paddle detector, neutron flux data from Oulu and barometric pressure of Atlanta, therefore no extra processes were applied on these three data sets.

Third, it is necessary to use the same time frame for the muon flux, neutron flux and effective temperature data. All three data sets were matched with their time stamps and merged into a matrix. The whole matrix was chunked into three periods like the muon flux data. The data were aggregated within a day. Whenever there is a missing value in any row of the data matrix, the whole row was dropped. This is because of the package that calculates the PLS regression does not accept any missing values. The muon data (0°, 30°), neutron data and effective temperature in the data matrix are shown in Figure 5.9. Because both the muon and neutron data are pressure corrected, the barometric pressure was not

included in this analysis.

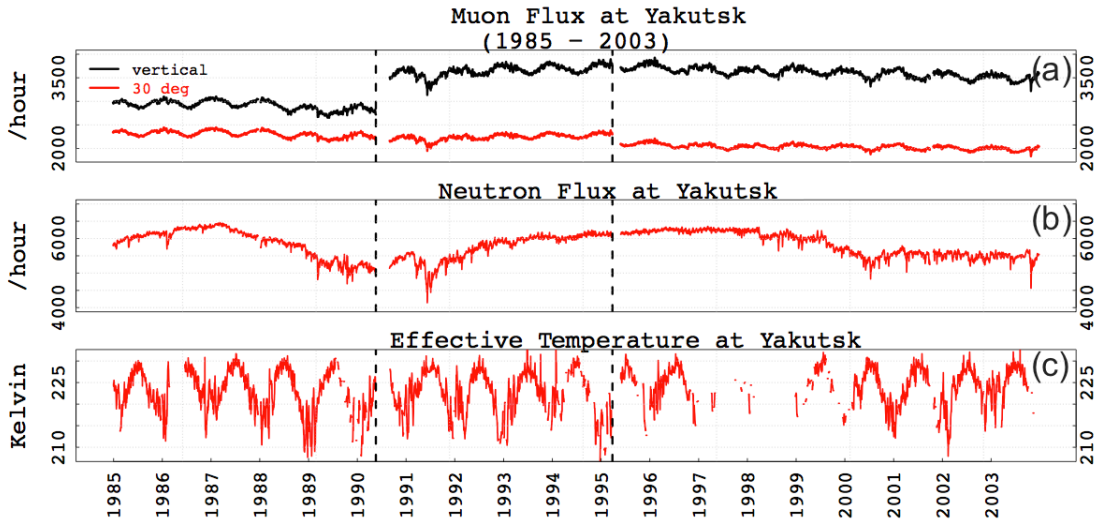

Figure (5.9) Time series of muon flux, neutron and effective temperature at Yakutsk from 1985 to 2003. All three data sets are chunked into three periods because of sudden muon counts jumps among the three periods, which are separated by black dashed lines in the plots. (a) Muon flux measured at ground level of Yakutsk (hourly counts averaged over a day). Black curve is muon flux from 0° zenith angle and red curve is 30°. (b) Neutron flux measured at Yakutsk (hourly counts averaged over a day). (c) Effective temperature (averaged over a day) calculated with weighting functions obtained with simulation and temperature profiles of Yakutsk.

Please note that the percent variations of the four variables were used in the regression. For each period, the reference point of the percent variations of each variable was chosen to be the average value of the variable within the period correspondingly.

Regression Models In the analysis of data from Yakuts, the predictors include percent variations of muon flux (δI_μ), neutron flux (δI_N). Higher order of the two variables and interaction terms of them were included as well: δI_N^2, δI_μ^2, $\delta I_N \cdot \delta I_\mu$, δI_N^3, δI_μ^3, $\delta I_N^2 \cdot \delta I_\mu$, $\delta I_N \cdot \delta I_\mu^2$.

Every combination of the terms listed above was tested. The combinations were grouped by number of predictors and everyone of them were labeled in the analysis code with a name in the format of "Ma.b", where a is the number of predictors in the model and b is the

sequence in the group. There are 9 groups and 511 candidate models in total in this analysis.

Model Selection As discussed in section 4.3.3, the estimation of RMSE and its σ can be affected by the two CV parameters, n (repeat times) and k (fold numbers). There are several general rules to follow when making choice of n and k. The choice of k is affected by the size of observations. It is recommended that choosing a divisor of the sample size as k value is a good practice. Larger k means less bias towards overestimating the true expected error (as training folds will be closer to the total dataset) but higher variance and higher running time. Another way of lower the variance can be repeating the CV with new random splits. But too many repeat times of CV can increase the risk of creating duplicated pattern/data structure. All of these general rules have more weights with small observation size, such as in the case of the four-paddle detector's data.

Since one of the purposes of this analysis is to compare the predictive power between the two incoming angels, the bias of the RMSEs are secondary to the variances. This analysis chooses 30 times repeated 3-fold CV to estimate RMSE of all 511 candidate models. Results of CV with muon flux data from the three periods are shown in Figure 5.10, 5.11 and 5.12 respectively.

First of all, it is shown from all three period that there is a big decrease in the RMSE from model M1.1 ($\delta T_{eff} = a \cdot \delta I_{\mu}$) to M2.1 ($\delta T_{eff} = a \cdot \delta I_{\mu} + c \cdot \delta I_N$). By including the neutron flux in the model can significantly increase the predictive power. This is true for muon flux from both directions and for all three periods. Second, although there are some decreases in RMSE with increasing model complexities, the changes in RMSE are relatively small. This trend is more apparent for Figure 5.10 and 5.12. It is true that M2.1 is not the true model of this analysis, of course, but it is also true that one can only find a model that has very similar predicting performance as the true model with the statistical modeling techniques. For all three periods, M2.1 has very similar predictive powers as other more complexed models, meanwhile its predicting performance is much better than simpler models. Therefore M2.1 was finally chosen as the predictive model for latter analysis.

Predictive Power vs Zenith Angles It has been shown in last section that model M2.1 was chosen as the statistical model for Yakutsk's dataset because of its advantages in predictive power and simplicity. It was then used in the comparisons of the predictive power of muon flux measured from two incoming angels for all three periods. Using 30 times repeated 3-fold CV and model M2.1, RMSE of all six datasets (2 incoming angles $\times$ 3 periods) were calculated. The results are shown in Figure 5.12.

It can be seen that except for (a) in Figure 5.13, both the 1991 to 1995 (b) and 1996 to 2003 (c) periods show that there is no significant differences in predictive power between the two directions. Because of the consistency between plot (b) and plot (c), the abnormal results of plot (a) should be interpreted as result of factors other than limitations of weighting functions such as detector efficiency changes or strong magnetic files effects. Moreover, the limitations of the weighting functions would cause reduced predictive power (larger RMSE) on larger incoming angles, therefore the smaller RMSE of 30° than 0° in plot (a) should not be the result of weighting functions. Now it is safe to say that weighting functions derived in this analysis do not cause any significant predictive power differences between muon fluxes coming from small zenith angles (0° to 30°).

Effective Temperature Reconstruction with Yakutsk's Data PLS regressions were carried out to estimate the coefficients (a and b) of M2.1 for all three time periods. A and b are -0.605 and 0.285 respectively for period 1985 - 1990, -0.828 and 0.221 for period 1990 - 1995 and -0.687 and 0.257 for period 1995 - 2003. Using these three models and cosmic ray flux variations (muon and neutron), the effective temperatures in the three periods were rebuilt. As shown in Figure 5.14, the black dots are the effective temperature that calculated with the atmospheric temperature vertical profiles measured by radiosonde and the weighting functions that derived from simulation. The red dots are the effective temperature that predicted with cosmic ray muon and neutron flux. It can be seen that the predicted effective temperature catches most of the variations patterns of those derived from radiosonde measurements, especially the seasonal variations. Besides, variation pat-

terns happened in medium terms (several weeks) like the sudden increases in the effective temperature during winter time (possibly SSWs) can also be seen in both. The radiosonde derived effective temperature show many local (several days) variations which are more violent during winter time (troughs of the wave). Due to the limited weather balloon launch times in everyday (twice), the calculated effective temperature of everyday is not a good representative of the daily averaged effective temperature. This is more apparent during winter time since atmospheric temperature show more changes in winter. Therefore these local variations should not be expected in the prediction. Larger frequency of atmospheric profile measurements are needed in order to test the performance of the models on predicting short term variations.

5.3.2 Data Analysis for the Four-Paddle Detector

Preprocessing of the Data As mentioned in section 5.2.2, cosmic ray flux measurement was carried out by the four-paddle detector at two locations (basement of NSC and top floor of 25 Park Place) and muon intensity coming from different zenith angles can be derived from data of both locations. On the one hand, because of the attenuation and absorption to muon particles by materials in the building such as concrete and bricks, the mean energy of muon particles that were detected in room 122 of NSC leans towards to the higher end of the energy spectrum. On the other hand, mean energy of muon particles detected on the 26th floor of 25 Park Place is more tilted to the lower end since more low energy muon particles were detected by the detector. By comparing the predictive power of the two datasets, it is possible to find a position for muon flux measurement that is more suitable for effective temperature monitoring. Similarly, the predicting performances of muon flux coming from different zenith angles can also be evaluated. In order to make reasonable comparison, several preprocessing procedures were carried out to eliminate possible errors caused by irrelevant factors.

It is necessary to use the same time frame for the muon flux measurements at the two locations to ensure that the two datasets have the same amount of data points and within the

same season in a year. The measurement at basement of NSC took place from 03/18/2014 to 09/30/2014. The top floor measurement started from 10/27/2014 and maintained continuously operation till 12/07/2015. So the final data used in the regression analysis were chosen to be within the period of 03/18/2014 to 09/30/2014 (basement) and 03/18/2015 to 09/30/2015 (top floor).

Similarly, the percent variations of the four variables were used in the regression. For each period, the reference point of the percent variations of each variable was chosen to be the average value of the variable within the corresponding period.

Regression Models In the analysis of the four-paddle detector's data, the predictors include percent variations of muon flux (the four-paddle, δI_μ), neutron flux (Oulu, δI_N), barometric pressure (Underground Weather, δP). I also considered higher order of the three variables and interaction terms of the three variables: δI_N^2, δP^2, $\delta I_N \cdot \delta P$. Since it has been proven that the correlation between the percent variations of effective temperature and percent variations of muon flux is linear, the higher order of δI_μ and any interaction terms include it were not taken into account.

Every combination of the terms listed above was tested. The combinations were grouped by number of predictors and everyone of them were labeled in the analysis code with a name in the format of "Ma.b", where a is the number of predictors in the model and b is the sequence in the group. For example, there are 20 candidate models in the group of three ($C_6^3 = \frac{6!}{3!3!} = 20$), and model "M3.1" is $\delta T_{eff} = a \cdot \delta I_\mu + c \cdot \delta I_N + b \cdot \delta P$, where a, b, c are coefficients to be determined with regression. There are 6 groups and 63 candidate models in total in this analysis of the four-paddle detector data.

Optimizing CV Parameters In order to find the optimized n and k, tests with the four-paddle detector's data were carried out. Cosmic ray flux with incoming angle of $17^\circ - 28^\circ$ (2-fold minus 3-fold) during period 03/18/2014 - 09/30/2014 was used in the tests. The regression model is $\delta T_{eff} = a \cdot \delta I_\mu + c \cdot \delta I_N + b \cdot \delta P$. As shown in Figure 5.15 (a) and (b), the larger the subset number, the less the RMSE but the larger the standard

deviation of RMSE. From Figure 5.15 (c) and (d), it can be seen that by increasing n, neither RMSE nor σ shows any decreasing trend. But except for tests of 2 and 3 folds, both RMSE and σ of other tests are relatively flat with n larger than 30. Although using more folds can decrease the bias on estimation of prediction error, it is difficult to distinguish the predictive power of candidate models when σ is too large. In this study, the main purpose is to make comparison of candidate models' predictive performances, which means bias on the estimation is secondary to the size of σ. Therefore the value of k was finally chosen as 5 and n was 30.

Model Selection RMSE of all 63 candidate models were estimated with 30 times repeated 5-fold CV. Muon flux measured at both the NSC and 25 Park Place were used in the tests. Results of CV with muon flux data from NSC are shown in Figure 5.16. More informations of the plot can be found in the caption of the plot.

It is shown in Figure 5.16 that the minimum RMSE of each group decrease with more predictors and become flat at group three. This means that some models with three predictors are more accurate than any of those with less predictors and their predictive performances should be at least as good as more complicated models (models with more predictors). Following the rules of Occam's Razor: to choose the simplest model in the candidates that adequately accommodates the data, that is group three, group of models with three predictors.

Although Figure 5.16 shows that model "M3.5" ($\delta T_{eff} = a \cdot \delta I_\mu + c \cdot \delta I_N^2 + b \cdot \delta P$) has the minimum RMSE value in group three, it worth a closer look at the results of the whole group, which is shown in 5.17. Model M3.5 has the smallest RMSE but not significantly smaller than other models such as M3.1 ($\delta T_{eff} = a \cdot \delta I_\mu + c \cdot \delta I_N + b \cdot \delta P$), M3.6 ($\delta T_{eff} = a \cdot \delta I_\mu + c \cdot \delta P + b \cdot \delta P^2$) and M3.7 ($\delta T_{eff} = a \cdot \delta I_\mu + c \cdot \delta I_N + b \cdot \delta P \cdot I_N$). If taking the error bars into account, it is safe to say that model M3.1, M3.5, M3.6 and M3.7 have similar performances. Here, again, rule of Occam's Razor comes into effect. The simplest model M3.1 (has neither higher orders nor interaction terms) was chosen as the final model for the dataset of NSC and for all three

muon flux incoming angle ranges.

Results of CV with muon flux data from 25 Park Place are shown in Figure 5.18. Again group three shows a much better predictive performance than group one and two and as good as group four to group six. Clearly group three stands out from all six groups, in a more significant way than the dataset of NSC. The comparison among all models with three predictors are shown in Figure 5.19. For this dataset M3.1 has much smaller RMSE than any other models. And this conclusion is true for tests using data of muon flux from all three incoming angle ranges.

To summarize the model selection section for the the four-paddle detector's data, either the muon flux was measured in the basement (NSC room 122) or on the top floor (25 Park Place), and no matter which zenith angle the muon particles are coming from, effective temperature of Atlanta area can be predicted/reconstructed using model M3.1 ($\delta T_{eff} = a \cdot \delta I_{\mu} + c \cdot \delta I_N + b \cdot \delta P$), with predictive errors no less than any other models tested.

This choice of the model is actually very consistent with the model chosen for dataset of Yakutsk. Since both the muon flux and neutron flux in Yakutsk's datasets are pressure corrected, it is equivalent as having the barometric pressure in model M2.1.

Predictive Power vs Locations and Acceptance Angles It has been shown in last section that model M3.1 is the best model out of all 63 candidates in the aspects of both predictive power and simplicity. M3.1 was then used in the comparisons of the predictive power of muon flux measured at two locations and from multiple incoming angels. Using 30 times repeated 5-fold CV and model M3.1, RMSE of all six datasets (3 acceptance angles $\times$ 2 locations) were calculated. Note that the same data of neutron flux, effective temperature and barometric pressure were used in all of the CV. The results are shown in Figure 5.20.

Three patterns can be seen from Figure 5.20. First, RMSEs of NSC 122 datasets (red) show a moderate increasing trend with decreasing incoming angle. Second, there is clearly a decreasing trend in the RMSEs of 25 Park Place datasets (black) with incoming angle increases. Third, there are no apparent difference between the RMSE of 2-3 fold of NSC 122

datasets and 4-fold of 25 Park Place datasets. These patterns are results of two variables in this analysis, location and acceptance angel.

There are two possible mechanisms of how the variations in acceptance angle can affect the predictive power. At first, there might exist angular dependence of temperature effect on muon flux variations. Those muon flux with temperature effect more pronounced would have higher signal-to-noise ratio than others and thereby lead to better predicting performance. Moreover, effective temperature calculated in this analysis is more correlated with muon flux coming from small zenith angles than large zenith angles. The weighting function used in this analysis were obtained with simulations, in which the primary particles were launched vertically down to the surface of the Earth and the incoming zenith angles of most of secondary muon particles that arrived to the ground are smaller than 5 degree. In general, this should be a reasonable estimation for muon fluxes from small incoming angles but not as accurate for large incoming angles.

Another variable in the analysis is location. As discussed in section 5.3.2, muon flux measured in NSC room 122 should have higher average energy than those in 25 Park Place because of the attenuation and absorption of the materials in NSC building. It is known that negative temperature effect is stronger on low energy muons. Under the condition of having same counting rate, muon flux of lower average energy should have higher signal-to-noise ratio than those of higher average energy and therefore results in smaller RMSE in the CV test. In addition, weighting function can also play a role in the location caused predictive power differences. In the simulation, the attenuation or absorption effects by materials in the buildings were not implemented. Those effects are stronger for the NSC 122 datasets, because of more building materials above the detector when it was in NSC. Theoretically, the weighting function derived in this analysis should be more accurate for the 25 Park Place datasets and leads to smaller RMSE of 25 Park Place datasets than that of NSC datasets.

In this analysis, both angular dependence of temperature effect and limitation of the weighting function can play important roles. In order to interpret the patterns in Figure 5.20, correlation coefficients of temperature effect were calculated with multi-variate regressions

and model $\delta I_\mu = \alpha \cdot \delta T_{eff} + \gamma \cdot \delta I_N + \beta \cdot \delta P$, where α, β and γ are the temperature coefficient, barometric coefficient and primary coefficient respectively. Results are shown in Figure 5.21. Please note that temperature effects for all of the datasets are negative effect, which means the smaller the temperature coefficient (α), the stronger the temperature effect. The decreasing trend in α of NSC 122 datasets with decreasing acceptance angels indicates that temperature effect is stronger with smaller acceptance angle. And pattern in 25 Park Place datasets suggests that temperature effect is stronger on small (4-fold, $0^\circ - 23^\circ$) and large (2-3 fold, $40^\circ - 60^\circ$) acceptance angels and weaker on medium (3-4 fold, $23^\circ - 40^\circ$) acceptance angle. In addition, except for the 2-3 fold, there is no significant difference on α between the two locations.

As discussed above, a stronger temperature effect should result in a smaller RMSE and the limitation of the weighting function in this analysis should lead to smaller RMSE (more predictive power) on smaller acceptance angles. The first pattern in Figure 5.20, RMSEs of NSC 122 datasets increase with decreasing acceptance angles, is opposite to the anticipation of both factors. The temperature effect of NSC 122 datasets have increasing temperature effect with decreasing acceptance angles (5.21, red), which should cause decreasing RMSEs in Figure 5.20 (red). Meanwhile, the limitation of the weighting function should also lead to decreasing RMSEs in Figure 5.20 (red). The current results suggest that neither the strength of the temperature effect nor the limitation of weighting functions dominates the effects on predictive power of muon flux measured in room 122 of NSC. The possible reason for the decreasing trend is the difference in counting rate of the muon fluxes. For the muon fluxes that have small counting rate, their signal-to-noise ratio should be small as well. Therefore the significantly dropped average counts in 3-4 fold and 4-fold cause the weaken predicting performances, i.e. larger RMSE. In conclusion, muon flux counting rate dominantly affects the predicting performances of muon flux measured in room 122 of NSC. This indicates that weighting functions derived in this analysis do not cause significant predictive power differences for incoming zenith angles ranging from 0° to at least 28°. It is also suggested that variations in predicting performances caused by strength of temperature effects is secondary

to those caused by the changes in counts. This trend in the RMSEs of NSC 122 datasets might be a combined effects of all three factors, but it is also possible that the small counting rate caused predictive power decrease is moderate and the effects of the other two factors are negligible. This will be further studied with the Yakutsk's datasets.

Similarly, the second trend in Figure 5.20, RMSEs of 25 Park Place datasets decrease with decreasing acceptance angles, should be mainly caused by the limitation of the weighting function. The arch shape of temperature effect shown in Figure 5.21 (black) suggests that its effects on predictive power of this group of datasets are not as significant as weighting function. In addition, the decreasing RMSEs with decreasing acceptance angles shows that the counting rate drop on the 3-4 fold and 4-fold do not affect the predicting performance as much as other facotors. So for the 25 Park Place datasets, the limitation of weighting function dominates the effects on predicting performances. Again, the changes in the strength of temperature effect do not make important contribution to the predictive power. And it also suggests that for the current predictive model, a counting rate of 3200 per hour might be sufficient.

The third pattern, RMSE of 2-3 fold ($17^\circ - 28^\circ$) in NSC 122 datasets is very similar to RMSE of 4-fold ($0^\circ - 23^\circ$) in 25 Park Place datasets. Although it is not a exact apple to apple comparison, it is still worth to take a look. As discussed above, the variations of temperature effect strength do not have significant effect on the predictive power of the muon flux. In addition, a counting rate of 3200 per hour is sufficient for the predictive model in this analysis. So the effects of these two factors on predictive power of muon fluxes can be ignored in this comparison. With excluded the two factors, the third pattern shows that using the weighting function for measurements in both NSC room 122 and the top floor of 25 Park Place do not significantly change the predictive power of the muon fluxes, as long as the acceptance angle is at least smaller than 28°.

In summary, the weighting function obtained from simulation in this analysis maybe not significantly affect the predictive power of muon fluxes with small incoming zenith angles (less than 28°, will be verified with data from Yakutsk). In addition, the predicting performance

would moderately decrease when the counting rate are too low. Moreover, the variations in the strength of temperature effects do not play a important role in this analysis. The weighting function can be used for the analysis of data measured both on the top floor and in the basement in this study without causing any significant predictive power decreasing, as long as the incoming angles are small enough and counts are large enough.

Effective Temperature Reconstruction with Four-Paddle Detector's Data It is shown above that out of 63 multi-variate polynomial models, model M3.1 is the best one to use for effective temperature prediction/reconstruction. It is also shown that muon flux coming from zenith angles less than 28° with sufficient counting rate has the highest precision when predicting effective temperature.

Muon flux of 4-fold coincidence (0° - 23°) during period of 03/28/2015 - 09/30/2015 is the best dataset for prediction, and all datasets during the period of 03/28/2014 - 09/30/2014 have very similar predicting performances. To be consistent, muon flux of 4-fold coincidence of both periods were used to train model M3.1 and to make predictions to effective temperature at Atlanta.

The coefficients (a, b and c) in model M3.1 were obtained for both periods with PLS regression and data in each period. Eq. 5.2 is the final model for period of 2014 and Eq. 5.3 is for period of 2015.

$$\delta T_{eff} = -0.239\delta I_{\mu} + 0.044\delta I_N - 0.505\delta P \tag{5.2}$$

$$\delta T_{eff} = -0.369\delta I_{\mu} + 0.160\delta I_N - 0.850\delta P \tag{5.3}$$

The effective temperature can be calculated with the following equation,

$$T_{eff} = (\delta T_{eff}/100 + 1) \cdot \bar{T}_{eff} \tag{5.4}$$

where $\bar{T}_{eff}$ is 221.8 K for 2014 period and 222.1 K for 2015 period. The effective temperature were calculated for periods 03/11/2014 - 10/07/2014 and 10/24/2014 -12/09/2015 with Eq.

5.2 and 5.3 respectively. The results are shown in Figure 5.22.

As shown in Figure 5.22, while the general trend of the effective temperature are captured by the reconstruction, plenty of discrepancies can be seen between the ones calculated with radiosonde data and the ones calculated with the predictive model and cosmic ray flux. There are several possible sources of the discrepancies. The first one is that neutron flux were not measured simultaneously at the same location. As mentioned in early section, neutron flux data used in this analysis was measured at Oulu, Finland. Because of the low geomagnetic rigidity cut-off at Oulu (3.6 GV for Atlanta and 0.8 GV for Oulu), neutron flux measured at Oulu have more fluctuations caused by solar activities than that of Atlanta, which causes more discrepancies. The second one is the insufficient training period. There are relative large discrepancies during the periods that were not included in training. This may suggests that all seasons in a year should be included when train the model. It is reasonable since the effective temperature has seasonal variation with a period of a year.

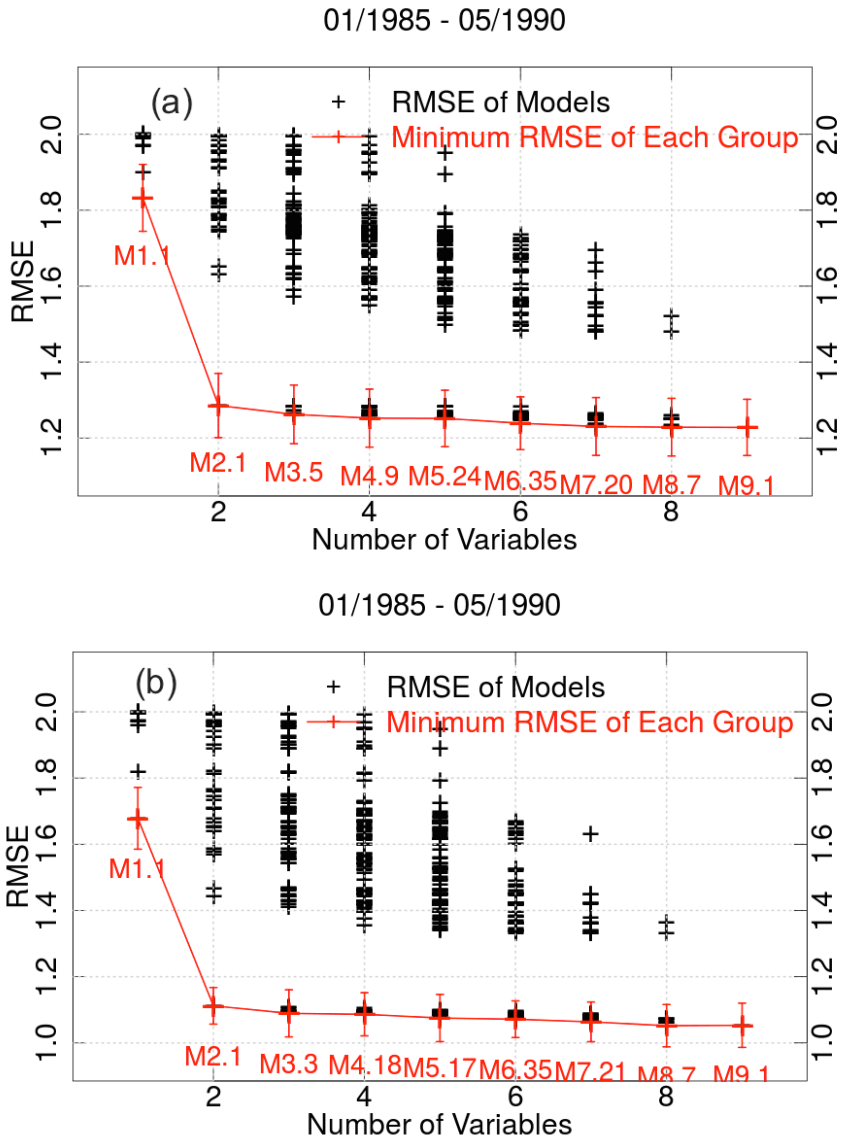

Figure (5.10) RMSE of all 511 candidate models calculated with 30 times repeated 3-fold CV. The candidates are divided into nine groups according to their number of predictors. Each of the black or red plus symbol represents one candidate. The models that have the least RMSE in their group are plotted with red plus, whose σ and names (such as M1.1) are all labeled with red color. In plot (a) and (b), muon flux data used in the test are those with incoming zenith angles of 0° and 30° respectively. Also, both plots used same effective temperature and neutron flux. The period of the datasets are from 01/01/1985 to 05/24/1990.

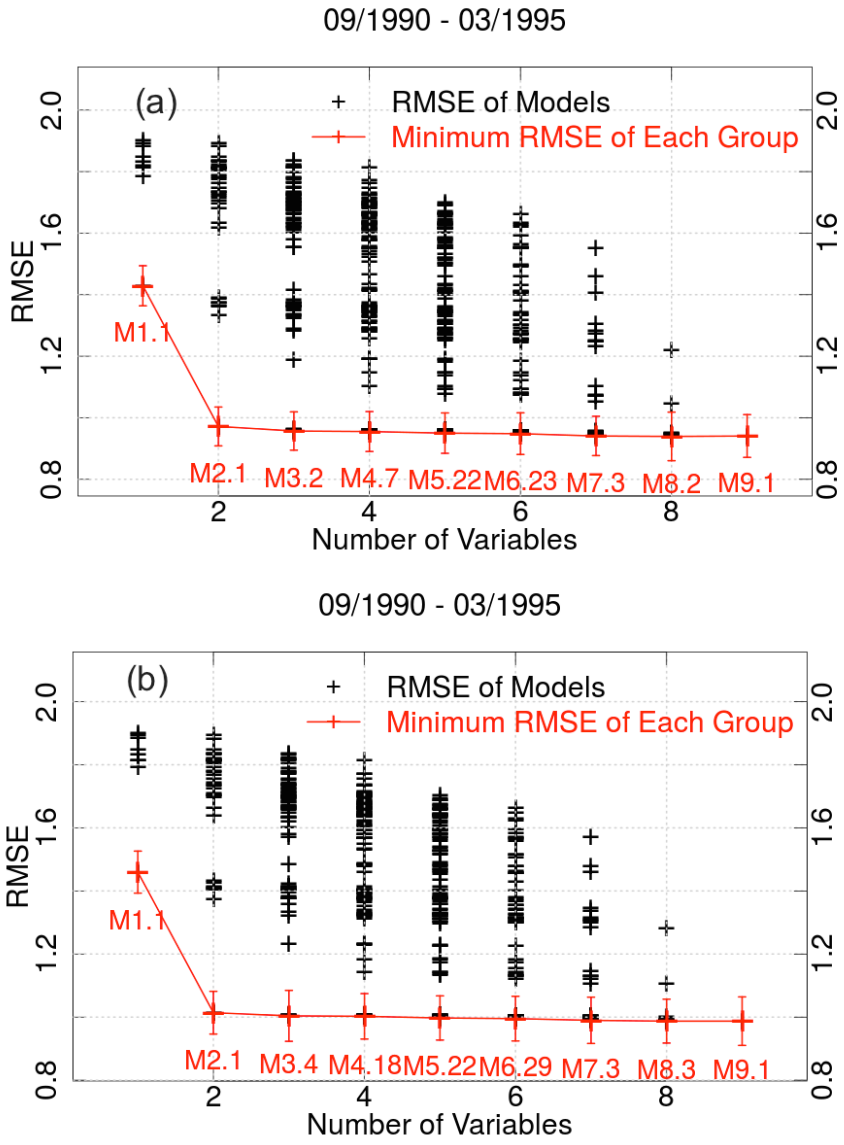

Figure (5.11) RMSE of all 511 candidate models calculated with 30 times repeated 3-fold CV. The period of the datasets are from 09/03/1990 to 03/31/1995.

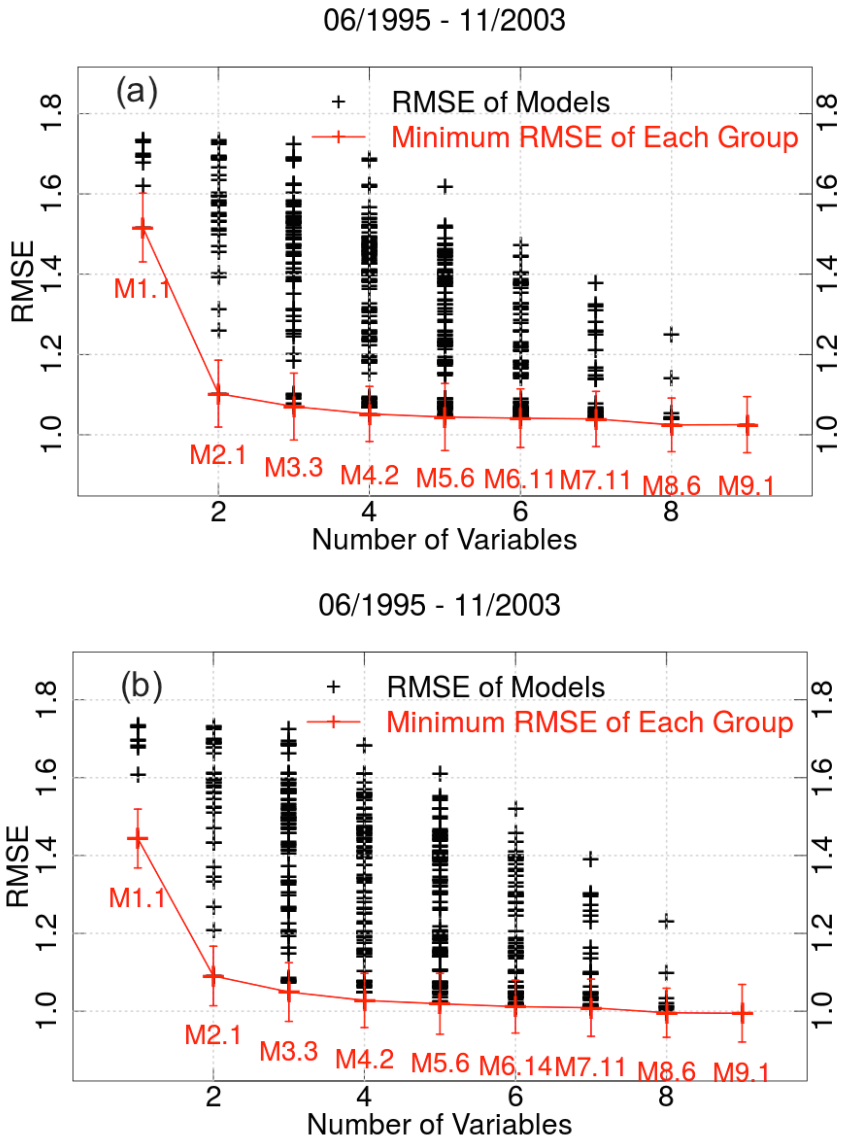

Figure (5.12) RMSE of all 511 candidate models calculated with 30 times repeated 3-fold CV. The period of the datasets are from 06/01/1995 to 11/30/2003.

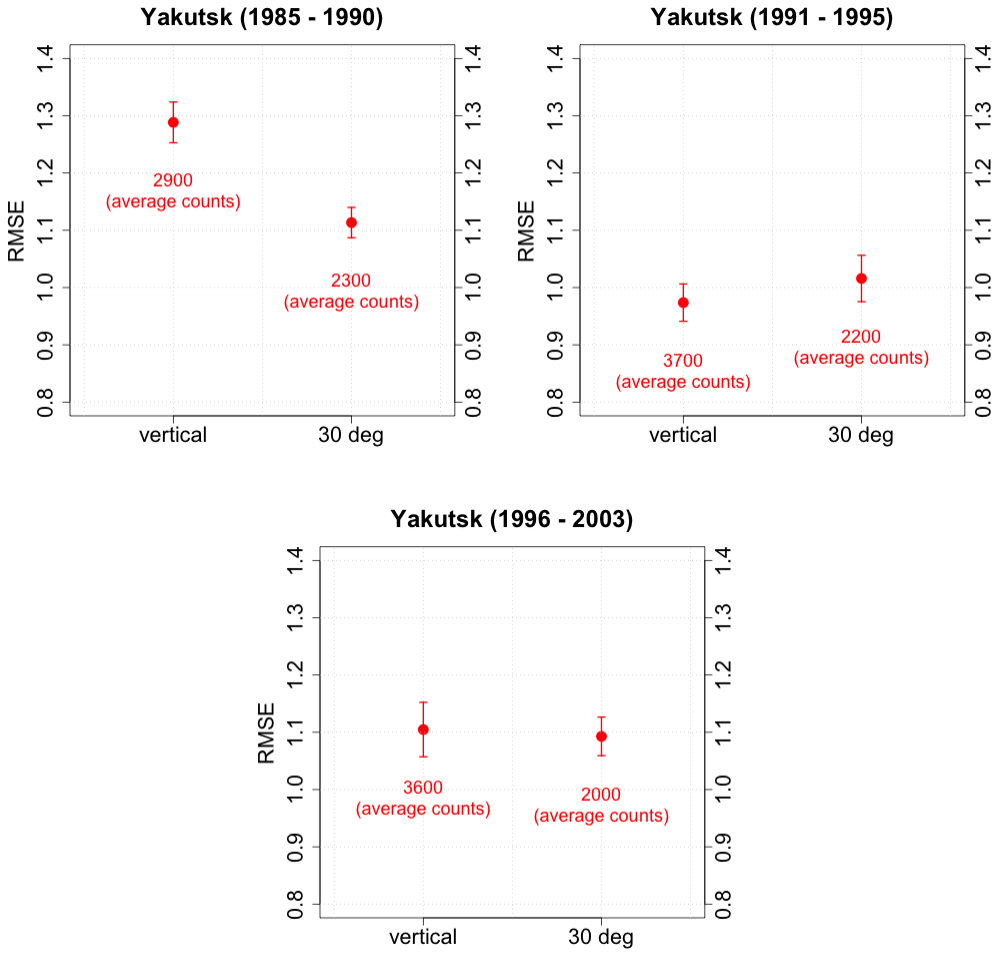

Figure (5.13) RMSE of model M2.1 calculated by 30 times 3-fold CV over six datasets of muon flux measured by Yakutsk muon telescope at ground level. The dataset was divided into three periods as indicated on the top of each figure.

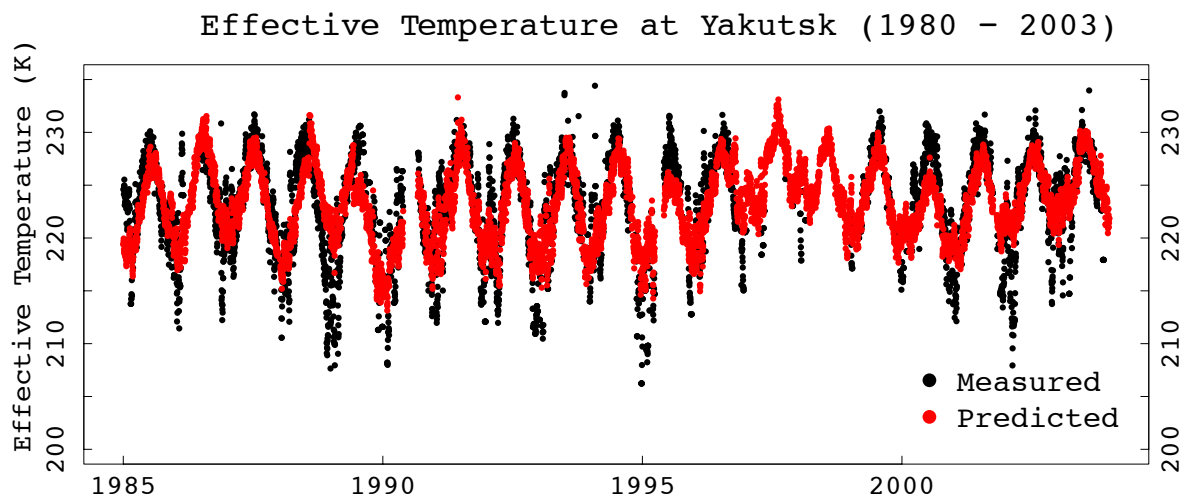

Figure (5.14) Time series plot of effective temperatures at Yakutsk from 1985 to 2003. The black dots are effective temperatures that calculated with radiosonde data and weighting functions obtained with Geant4 simulation. The red dots are effective temperatures that predicted with cosmic ray data and predictive model M2.1. The tick marks of each year on the x axis is the first day of the year.

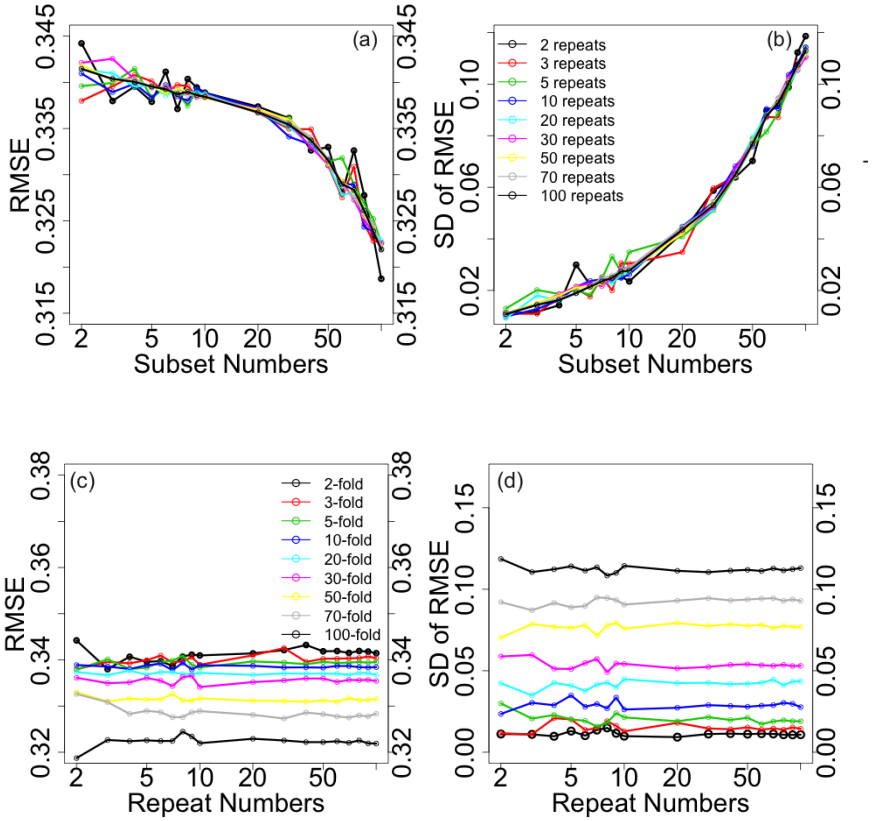

Figure (5.15) RMSE and its σ (SD of RMSE) estimated by repeated k-fold CV with varying fold numbers and repeat numbers. Testing dataset is cosmic ray muon flux coming from $17^\circ - 28^\circ$ measured by the four-paddle detector during the period 03/18/2014 - 09/30/2014. The model used in the test is $\delta T_{eff} = a \cdot \delta I_\mu + c \cdot \delta I_N + b \cdot \delta P$. (a) and (b) share the same legend, in which different color represent different n. (c) and (d) use the same legend labeling k with colors. (a) RMSE vs subset numbers (k); (b) σ of RMSE vs subset numbers (k); (c) RMSE vs repeat times of CV (n); (d) σ of RMSE vs repeat times of CV (n)

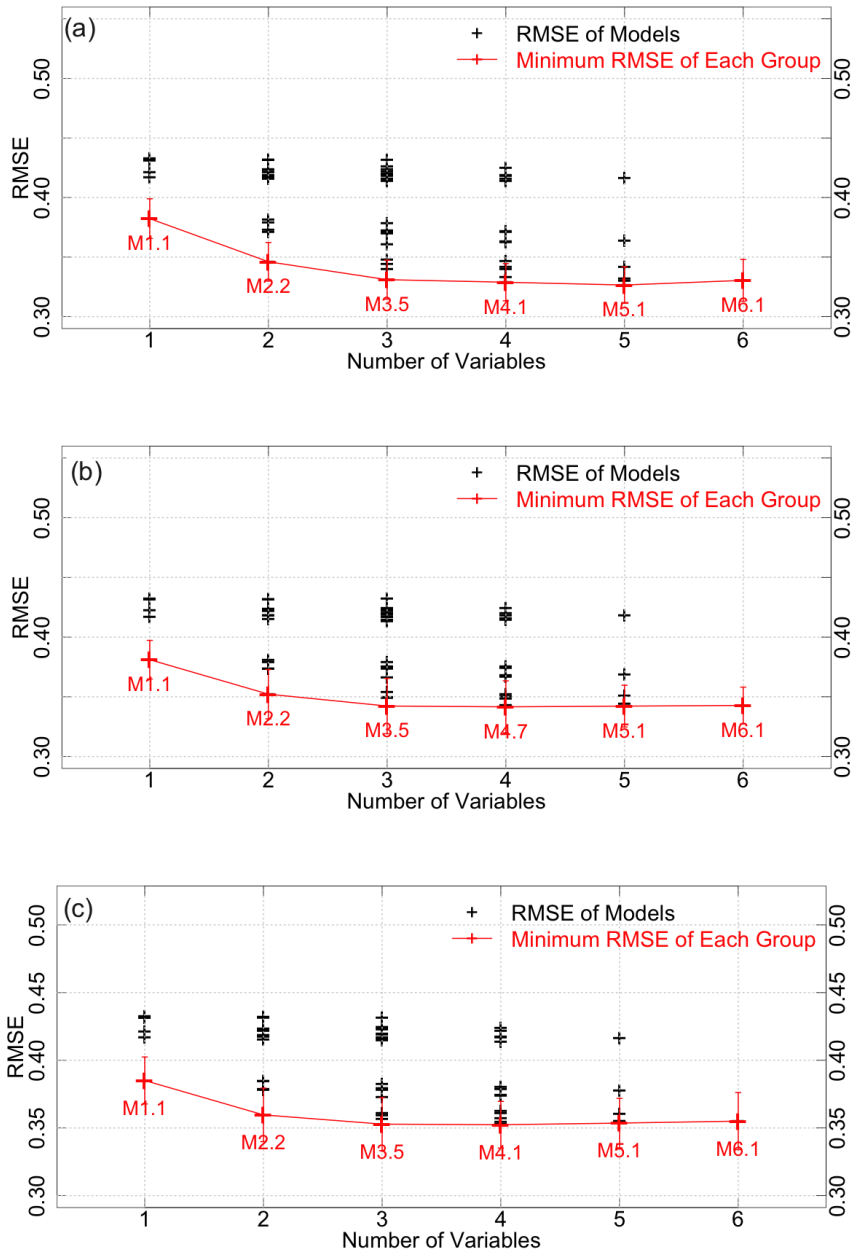

Figure (5.16) RMSE of all 63 candidate models calculated with 30 times repeated 5-fold CV. The candidates are divided into six groups according to their number of predictors. Each of the black or red plus symbol represents one candidate. The models that have the least RMSE in their group are plotted with red plus, whose σ and names (such as M1.1) are all labeled with red color. In plot (a), (b) and (c), muon flux data used in the test are those measured by the four-paddle detector with incoming angle ranges of 17° to 28°, 12° to 17° and 0° to 12° respectively. Also, all three plots used effective temperature calculated from radiosonde measurement at Peachtree City Observation, neutron flux measured at Oulu and barometric pressure from Underground Weather, LLC. The period of the datasets are from 03/18/2014 to 09/30/2014.

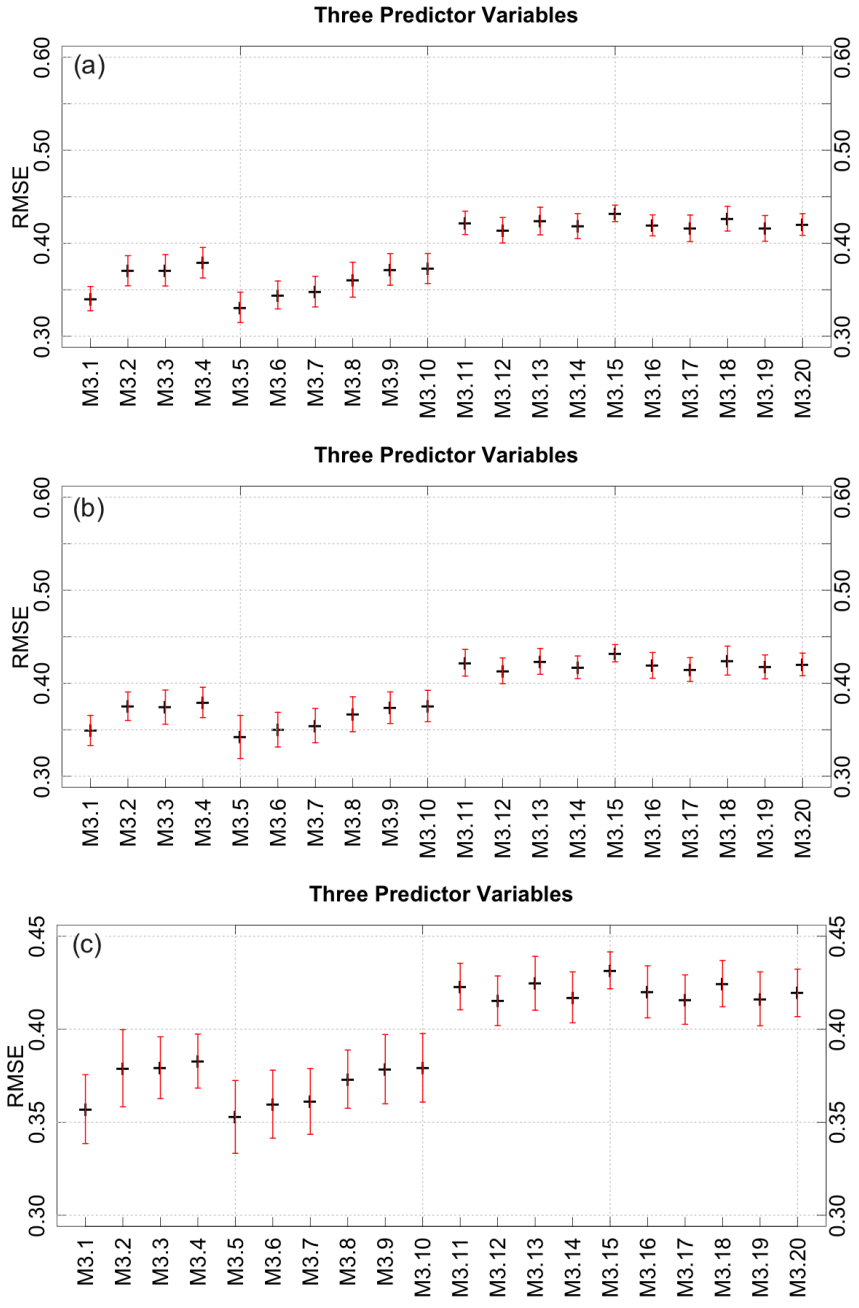

Figure (5.17) RMSE of models that have three predictors. In plot (a), (b) and (c), muon flux data used in the test are those measured by the four-paddle detector with incoming angle ranges of 17° to 28°, 12° to 17° and 0° to 12° respectively. Also, all three plots used effective temperature calculated from radiosonde measurement at Peachtree City Observation, neutron flux measured at Oulu, and barometric pressure from Underground Weather, LLC. The period of the datasets are from 03/18/2014 to 09/30/2014.

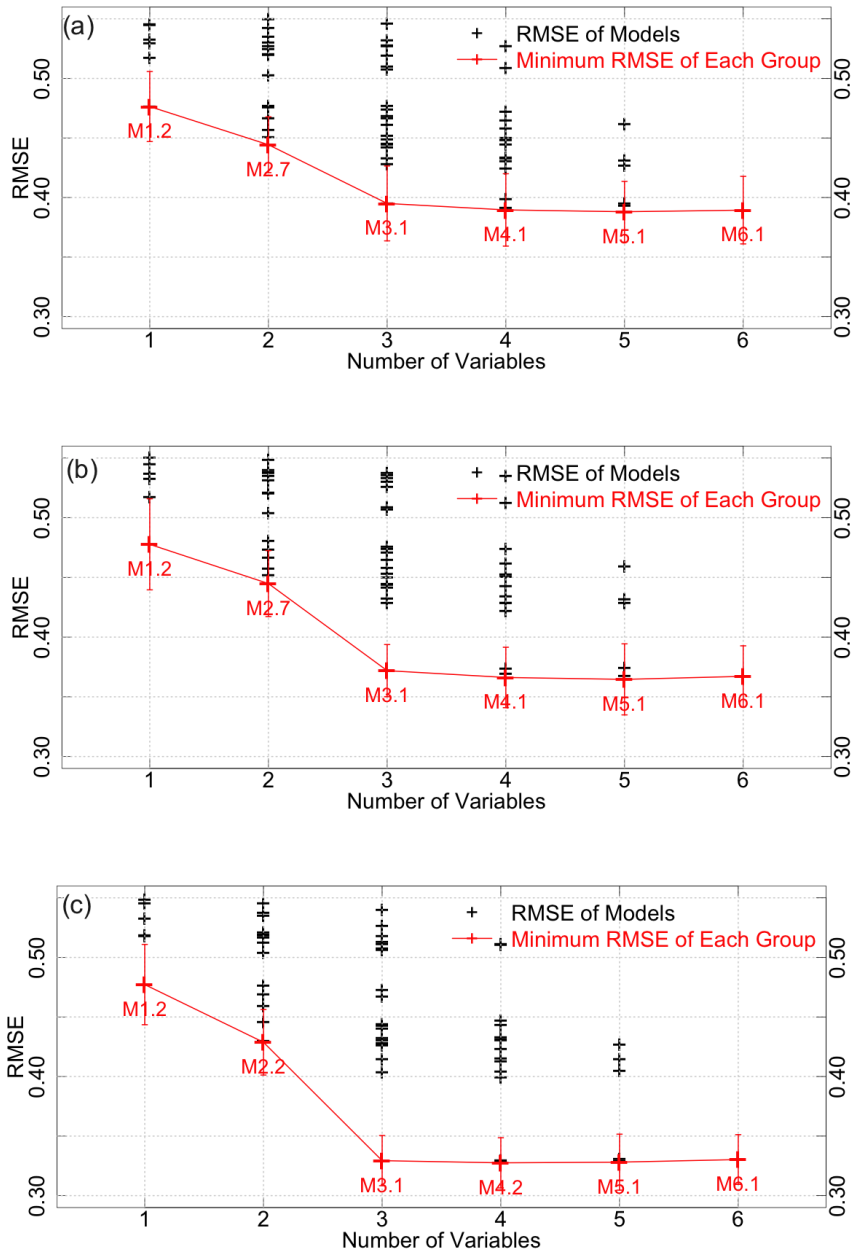

Figure (5.18) RMSE of all 63 candidate models calculated with 30 times repeated 5-fold CV. The candidates are divided into six groups according to their number of predictors. Each of the black or red plus symbol represents one candidate. The models that have the least RMSE in their group are plotted with red plus, whose σ and names (such as M1.1) are all labeled with red color. In plot (a), (b) and (c), muon flux data used in the test are those measured by the four-paddle detector with incoming angle ranges of 40° to 60°, 23° to 40° and 0° to 23° respectively. Also, all three plots used effective temperature calculated from radiosonde measurement at Peachtree City Observation, neutron flux measured at Oulu, and barometric pressure from Underground Weather, LLC. The period of the datasets are from 03/18/2015 to 09/30/2015.

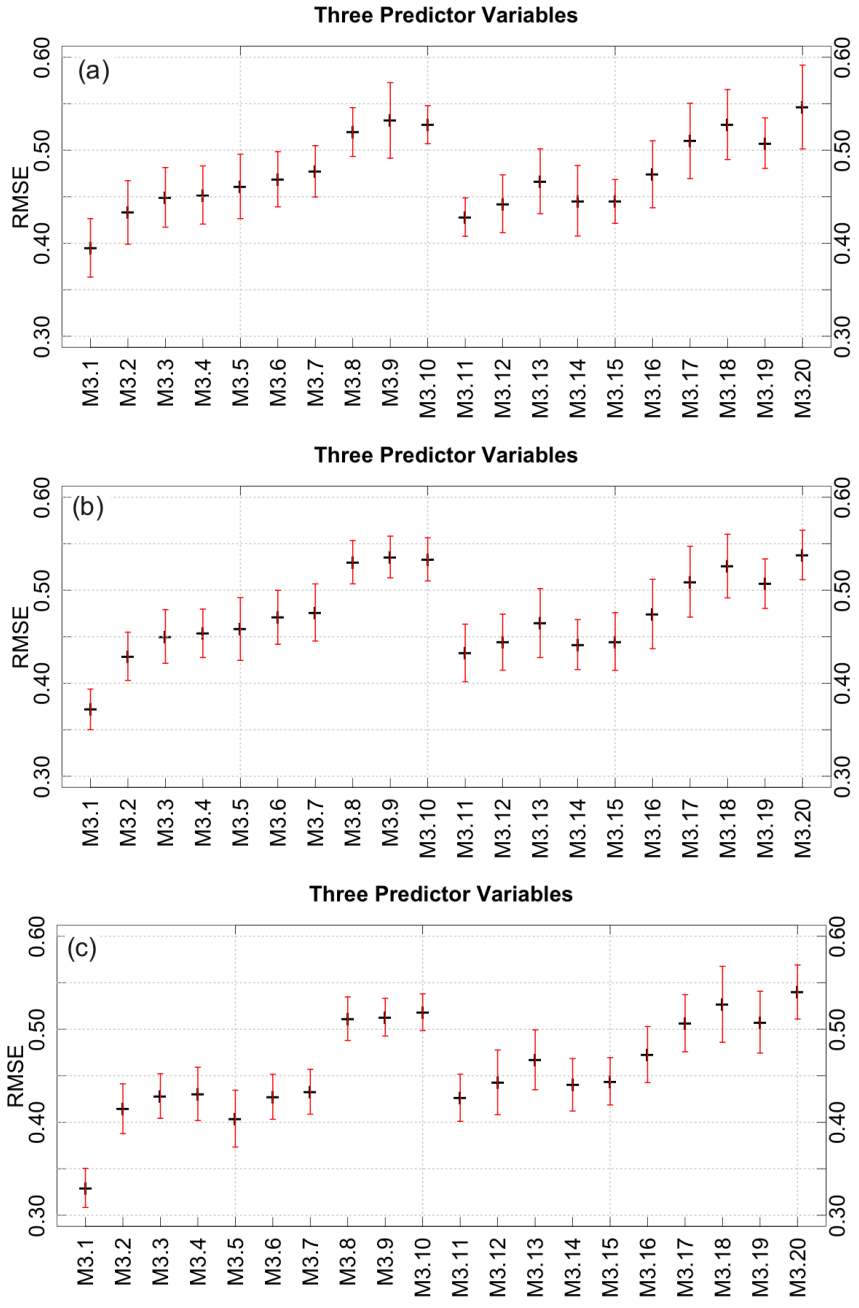

Figure (5.19) RMSE of models that have three predictors. In plot (a), (b) and (c), muon flux data used in the test are those measured by the four-paddle detector with incoming angle ranges of 40° to 60°, 23° to 40° and 0° to 23° respectively. Also, all three plots used the same data of effective temperature calculated from radiosonde measurement at Peachtree City Observation, neutron flux measured at Oulu, and barometric pressure from Underground Weather, LLC. The period of the datasets are from 03/18/2015 to 09/30/2015.

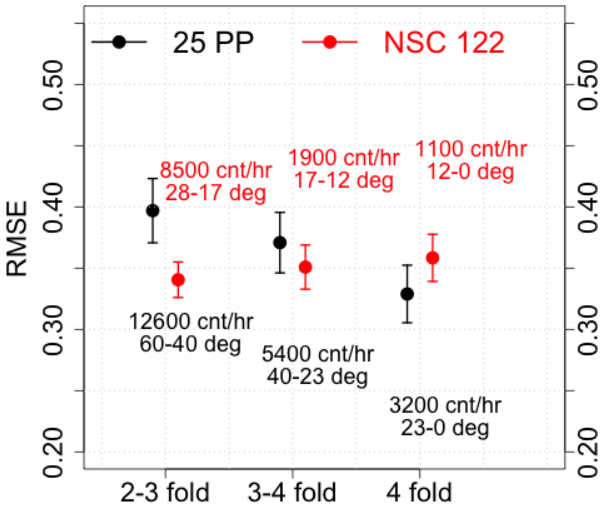

Figure (5.20) RMSE of model M3.1 calculated by 30 times 5-fold CV over six datasets of muon flux measured by four-paddle detector. Results of datasets in the period of 03/18/2014 - 09/30/2014 (measured in NSC 122) are red solid circles and period of 03/18/2015 - 09/30/2015 (measured at 25 Park Place) are black solid circles. X-axis are the ways how the muon flux variation were derived from the original four-paddle detector measurements. Such as "2-3 fold" means the muon flux was calculated by subtracting counts of 3-fold coincidence from counts of 2-fold coincidence. The average counting rate and acceptance angles are labeled with the same color above/bellow their corresponding circles.

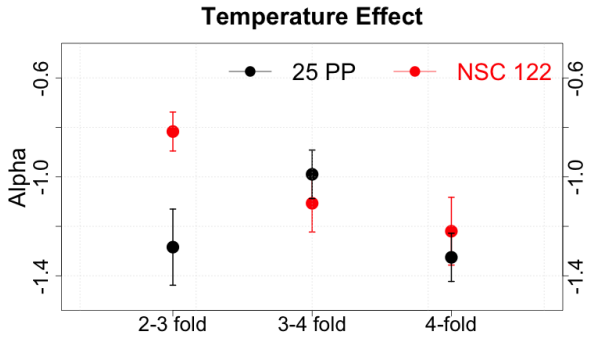

Figure (5.21) Temperature coefficients of the muon fluxes calculated with multi-variate regressions. Muon flues were measure at NSC room 122 (03/18/2014 - 09/30/2014, red circles) and 26th floor of 25 Park Place (03/18/2015 - 09/30/2015, black circles).

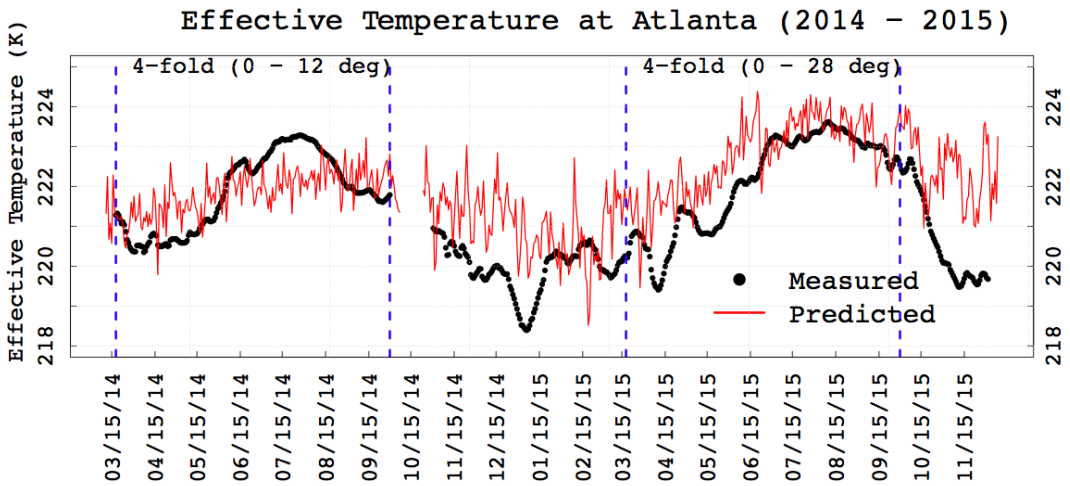

Figure (5.22) Effective temperature of Atlanta from 03/11/2014 to 12/09/2015. Black circles are effective temperatures calculated with radiosonde data and weighting functions obtained with simulation (2 weeks moving average). Red curves are effective temperature calculated with muon flux of 4-fold coincidence, neutron flux, barometric pressure and Eq. 5.2, 5.3 and 5.4. Eq. 5.2 was obtained with data from period 03/18/2014 - 09/30/2014 (period between first and second blue dashed lines), and equations 5.3 was obtained with data from period 03/18/2015 - 09/30/2015 (period between third and fourth blue dashed lines).

CHAPTER 6

CONCLUSION AND FUTURE WORK

A new method was shown in this work to calculate effective temperature using muon fluxes measured at any location on the surface of the Earth. It was proved theoretically that weighting function can be calculated with quantities N_i and N_{total}, which can be derived in simulation. In this work, the weighting functions for Atlanta and Yakutsk were derived by parametrizing their air density profiles and cut-off rigidities in the simulation program developed by the Nuclear Physics Group at GSU. Both weighting functions peak at 100 to 200 hPa region, which indicates the strong dependence of effective temperature variations on temperature changes in the tropopause and lower stratosphere.

With the data from Atlanta and Yakutsk, prediction errors of up to hundreds of general linear models were estimated with k-fold cross validation. The results shows that a simple linear model with variations of muon flux, neutron flux and barometric pressure (not necessary for pressure corrected cosmic ray data) as predictors can be used for effective temperature prediction/reconstruction with about the same predicting errors as more complicated models. In addition, the results of both datasets show that effective temperature derived with this method can be predicted using muon flux coming from less than 30° zenith angle with similar predicting errors. For muon flux with larger incoming angles, it is necessary to improve the simulation accordingly to obtain a more suitable weighting function for the effective temperature. Moreover, it was shown by the four-paddle detector's measurement that differences in the strength of the temperature effect between muon flux with different incoming angles do not cause significant predictive power changes. It was also shown that the muon flux measured in the basement can predict effective temperature as precisely as the measurements on the top floor when the counting rates are large enough.

Effective temperature variations at both Atlanta and Yakutsk were reconstructed with

the statistical modeling method. The reconstructed effective temperature can reproduce both long term (years) and medium term (weeks) variations of the ones calculated with radiosonde measurements. In order to test the predicting performance of the models in short term (days), more direct measurements of the atmospheric temperature profile are necessary. The result of this work indicates the possibility of monitoring the effective temperature with simultaneous measurement of cosmic ray muon and neutron flux. It also suggests one possible usage of the early datasets from the worldwide cosmic ray network, which is helping with reconstructing the effective temperatures finding the SSW events in the historical archive. At last, this study suggests that cosmic ray detectors can complement the current detection of SSWs and stratosphere observation techniques and help improve the weather forecasts.

REFERENCES

[1] P. J. Kocin and L. W. Uccellini, "A snowfall impact scale derived from northeast storm snowfall distributions," *AMERICAN METEOROLOGICAL SOCIETY*, pp. 177 – 194, 2004.

[2] Wikipedia, "January 2016 united states blizzard," 2016. [Online; last modified on 18-June-2016].

[3] Wikipedia, "January 2016 east asia cold wave," 2016. [Online; last modified on 18-June-2016].

[4] NOAA, "The northeast snowfall impact scale (nesis)." `https://www.ncdc.noaa.gov/snow-and-ice/rsi/nesis`, 2016.

[5] M. P. Baldwin and T. J. Dunkerton, "Stratospheric harbingers of anomalous weather regimes," *Science*, vol. 294, no. 5542, pp. 581–584, 2001.

[6] NOAA, "Arctic oscillation (ao)." `https://www.ncdc.noaa.gov/teleconnections/ao/`, 2016.

[7] D. G. Andrews, J. R. Holton, and C. B. Leovy, "Middle atmosphere dynamics," *Academic Press*, p. 489 pp, 1987.

[8] V. Limpasuvan, D. W. J. Thompson, and D. L. Hartmann, "The life cycle of the northern hemisphere sudden stratospheric warmings," *Journal of Climate*, vol. 17, pp. 2584–2596, 2004.

[9] L. P. Goncharenko, A. J. Coster, J. L. Chau, and C. E. Valladares, "Impact of sudden stratospheric warmings on equatorial ionization anomaly," *Journal of Geophysical Research*, vol. 115, p. A00G07, 2010.

[10] D. Pancheva and P. Mukhtarov, "Strong evidence for the tidal control on the longitudinal structure of the ionospheric f-region," *Geophsical Research Letters*, vol. 37, p. L14105, 2010.

[11] P. Adamson, C. Andreopoulos, *et al.*, "Observation of muon intensity variations by season with the minos far detector," *Phys. Rev. D*, vol. 81, p. 012001, 2010.

[12] S. Tilav, P. Desiati, T. Kuwabara, D. Rocco, F. Rothmaier, M. Simmons, and H. Wissing, "Atmospheric variations as observed by icecube," *Proceedings of the 31st ICRC*, vol. 5, p. 11111, 2009.

[13] P. Carlson and A. A. Watson, "Erich regener and the ionisation maximum of the atmosphere," *Hist. Geo Space Sci.*, vol. 5, pp. 175–182, 2014.

[14] V. F. Hess *Phys. Zeits.*, vol. 12, p. 998, 1911.

[15] W. Baade and F. Zwicky, "Cosmic rays from super-novae," *Proc. National Academy of Sci.*, vol. 20, pp. 259–263, 1934.

[16] K. Sakurai, *Physics of Solar Cosmic Rays*. University of Tokyo Press, 1974.

[17] J. Beringer and J. F. A. et al., "Cosmic rays," *Particle Data Group, Phys. Rev. D*, vol. 86, p. 010001, 2012.

[18] R. Bellotti and et al., "Measurement of the negative muon spectrum between 0.3 and 40 gev/c in the atmosphere," *Phys. Rev.*, vol. D53, p. 35, 1996.

[19] R. Bellotti and et al., "Balloon measurements of cosmic ray muon spectra in the atmosphere along with those of primary protons and helium nuclei over mid latitude," *Phys. Rev.*, vol. D60, p. 052002, 1999.

[20] M. Boezio and et al., "Measurement of the flux of atmospheric muons with the caprice94 apparatus," *Phys. Rev.*, vol. D62, p. 032007, 2000.

[21] M. Boezio and et al., "Energy spectra of atmospheric muons measured with the caprice98 balloon experiment," *Phys. Rev.*, vol. D67, p. 072003, 2003.

[22] S. Coutu and et al., "Energy spectra, altitude profiles and charge ratios of atmospheric muons," *Phys. Rev.*, vol. D62, p. 032001, 2000.

[23] S. H. et al., "Measurements of primary and atmospheric cosmic - ray spectra with the bess-tev spectrometer," *Phys. Lett.*, vol. B594, p. 35, 2004.

[24] W. O. Davis, "Energy and density distribution of cosmic-ray neutrons," *Phys. Rev.*, vol. 80, pp. 150–154, 1950.

[25] W. N. Hess, H. Patterson, and R. Wallace, "Cosmic ray neutron energy spectrum," *Phys. Rev.*, vol. 2, p. 116, 1959.

[26] M. S. Gordon, P. Goldhagen, K. P. Rodbell, T. H. Zabel, H. H. K. Tang, J. M. Clem, and P. Bailey, "Measurement of the flux and energy spectrum of cosmic-ray induced neutrons on the ground," *Nuclear Science*, vol. 51, no. 2, pp. 3527–3434, 2004.

[27] P. Goldhagen, J. M. Clem, and J. W. Wilson, "The energy spectrum of cosmic-ray induced neutrons measured on an airplane over a wide range of altitude and latitude.," *Rad. Prot. Dos.*, vol. 110, no. 1-4, pp. 387–392, 2004.

[28] M. R. Moser, J. M. Ryan, L. Desorgher, and E. O. Fluckiger, "Atmospheric neutron measurements in the 10-170 mev range," *Proceedings of the 29th International Cosmic Ray Conference*, vol. 2, pp. 421–424, 2005.

[29] Y. Miyazaki and M. Wada, "Simulation of cosmic ray variation due to temperature effect," *Acta Phys. Acad. Sci. Hung.*, vol. 29(2), pp. 591–595, 1970.

[30] M. Bertaina, L. Briatore, A. Longhetto, G. Navarra, and E.-T. Coll., "The atmospheric muon flux in correlation with temperature variations in the low stratosphere (50-200mb)," *Proceedings of the 30th ICRC*, vol. SH.3.6, pp. 713 – 716, 2008.

[31] V. Borog, O. V. Belonosova, A. A. Davydov, G. M. Kruchenitskii, S. P. Perov, and V. G. Yanke, "Study of atmospheric temperature at different altitudes using muon angular distribution at sea level," *Proceedings of the 29th ICRC*, vol. 2, pp. 381 – 384, 2005.

[32] L. Dorman, *Cosmic Rays in the Earth's Atmosphere and Underground*. Kluwer Academic Publishers, 2004.

[33] L. Dorman, *Cosmic Ray Variations*. State Publishing House, Moscow, 1957.

[34] P. Barrett *et al.*, "Interpretation of cosmic-ray measurements far underground," *Reviews of Modern Physics*, vol. 24, pp. 133–178, 1952.

[35] M. Ambrosio *et al.*, "Seasonal variations in the underground muon intensity as seen by macro," *Astroparticle Physics*, vol. 7, pp. 109–124, 1997.

[36] R. Abbasi, Y. Abdou, T. Abu-Zayyad, and et al., "Measurement of the anisotropy of cosmic-ray arrival directions with icecube," *The Astrophysical Journal Letters*, vol. 718, no. 2, pp. L194–L198, 2010.

[37] B.-H. Mevik and R. Wehrens, "Introduction to the pls package." `https://cran.r-project.org/web/packages/pls/vignettes/pls-manual.pdf`, 2015.

[38] W. N. Venables, D. M. Smith, and the R Core Team, "An introduction to r." `https://cran.r-project.org/doc/manuals/R-intro.pdf`, 2016.

[39] M. Kuhn, "A short introduction to the caret package." `https://cran.r-project.org/web/packages/caret/vignettes/caret.pdf`, 2015.