12-2010

Statistical Analysis of the USU Lidar Data Set with Reference to Mesospheric Solar Response and Cooling Rate Calculation, with Analysis of Statistical Issues Affecting the Regression Coefficients

Troy Alden Wynn
Utah State University

STATISTICAL ANALYSIS OF THE USU LIDAR DATA SET WITH REFERENCE TO MESOSPHERIC SOLAR RESPONSE AND COOLING RATE CALCULATION, WITH ANALYSIS OF STATISTICAL ISSUES AFFECTING THE REGRESSION COEFFICIENTS

by

Troy Alden Wynn

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Physics

Approved:

_____	_____
Dr. Vincent B. Wickwar Major Professor	Dr. James T. Wheeler Committee Member
_____	_____
Dr. Eric D. Held Committee Member	Dr. Robert W. Schunk Committee Member
_____	_____
Dr. Daniel C. Coster Committee Member	Dr. Byron R. Burnham Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2010

ABSTRACT

Statistical Analysis of the USU Lidar Data Set with Reference to Mesospheric Solar Response and Cooling Rate Calculation, with Analysis of Statistical Issues Affecting the Regression Coefficients

by

Troy A. Wynn, Doctor of Philosophy

Utah State University, 2010

Major Professor: Dr. Vincent B. Wickwar
Department: Physics

Though the least squares technique has many advantages, its possible limitations as applied in the atmospheric sciences have not yet been fully explored in the literature. The assumption that the atmosphere responds either in phase or out of phase to the solar input is ubiquitous. However, our analysis found this assumption to be incorrect. If not properly addressed, the possible consequences are bias in the linear trend coefficient and attenuation of the solar response coefficient.

Using USU Rayleigh lidar temperature data, we found a significant phase offset to the solar input in the temperatures that varies ±5 years depending on altitude. In addition to introducing a phase offset into the linear regression model, we argue that separating what we identify as the solar-noise is to be preferred because (1) the solar-noise can contain important physical information, (2) its omission could lead to spurious conclusions about the significance of the solar-proxy coefficient, and (3) its omission could also bias the

solar proxy coefficient.

We also argue that the Mt. Pinatubo eruption caused a positive temperature perturbation in our early mesopause temperatures, exerting leverage on the linear trend coefficient. In the upper mesosphere, we found a linear cooling trend of greater than −1.5 K/year, which is possibly exaggerated because of leverage from the earlier temperatures and/or collinearity. In the middle mesosphere we found a cooling trend of −1 K/year to near zero.

We use the autocorrelation coefficient of the model residuals as a physical parameter. The autocorrelation can provide information about how strongly current temperatures are affected by prior temperatures or how quickly a physical process is occurring.

The amplitudes and phases of the annual oscillation in our data compare favorably with those from the OHP and CEL French lidars, as well has the HALOE satellite instrument measurements. The semiannual climatology from the USU temperatures is similar to that from the HALOE temperatures. We also found that our semiannual and annual amplitudes and phases compare favorably with those from the HALOE, OHP, and CPC data.

The computer code used to generate the author's figures included in this dissertation is given in Appendix I.

(280 pages)

ACKNOWLEDGMENTS

I would like to thank the Rocky Mountain Space Grant Association for supporting me for several years while I conducted this research and the USU Physics Department for providing various assistantships and teaching opportunities during my time here at USU.

I would also like to thank Dr. Vincent Wickwar and Dr. Joshua Herron for their assistance and for making the data available to me, as well as Bill Randel at NCAR for providing the CPC and ERA temperature data.

I give special thanks to my parents who have been ever supportive and patient, and to all my friends and colleagues for their help and support.

Troy A. Wynn

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1. Carbon Dioxide and Climate

Over the past several decades there has been increasing interest in the effects of green house gases on the Earth's climate and temperature. There is particular concern that increased carbon dioxide levels could drastically alter the terrestrial atmosphere's mean temperature structure. Because carbon dioxide is not chemically active, it is expected to persist in the Earth's atmosphere for decades to centuries. CO_2 measurements from Mauna Loa Observatory, Hawaii show a 22% increase in CO_2 levels from 1959 to 2008 [*Tans*, 2008]. Ice core measurements indicate CO_2 levels have increased 33% since 1850 [*Ledley et al.*, 1999], with the bulk of the increase having occurred in the last fifty years. Climatological simulations predict that in response to increased CO_2 levels the middle atmosphere will cool and the troposphere will warm [*Roble and Dickinson*, 1989; *Rind et al.*, 1990; *Khosravi et al.*, 2002; *Gruzdev and Brasseur*, 2005; *Fomichev et al.*, 2007]. Consequently atmospheric scientists are looking for troposphere warming and middle atmosphere cooling as evidence of what is colloquially referred to as global warming, a term concomitant with discussions about global climate changes.

Climate models suggest the effects of global climate change will not be evenly distributed, geographically or seasonally. Some regions might experience increased precipitation while others might experience less; some regions might experience warmer winters and others could experience little or no change during summer; some regions might become warmer and wetter while others could become cooler and dryer. The net

effect, however, is elevated global temperatures in the troposphere. Some of the consequences are increased evaporation, changes in freshwater availability and ground water replenishment, changes in ocean temperatures that could affect fish and other aquatic species, and a rise in sea levels that could affect costal and river basin areas [*Parry et al.*, 2007].

Naturally the possibility of damage to the ecosystem and its consequent human cost coupled with the possibility of increased government intervention makes the issue serious and politically controversial. This has fueled skepticism about the seriousness and reality of global warming. I once tried to explain global warming to a friend. He countered my arguments with a news article from the 1970s that mentioned a predicted global cooling. It turns out that in 1975 *Newsweek* published an article called The Cooling World and in 1974 *Time* published an article called Another Ice Age? (*Newsweek*, 28 April 1975; *Time*, 24 June 1974).

There is a great deal of climate skepticism among the public. And, unfortunately, the recent e-mail scandal that emerged in November 2009 only encouraged these doubts. Hundreds of hacked e-mails at the Climatic Research Unit (CRU) at the University of East Anglia (UEA) in Norwich, UK were released to the public, some of which contained comments that, at face value, suggested members of the CRU had attempted to keep publications not in harmony with the global warming thesis out of the next International Panel on Climate Change (IPCC) report. Dubbed Climategate, this scandal naturally caused a great deal of embarrassment for climate scientists and fueled skepticism about the verity of the global warming thesis. A December 2009 Rasmussen Report indicates that 52% of Americans believe there is significant disagreement about global warming in

the scientific community [*Rasmussen Reports*, 2009]. According to a March 2010 Gallup poll 52% of Americans believe that most scientists believe global warming is occurring, down from 65% just two years ago. This report also indicates that 50% of Americans believe increases in global temperatures are due to human activities, down from 61% in 2003 [*Newport*, 2010]. This emphasizes the importance of a thorough statistical analysis of all available data. (See also *BBC* [2010], *Leiserowitz et al.*[2010], and *Spence et al.* [2010] for more polling results.)

Though an independent panel set up by the UEA in consultation with the Royal Society concluded that the CRU research was honest and scientifically justified, they did express regret that "so few professional statisticians have been involved in this work because it is fundamentally statistical" [*Oxburgh et al.*, 2010].

It should be added that the IPCC report is not without its flaws. After reviewing some of the criticisms of the latest IPCC report, an editorial from *The Economist* online concludes, "A suspicion thus gains ground that the way in which the IPCC synthesises, generalises and checks its findings may systematically favour adverse outcomes in a way that goes beyond just serving the needs of policy makers" (5 July 2010).

The controversy surrounding the issues of global warming is intense. This only strengthens the need for the greatest possible precision in the statistical analysis of the available data.

2. Statistical Issues

Several statistical issues affecting the least squares regression technique have already been addressed in the literature. The effects of autocorrelation have been discussed by

Frederick [1985], *Tiao et al.* [1990], *Krzyścin* [1997], and *Weatherhead et al.* [1998]. *Weatherhead et al.* also reviewed the effects of interventions such as volcanic eruptions, as well as instrumentation adjustments that might affect linear trend estimates. *Kerzenmacher et al.* [2006] addresses the problems of bias that accompany model selection. *Mäder et al.* [2007] mentions that collinearity could negatively impact a stepwise variable selection procedure. *Kerzenmacher et al.* also did Monte Carlo simulations to calculate the errors that occur from model underspecification, though it is likely he used simulations because he did not distinguish between coefficient bias and coefficient attenuation, as the bias calculation is rather simple.

One subject not covered in the literature is the problem of linear-trend bias that arises from including a fixed-phase solar proxy in a temperature model under circumstances where a variable phase proxy is more appropriate. By fixed proxy we mean a solar proxy that does not change its phase with altitude; it is fixed. By variable phase we mean a proxy that contains a phase that can vary continuously with altitude, such as $\sin[\omega t + \varphi(z)]$.

Model misspecification occurs when a model omits important variables and/or includes unimportant ones. If the model is under specified, that is, if important variables are omitted and no unimportant variables are included, then the model coefficients will be biased. If the model is over specified, that is, if all essential model variables are included along with some unimportant ones, then the correctly specified variables are unbiased and the coefficients of the unimportant variables are zero. If the model is cross specified, that is, if some correct variables are excluded and unimportant variables are included, then the correctly specified variables will be biased and the incorrectly specified variables

are attenuated versions of the omitted, true-model coefficients.

Some researchers will deliberately over specify a model to reduce the danger of coefficient bias. However, this introduces the danger of collinearity, or illconditioned data. If two variables have near linear dependence, this can lead to coefficient correlation: If one coefficient is high, then the other will be low or high depending on the sign of the correlation. Of interest here is the correlation that exists between the solar proxy and linear trend coefficients. (These two variables are found in nearly all middle atmosphere temperature models.) Depending on the phase of the solar cycle relative to the data, a significant correlation problem can arise, and it is possible to identify cases that do not suffer from this difficulty. Thus greater confidence may be placed in those results. The main criteria for identification of collinearity problems depend principally on the phase of the solar cycle relative to the data and the length of the data set.

As already mentioned, the magnitude of the atmospheric solar response is often extracted using least squares regression techniques. To approximate the solar input a solar proxy is included in the model—typically the F10.7 radio emission; Mg II core-to-wing ratio; or sunspot number. The resulting regression coefficient may be positive or negative. A positive coefficient indicates an in-phase atmospheric response. A negative coefficient indicates an inverted response. Positive and negative solar responses have been found in both stratospheric and mesospheric temperatures. One problem with using a fixed-phase proxy is that, in doing so, an implicit assumption is made: The atmosphere is responding either in phase or out of phase to the solar input. And besides failing to capture a phase offset it creates another problem. A positive coefficient at one altitude and a negative coefficient at another means the coefficient value must go through zero at

some point, potentially masking changes in the amplitude of the atmospheric solar response. Moreover, if a fixed-phase proxy model is applied when a variable-phase model is more appropriate, the fixed-phase proxy assumption amounts to model misspecification which can introduce significant bias to the linear trend coefficient and attenuate the solar proxy coefficient.

Because of the ongoing effort invested in detecting changes in the Earth's climate, knowing when models do and do not suffer from bias, attenuation, or collinearity can be of value to those in the scientific community trying to detect cooling trends and model the Earth's complex chemical, dynamical, and radiative processes.

3. Review of Literature

Because of the complexity of the Earth's atmosphere, climate models are needed to assess the long-term effects of increased carbon dioxide levels. A CO_2 doubling simulation conducted by *Fomichev et al.* [2007] using the Canadian Middle Atmosphere Model (CMAM) compared two different CO_2 levels. The first was at a baseline level of 338 ppmv, corresponding approximately to 1986 levels. In the second the CO_2 level was doubled from the baseline value. Significant differences between the two model runs were then noted. Their simulations predict that for January and July at 40ºN (approximately the same latitude as the USU lidar) the mesosphere should cool from 4 to 10 K. If the current rate of CO_2 increase continues then a CO_2 doubling from its 1986 level should occur around the year 2100, putting the cooling rate between −0.35 and −0.9 K/decade. *Gruzdev and Brasseur* [2005] did a simulation using the SOCRATES model where they investigated mesosphere thermal and chemical responses to changes in

greenhouse gas concentrations that occurred over the past 50 years. They predicted a 3 to 7 K mesopause cooling and a 4 to 6 K cooling in the middle mesosphere. This amounts to a cooling rate of −0.6 to −1.4 K/decade for the mesopause and nearly the same rate for the middle mesosphere. Using the Spectral Mesosphere/Lower Thermosphere model, *Akmaev and Fomichev* [2000] investigated the thermal response of the atmosphere to changes in CO_2 levels that occurred from 1955 (313 ppm) through 1995 (360 ppm). They found a mesosphere cooling of approximately 3 K, which amounts to a cooling rate of −0.75 K/decade. *Rind et al.* [1990] investigated the impact of a doubled CO_2 climate on the thermal atmosphere using the GISS Global Climate Middle Atmosphere Model (GCMAM). Starting with a CO_2 baseline of 315 ppm (corresponding to approximately 1959 levels) then increasing it to 630 ppm (corresponding to approximately 2090 by my estimate) they found a mesosphere cooling of 4 to 11 K, which corresponds roughly to −0.3 to −0.8 K/decade. In an updated paper *Rind et al.* [1998] reproduced some of their earlier CO_2 doubling simulations using the latest version of the GCMAM model and found a mesosphere cooling between 8 and 12 K at 40ºN which, based on the CO_2 levels used in their simulation, amounts to −0.06 to −0.09 K/year. When all are taken together these simulations indicate a middle atmosphere cooling rate in the neighborhood of −0.3 to −1.4 K/decade.

Middle atmosphere cooling has been found by many researchers. However, the magnitudes are, needless to say, all over the place. Table 4 from *Beig et al.* [2003] lists mesosphere (50-79 km) linear trends from various data collection sites. The cooling rates range from −10 to −0.24 K/decade. Table 5 from the same paper lists several mesopause (80-100 km) temperature trends ranging from −10.5 to +5 K/decade. (Figure 1.1 for

histograms.) Published radiosonde temperatures also indicate significant variability. Table 3 from *Ramaswamy et al.* [2001] lists several lower stratosphere cooling trends ranging from −4.5 to +4.1 K/decade, with a median value of −0.4 K/decade. In both *Ramaswamy et al.* and *Beig et al.* the majority of the temperature trends are negative and consistent in sign with the predicted middle atmosphere cooling.

In addition to increased levels of CO_2, the variation of solar input during the approximate 11-year solar cycle also has a direct impact on middle atmosphere temperature structure, principally through exothermic chemical reactions involving ozone and atomic oxygen, which are strong absorbers of UV radiation. And, as with cooling rates, calculating the atmospheric response to variations in solar irradiance requires computer models to handle the complex physical and chemical processes.

One simulation by *Huang and Brasseur* [1993] shows a temperature variation of 1 to

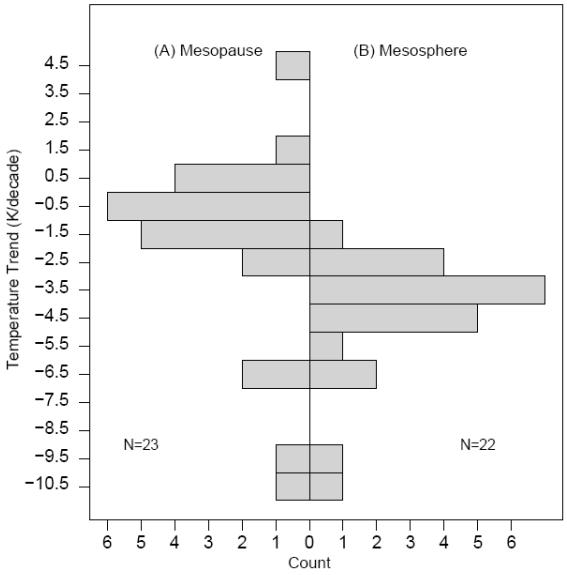


Figure 1.1. Histograms of reported temperature trends tabulated in *Beig et al.* [2003]. Histogram (a) is for temperature trends near the mesopause (80-100 km) from Table 5. Histogram (b) is for mesosphere trends (50-79 km) from Table 4. For cases where the temperatures were reported as −1.4 to −2.1 K/decade (for example) both the upper and lower limits were treated as data points.

10 K (depending on the altitude) for the mesosphere at 40ºN for the September model run. Using the NCAR Whole Atmosphere Community Climate Model, version 3 (WACCM3), *Marsh et al.* [2007] found a temperature change from solar maximum to solar minimum from 0.50 to 3 K between 45 and 90 km at 40ºN. Simulations run using the SOCRATES model by *Khosravi et al.* [2002] found a max-min mesosphere (50-90 km) solar response at 40ºN between 0.5 and 5 K.

Measurements indicate a more varied response. *Chanin* [2006] lists temperature responses from six sites ranging in location from 37ºN to 54ºN. The following examples are from 30 km. The solar response coefficient at Primrose Lake, Canada has a value of −8 K, the negative sign indicating that the atmospheric solar response is 180º out of phase with the solar input; at Riori, Japan it is statistically insignificant; at Volgograd, Russia it is −2 K but also statistically insignificant; at Wallops Island, Virginia it is approximately 2.5 K and statistically insignificant; at Shemya, Alaska it is negative and statistically insignificant; the OHP lidar (44ºN) in France shows 0 K atmospheric solar response. The altitudes where the solar response passes through zero are 50 km at OHP and 42 km at Shemya; it does not go through zero at Wallops Island ; at Volgograd it occurs at 24 and 50 km; at Ryori it stays near zero between 24 and 50 km; at Primrose Lake it is zero from 50 to 60 km.

Each of these groups used a fixed-phase proxy in their regression analysis, which opens up the issues of model specification and collinearity. Are these variations due to actual amplitude differences? Or are they due to phase differences with the solar input that differ geographically? Is the atmospheric solar response zero or is it 90º out of phase with the solar input? These questions are difficult if not impossible to address if a fixed-

phase proxy is used in the least squares analysis to determine the atmospheric solar response.

4. Research Objectives

Standard linear regression models usually include a solar proxy to approximate the solar UV input. We have found that if the atmosphere is not responding with a phase difference of 0 or π radians, then the linear trend and solar proxy coefficients are possibly, respectively, bias and attenuated. We derive and discuss equations estimating the bias and attenuation. These equations also allow us to identify conditions that might be immune from this problem.

Along with the assumption that the middle atmosphere responds in phase or out of phase to the solar input is the assumption that the atmosphere responds in phase to short-term solar variations, or solar noise. We argue that because large-scale solar responses are likely to be driven by dynamics and shorter scale variations are more likely to represent photochemistry and radiative transfer the 11-year solar variation and the solar-noise variations are best modeled as separate terms. Additionally, if these two variations are not separated, there exists a risk of a false positive on the significance of the solar proxy coefficient. A climatology of the solar-noise is also included.

Equations are derived that aid in understanding the collinearity problem and allow us to identify cases that are severe, as well as cases that do not suffer from coefficient correlation between the linear term and the solar proxy. We also report our findings on the amplitudes and phases of the annual oscillation (AO) and the semiannual oscillation (SAO), the linear trend coefficient, and the amplitudes and phases of the atmospheric

solar response. These sections are mostly comparative. An SAO climatology from our temperatures is also discussed and compared to an equivalent climatology from the Halogen Occultation Experiment (HALOE) instrument on the UARS satellite.

Tools were developed for estimating the autocorrelation coefficient that allowed us to estimate the residual autocorrelation and create an autocorrelation climatology, which provides additional insight into the seasonal variations of the residual standard deviation. We also investigate the possible effect the Mt. Pinatubo eruption may have had on the linear trend estimate.

5. Summary of Chapters

Chapter 2 contains a discussion of the collinearity issues, how they come about, and how they can affect regression models that contain a linear trend and a fixed-phase solar proxy. Equations for describing the collinearity are derived and discussed. Chapter 3 contains a derivation of the bias and attenuation of the linear coefficient and the atmospheric solar response coefficient, respectively. We briefly look at the amplitudes and phases of the atmospheric solar response and also some fixed-phase proxy results. We discuss cases where the amplitude and phase of the solar proxy will not bias the linear trend or attenuate the solar proxy coefficient. We argue that separating the solar-noise from the proxy is to be preferred and perform some statistical tests to support this claim. In Chapter 4 we discuss more fully the linear trend term and how collinearity and bias may be affecting the linear trend estimate. The eruption of Mt. Pinatubo is considered as a possible influence on our early temperatures and evidence is given to support this thesis. We also take a closer look at the amplitude and phase of the

atmospheric solar response and compare our results to what others have found. This includes fixed-proxy results. We also compare SAO climatologies from the USU and HALOE temperatures and look at the amplitudes and phases of the annual and semiannual oscillations. Chapter 5 contains an examination of the residuals from the full OLS (ordinary least squares) regression model used in this dissertation. This includes a residual standard deviation climatology. We compare our climatology to others from the literature. We look at the autocorrelation and discuss a climatology of the autocorrelation coefficient and use it to better understand the seasonal variation in the residual standard deviation. We develop and examine a climatology of the solar-noise response and use it to make inferences about the seasonal dependence of the atmospheric response to the solar-noise. Chapter 6 contains a summary and statement of future work. Additional information is included in the appendices.

6. Global Warming Basics

6.1. Introduction

Theories positing a connection between elevated atmospheric CO_2 levels and increased global temperatures have existed for over 100 years [*Callendar*, 1938; *Held and Soden*, 2000]. Though in the past there has been some debate as to whether or not industrialization would produce global warming or global cooling, over the past several decades the literature has increasingly favored the global warming thesis, which states that significant increases in the amount of atmospheric CO_2 results in elevated global temperatures in the troposphere.

In addition to CO_2, other important greenhouse gases are ozone, water vapor,

methane, and nitrous oxide. The dominant radiatively active gases are water vapor, carbon dioxide, and ozone, followed by methane and nitrous oxide. Several authors have emphasized that water vapor is the most important greenhouse gas [*Held and Soden*, 2000; *Soden*, 2005]. By this they mean that water vapor dominates troposphere radiative heating and cooling mechanisms. When the atmosphere is divided into spheres (troposphere, stratosphere, mesosphere), each region has its own unique combination of dominant chemical, radiative, and dynamical processes. (For atmosphere regions see Figure 1.2.) The warming is confined to the lower atmosphere (the troposphere), which contains nearly all the atmospheric water vapor. Carbon dioxide and ozone dominate the radiative-thermal properties of the stratosphere and mesosphere, commonly referred to as the middle atmosphere. While other gasses do influence the radiative and thermal properties of the atmosphere, the increase of atmospheric CO_2 concentrations from preindustrial times coupled with its radiative properties has lead many climate scientists to believe that global climate change is principally driven by increased CO_2 levels.

Atmosphere models predict that increasing CO_2 concentrations increase heat retention in the troposphere and heat loss in the stratosphere and mesosphere. By the end of the century surface temperatures are expected to increase by about 1.5 to 4ºC while the middle atmosphere is expected to cool between 4 to 11ºC [*Rind et al.*, 1990, 1998; *Held and Soden*, 2000; *Fomichev et al.*, 2007].

6.2. Solar Heating and Terrestrial Cooling

The principal source of terrestrial heat is from the sun, which emits about 4×10^{26} watts of power. The Earth receives between 1412 to 1321 W/m² depending on the

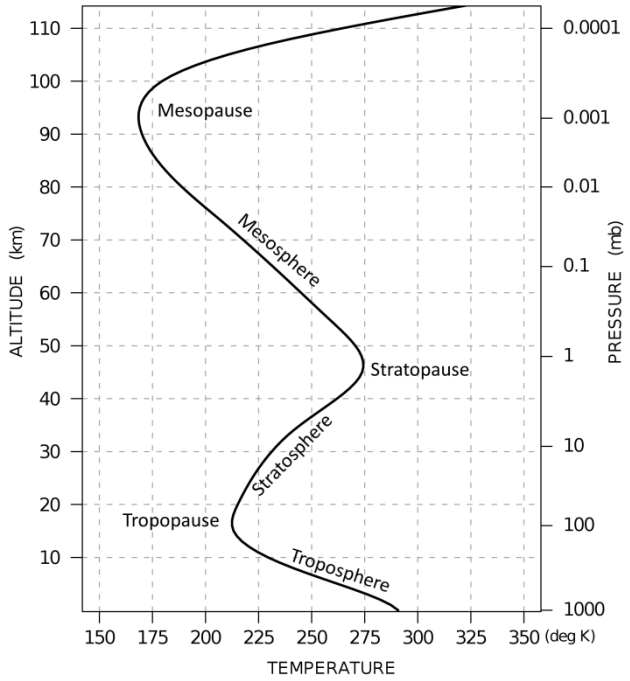


Figure 1.2. A model atmosphere temperature profile based on temperatures from the MSISe00 model.

Earth-sun distance. When averaged over the entire Earth's surface the incident energy flux is approximately 350 W/m^2. Roughly 50 to 70% of this reaches the Earth's surface. The rest is either reflected back into space or absorbed in the atmosphere or at the Earth's surface. For our purposes here, the incoming energy flux may be considered to be constant. This assumption simplifies the mechanism behind global warming: solar heating and terrestrial cooling. The sun warms the Earth and the Earth radiates heat back into space. If the heating and cooling balance, then the Earth is in radiative equilibrium and the average temperature is stable. If the cooling rate decreases, the planet will warm. If the cooling rate increases, the planet will cool.

A great deal of research indicates that increased CO_2 levels will increase global temperatures. In other words, increased CO_2 levels depress the net cooling rate and cause

the planet to retain more heat. The question now becomes why increased carbon dioxide levels decrease the global cooling rate.

The first thing to address is the temperature difference between the sun and the Earth. The temperature of the sun is approximately 5700 K and emits most of its radiation in the 250 to 2500 nm region with a peak in the visible spectrum around 500 nm. The average temperature of the Earth is about 290 K and emits radiation from 2.5 μm to well past 30 μm and peaks around 13 μm. Several atmospheric gases absorb radiation in this spectral range and can act to suppress the cooling rate. Figure 1.3 shows the spectral distribution of a blackbody at 220 K corresponding to a layer of atmosphere at 15 km and is nearly identical to the spectral distribution emitted from the Earth's surface. The peak of the distribution is near the 15 μm CO_2 absorption band. There is also a small CO_2 absorption band near 4 μm, but because it is far from the maximum it has a much smaller impact on the radiative properties of the atmosphere. There is also a significant O_3 peak near 9 μm.

6.3. The Role of Carbon Dioxide

The following discussion will be limited to the 12-18 μm CO_2 band. This narrow consideration provides a basic understanding of the physics involved, and though simplified, has pedagogical value.

In any gas, a strong absorption band also corresponds to a strong emission band. If a gas strongly absorbs in the 15 μm region, it will also strongly emit in that region. Likewise, a weak absorption band corresponds to weak emission band. An absorption coefficient of one indicates that a gas absorbs all external radiation passing through it and

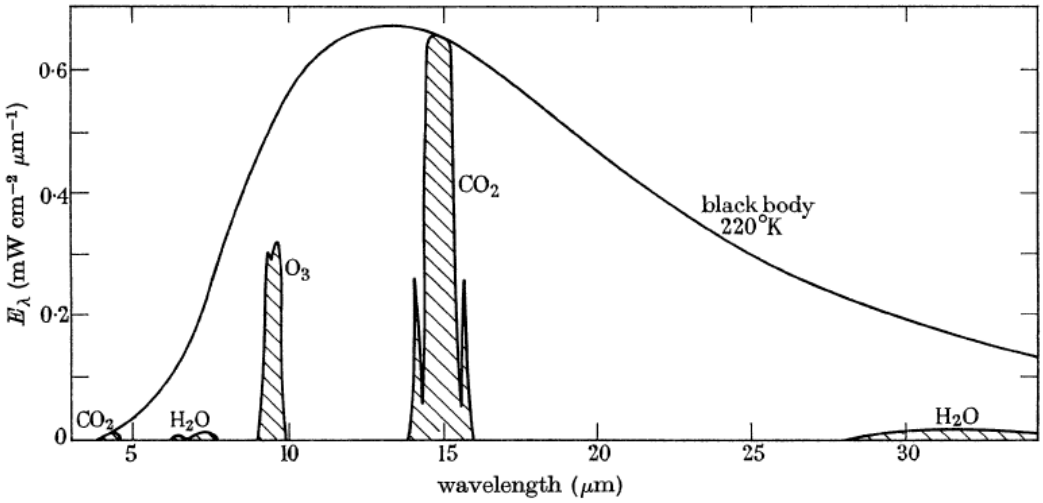


Figure 1.3. The spectral distribution from a layer of atmosphere at 15 km. (This figure is a reproduction of Figure 1 from *Houghton* [1965], p. 545. Used with permission of the Royal Society.)

that it emits radiation as a perfect blackbody. If the absorption coefficient is zero, then nothing is absorbed or emitted. If absorption is greater than zero but less than one, then the gas emits a fraction of the blackbody radiation and absorbs a fraction of the radiation passing through it.

For a given spectral band, if the atmosphere has zero absorption, then the downward radiation is zero because the atmosphere does not emit in that band, and all of the upward radiation originating from the Earth's surface passes into space. Thus the upward radiation is equal to the blackbody level corresponding to the Earth's temperature and the downward radiation is zero. If atmospheric absorption increases, then the downward radiation increases but emits at less than blackbody levels. The upward radiation from the Earth's surface is partially blocked, and to this is added the upwelling radiation from the atmosphere. The net effect is that the upwards radiation slightly exceeds blackbody levels and the downwards radiation is slightly less than blackbody levels. Increasing the amount

of atmospheric CO_2 increases the opacity of the atmosphere, decreasing the upwardradiation and increasing the downward radiation. The upward radiation decreases toward blackbody levels and the downward radiation increases toward blackbody levels. The net effect is a decrease in the radiation emitted into space. The cooling rate is depressed and the planet warms.

The actual temperature change is more complicated. The temporal cooling rate in K/year is given by

— —— ,

where ρ is the density, c_p is the specific heat at constant pressure, F↑ is the upwards radiation and F↓ is the downward radiation [*Lenoble*, 1993]. The temporal evolution of the temperature is proportional to the divergence of the net flux. When applying these equations to the problems just described, the heating and cooling rate can be estimated. Since the flux is temperature dependant and the temperature varies with altitude, the greatest heating and cooling occurs where the temperature gradients are greatest, near stratopause.

6.4. The Role of Water Vapor

A theoretical case for global warming can be made purely in terms of increased CO_2 levels, but water vapor is also important. Though the two dominant radiative gases in the troposphere are carbon dioxide and water vapor, water vapor emits a great deal more heat than carbon dioxide does. One reason for this is that water vapor emits radiation over a very broad spectral range whereas the radiative effects of CO_2 are restricted to well-

defined regions. Additionally, water vapor density decreases rapidly with decreasing temperature—colder air holds less water vapor. Consequently there is little water vapor in the upper troposphere and above. The temperature of the troposphere drops off at a rate of about 6.5 K/km. This temperature gradient causes the water vapor density to decrease with increasing altitude and creates a transition from opacity to transparency that occurs much faster than if the water vapor were uniformly mixed. For uniformly mixed gases this transition occurs over a greater altitude range. Consequently, adding a uniformly mixed gas to an atmosphere with a significant amount of water vapor suppresses the cooling effect of the water vapor by slowing the transition from opacity to transparency. *Clough and Iacono* [1995] note that "increases in uniformly mixed gases have the effect of reducing the cooling rate associated with water vapor alone while at the same time increasing the downward flux at the surface...in the region from 640 to 690 cm^{-1} the radiative effects of water vapor are effectively eliminated due to the strong absorptive properties of carbon dioxide."

Carbon dioxide has a constant mixing ratio up to about 80 km. According the calculations done by *Clough and Iacono* [1995] carbon dioxide moderates the strong cooling associated with water vapor due to its strong absorptive characteristics coupled with the fact that uniformly mixed trace gases, like CO_2, in atmospheres containing significant amounts of water vapor offset the water vapor cooling and increase the downward flux.

In the middle atmosphere the situation is somewhat different. A great deal of the outgoing longwave radiation in the 15μm band is absorbed in the troposphere so there is less upwards radiation in that band. And in contrast to the troposphere the middle

atmosphere has very little water vapor. Consequently increased CO_2 levels increases both the outgoing flux and the cooling rate.

The global effect of increased carbon dioxide levels is warming in the troposphere and cooling in the middle atmosphere. This is illustrated in Figure 1.4, which shows the difference in atmosphere cooling rate according to altitude and spectral region from two CO_2 baseline levels of 335 ppm and 710 ppm. The scale is in terms of cooling so a negative number indicates warming.

The region of maximum troposphere warming surrounds the 667 cm^{-1} (15 μm) band and ranges from 550 cm^{-1} to 800 cm^{-1}, with a noticeable gap at the center of the band, which corresponds to the region where CO_2 has its strongest absorption. In contrast, middle atmosphere cooling in that band is significantly greater, 10 to 80 $K \cdot d^{-1}$ $(cm^{-1})^{-1}$ at the stratopause. Other important spectral regions are the 980 to 1080 cm^{-1}, 1080 to 1200 cm^{-1}, and 2050 to 2150 cm^{-1} ozone bands. However, according to the differences of the two baseline levels, these spectral regions have a much smaller influence on atmosphere cooling, a maximum of 0.5 $K \cdot d^{-1}$ $(cm^{-1})^{-1}$.

7. The USU Rayleigh Lidar

7.1. Scattering Theory

The scattering of light by particles that are very small compared to the wavelength is called Rayleigh scattering. It occurs when $2\pi r/\lambda \ll 1$, where r is the radius and λ the wavelength. The backscatter cross section of unpolarized light in air is given by [*Measures*, 1984]:

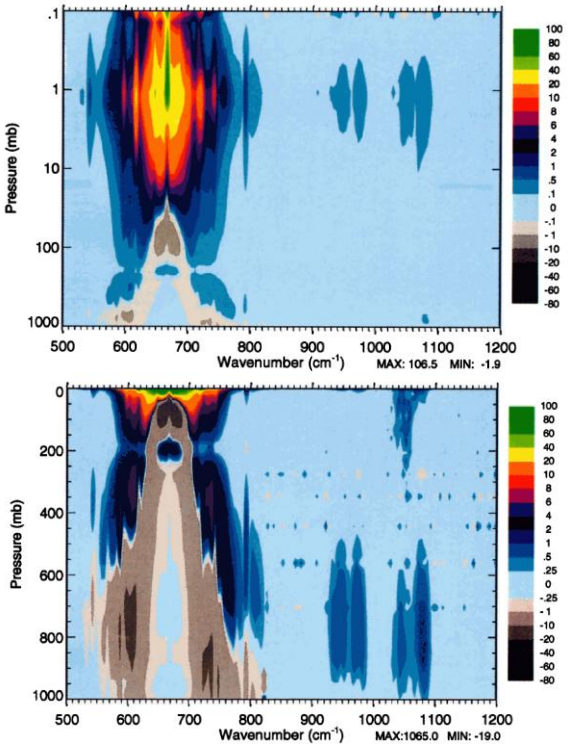


Figure 1.4. The change in heating and cooling rates by spectral region and altitude for the lower and middle atmosphere. These show the difference in cooling between CO_2 levels of 335 ppmv and 710 ppmv. The top and bottom figures are identical except for the vertical scale. (A wave number of 667 cm^{-1} corresponds to 15 μm.) Plate 5 from *Clough and Iacono* [1995], used with permission.

_____ ,

where λ is the wavelength, N is the number density, and n is the index of refraction of the gas. The return lidar signal for Rayleigh scattering from a layer of atmosphere at altitude z and thickness Δz is

_____ ,

where I_0 is the number of photons emitted, N(z) is the number density at altitude z, T is

the atmospheric transmittance, which is squared because of the return path; K is the optical efficiency of the system, and A is the telescope area [*Hauchecorne and Chanin*, 1980].

The temperature reduction equation is derived from hydrostatic equilibrium and the ideal gas law:

$$P = nkT, \tag{1.1}$$

$$dP/dz = -mgn, \tag{1.2}$$

where n is the number density, T is temperature, k is Boltzman's constant, P is the pressure, g is the gravitational constant, and m is the mean molecular mass. Substituting (1.1) into (1.2) and integrating between altitudes z_l and z_h yields

,

and lastly

,

where $T(z_h)$ is the initialization temperature and $n(z_h)$ is the measured, relative initialization number density at altitude z_h. The number densities are taken from the lidar equation

.

$T(z_h)$ is typically taken from a temperature climatology or an atmosphere model. The

mean molecular mass $m(z)$ is taken from the MSIS atmosphere model [*Beissner*, 1997; *Herron*, 2004]. Because the temperature integration includes a ratio of number densities, the constants 4π, k, A, I_0, σ_π^R, and Δz divide out leaving only a ratio of the signal strengths and altitudes.

The number density increases exponentially with decreasing altitude. After 15 km the initial temperature guess becomes insignificant because it is multiplied by $n(z_h)/n(z_l)$. This means that the system does not need external calibration and the temperatures from less than 80 km are very accurate. The temperature reduction was done by Dr. Joshua Herron. For more information about the temperature reduction see *Beissner* [1997] and *Herron* [2004, 2007].

7.2. Description of USU Lidar

A diagram of the USU lidar system is shown in Figure 1.5. The USU Rayleigh lidar consists of a Nd:YAG (neodymium:yttrium-aluminum-garnet) Spectra Physics laser. The

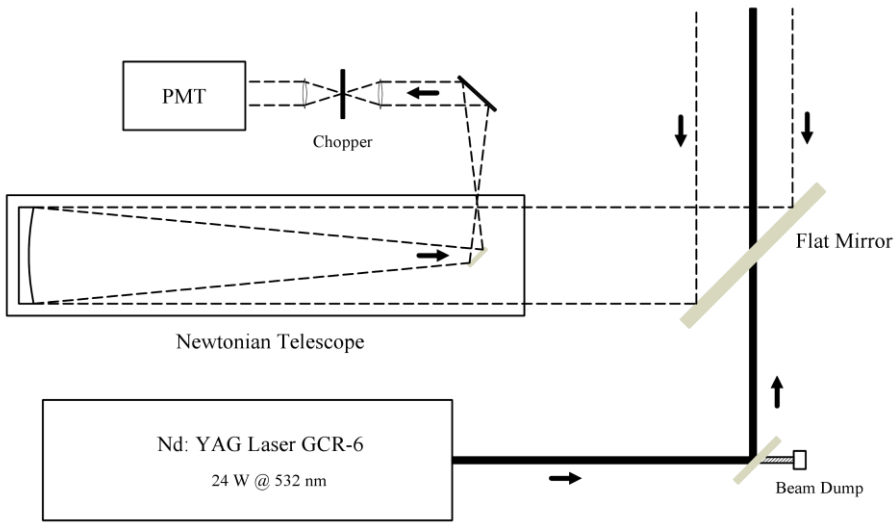


Figure 1.5. A diagram of the USU Rayleigh lidar system.

pulse width is 7-8 ns. The telescope is a 0.44 m (16 in) Newtonian with an effective area of 0.152 m^2 and a 2 to 3 mrad field of view and a 201 cm focal length. The collected light is focused into the plane of an optical chopper that blocks out strong lower altitude signals. It is then collimated and passes through a narrow band-pass filter and into a photo multiplier tube (PMT).

References

Akmaev, R. A., and V. I. Fomichev (2000), A model estimate of cooling in the mesosphere and lower thermosphere due to the CO2 Increase over the last 3–4 decades, *Geophys. Res. Lett.*, *27*(14), 2113–2116.

BBC (2010), *BBC Climate Change Poll - February 2010*. [online] Available from: http://news.bbc.co.uk/nol/shared/bsp/hi/pdfs/05_02_10climatechange.pdf (Accessed 28 June 2010).

Beig, G., et al. (2003), Review of mesospheric temperature trends, *Rev. Geophys.*, *41*(4), doi:10.1029/2002RG000121. [online] Available from: http://www.agu.org/pubs/crossref/2003/2002RG000121.shtml.

Beissner, K. C. (1997), *Studies of Mid-Latitude Mesospheric Temperature Variability and Its Relationship to Gravity Waves, Tides, and Planetary Waves*, Ph.D., Utah State University, Logan, Utah.

Callendar, G. S. (1938), The artificial production of carbon dioxide and its influence on temperature, *Quarterly Journal of the Royal Meteorological Society*, *64*(275), 223-240, doi:10.1002/qj.49706427503.

Chanin, M. L. (2006), Signature of the 11-year cycle in the upper atmosphere, *Space Science Reviews, 125(1)*, 261-272, doi: 10.1007/s11214-006-9062-5.

Clough, S. A., and M. J. Iacono (1995), Line-by-line calculation of atmospheric fluxes and cooling rates. II: Application to carbon dioxide, ozone, methane, nitrous oxide and the halocarbons, *J. Geophys. Res.*, *100*(D8), 16519-16535.

Fomichev, V. I., A. I. Jonsson, J. de Grandpré, S. R. Beagley, C. McLandress, K. Semeniuk, and T. G. Shepherd (2007), Response of the middle atmosphere to CO2 doubling: results from the canadian middle atmosphere model, *J. Climate*, *20*(7), 1121, doi:10.1175/JCLI4030.1.

Frederick, J. (1985), Measurement requirements for the detection of ozone trends, vol. 2362, pp. B1-B19, NASA, Goddard Space Flight Center. [online] Available from: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19850014801_1985014801.pdf.

Gruzdev, A. N., and G. P. Brasseur (2005), Long-term changes in the mesosphere calculated by a two-dimensional model, *J. Geophys. Res.*, *110*(D3), doi:10.1029/2003JD004410. [online] Available from: http://www.agu.org/pubs/crossref/2005/2003JD004410.shtml.

Hauchecorne, A., and M. Chanin (1980), Density and temperature profiles obtained by lidar between 35 and 70 km, *Geophys. Res. Lett.*, *7*(8), 565–568.

Held, I. M., and B. J. Soden (2000), Water vapor feedback and global warming, *Annual Rev. Energy and Envir.*, *25(1)*, 441-475, doi:10.1146/annurev.energy.25.1.441.

Herron, J. P. (2004), *Mesospheric Temperature Climatology Above Utah State University*, M.S., Utah State University, Logan, Utah.

Herron, J. P. (2007), *Rayleigh-Scatter Lidar Observations at USU's Atmospheric Lidar Observatory (Logan, UT): Temperature Climatology, Temperature Comparisons with MSIS, and Noctilucent Clouds*, Ph.D., Utah State University, Logan, Utah.

Houghton, J. T. (1965), Infrared Emission from the Stratosphere and Mesosphere, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *288*(1415), 545-555.

Huang, T., and G. Brasseur (1993), Effect of long-term solar variability in a two-dimensional interactive model of the middle atmosphere, *J. Geophys. Res.*, *98*(D11), 20413–20427.

Kerzenmacher, T. E., P. Keckhut, A. Hauchecorne, and M. Chanin (2006), Methodological uncertainties in multi-regression analyses of middle-atmospheric data series, *J. Environ. Monit.*, *8*(7), 682, doi:10.1039/b603750j.

Khosravi, R., G. Brasseur, A. Smith, D. Rusch, S. Walters, S. Chabrillat, and G. Kockarts (2002), Response of the mesosphere to human-induced perturbations and solar variability calculated by a 2-D model, *J. Geophys. Res.*, *107*(D18), 4358.

Krzyścin, J. (1997), Detection of a trend superposed on a serially correlated time series, *J. Atmos. Sol. Terr. Phys.*, *59*(1), 21-30, doi:10.1016/S1364-6826(96)00003-X.

Ledley, T. S., E. T. Sundquist, D. K. H. Schwartz, J. D. Fellows, and T. L. Killeen (1999), Climate change and greenhouse gases, *EOS*, *80*(39), 453.

Leiserowitz, A., E. Maibach, and C. Roser-Renouf (2010), *Climate Change in the American Mind: Americans' Global Warming Beliefs and Attitudes in January 2010*, Yale Project on Climate Change Communication, Yale University and George Mason University, New Have, CT. [online] Available from: http://www.cf.ac.uk/psych/home2/docs/UnderstandingRiskFinalReport.pdf (Accessed 28 June 1020).

Lenoble, J. (1993), *Atmospheric Radiative Transfer*, A. Deepak Pub., Hampton, Virginia.

Mäder, J. A., J. Staehelin, D. Brunner, W. A. Stahel, I. Wohltmann, and T. Peter (2007), Statistical modeling of total ozone: Selection of appropriate explanatory variables, *J. Geophys. Res.*, *112*(D11), doi:10.1029/2006JD007694. [online] Available from: http://www.agu.org/pubs/crossref/2007/2006JD007694.shtml.

Marsh, D. R., R. R. Garcia, D. E. Kinnison, B. A. Boville, F. Sassi, S. Solomon, and K. Matthes (2007), Modeling the whole atmosphere response to solar cycle changes in radiative and geomagnetic forcing, *J. Geophys. Res.*, *112*(D23), doi:10.1029/2006JD008306. [online] Available from: http://www.agu.org/pubs/crossref/2007/2006JD008306.shtml.

Measures, R. M. (1984), *Laser Remote Sensing: Fundamentals and Applications*, John Wiley and Sons, Malabar, Florida.

Newport, F. (2010), *Americans' Global Warming Concerns Continue to Drop*, The Gallup Organization. [online] Available from: http://www.gallup.com/poll/126560/americans-global-warming-concerns-continue-drop.aspx (Accessed 7 June 2010).

Oxburgh, R., K. Emanuel, L. Graumlich, D. Hand, H. Huppert, and M. Kelly (2010), *Report of the International Panel set up by the University of East Anglia to examine the research of the Climatic Research Unit*, University of East Anglia. [online] Available from: http://www.uea.ac.uk/mac/comm/media/press/CRUstatements/oxburgh.

Parry, M., O. Canziani, J. Palutikof, P. van der Linden, and C. Hanson (Eds.) (2007), *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK. [online] Available from: http://www.ipcc.ch/pdf/assessment-report/ar4/wg2/ar4-wg2-intro.pdf.

Ramaswamy, V., et al. (2001), Stratospheric temperature trends: Observations and model simulations, *Rev. Geophys.*, *39*(1), 71–122.

Rasmussen Reports (2009), *Americans Skeptical of Science Behind Global Warming*, Opinion Poll, Rasmussen Reports, United States. [online] Available from:

http://www.rasmussenreports.com/public_content/business/econ_survey_toplines/december_2009/toplines_climate_change_december_1_2_2009 (Accessed 3 June 2010).

Rind, D., D. Shindell, P. Lonergan, and N. K. Balachandran (1998), Climate change and the middle atmosphere. Part III: The doubled CO2 climate revisited, *Journal of Climate*, *11*(5), 876–894.

Rind, D., R. Suozzo, N. K. Balachandran, and M. J. Prather (1990), Climate change and the middle atmosphere. Part I: The doubled CO2 climate, *J. Atmos. Sci.*, *47*(4), 475-494, doi:10.1175/1520-0469(1990)047<0475:CCATMA>2.0.CO;2.

Robel, R., and R. Dickinson (1989), How will changes in carbon dioxide and methane modify the mean structure of the mesosphere and thermosphere?, *Geophys. Res. Lett.*, *16*(12), 1441-1444.

Soden, B. J. (2005), The radiative signature of upper tropospheric moistening, *Science*, *310*(5749), 841-844, doi:10.1126/science.1115602.

Spence, A., A. Venables, N. Pidgeon, W. Poortinga, and C. Demski (2010), *Public Perceptions on Climate Change and Energy Futures in Britain: Summary of Findings of a Survey Conducte in January-March 2010*, Understanding Risk Working Paper 10-01, Technical Report, Cardiff: School of Psychology, Castle, Cardiff, UK. [online] Available from: http://www.cf.ac.uk/psych/home2/docs/UnderstandingRiskFinalReport.pdf.

Tans, P. (2008), Trends in carbon dioxide, *Atmospheric Carbon Dioxide - Mauna Loa*. [online] Available from: http://www.esrl.noaa.gov/gmd/ccgg/trends/ (Accessed 5 February 2010).

Tiao, G. C., G. C. Reinsel, D. Xu, J. H. Pedrick, X. Zhu, A. J. Miller, J. J. DeLuisi, C. L. Mateer, and D. J. Wuebbles (1990), Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation, *J. Geophys. Res.*, *95*(D12), 20507-20517, doi:10.1029/JD095iD12p20507.

Weatherhead, E. C., et al. (1998), Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res.*, *103*(D14), 17149–17161.

CHAPTER 2

COLLINEARITY BETWEEN THE SOLAR PROXY AND LINEAR TREND COEFFICIENTS

Abstract. Collinearity arises in a linear model when one or more of the explanatory variables have near linear dependence with one or more of the other variables. Its effects can be correlated regression coefficients, inflated standard errors, difficulty in identifying significant model variables, and coefficient sensitivity to model specification.

This paper focuses on the specific problem of coefficient correlation between the solar proxy and time (linear) regressors in a simple ordinary least squares model. For this case, collinearity arises when there is a near linear relationship between the solar proxy and time regressors, and varies according to the phase of the solar proxy and length of the data set. If the solar proxy maximum occurs in the middle of the second half of the data set, there is significant positive correlation between the time and solar regressors. Conversely, if the solar proxy maximum occurs in the middle of the first half of the data set, they possess significant negative correlation. This leads to correlation between model estimates of the time and solar coefficients. The optimal phase of the solar proxy relative to the data is for the solar proxy maximum or minimum to occur at the time center of the data set. In that particular case the correlation between the linear and solar proxy terms is minimized, along with their coefficient correlation. When the data set spans approximately 1.3 solar cycles or more, correlation between the time and solar coefficients is minimal and may be ignored. The difficulties created by the presence of collinearity are independent of the magnitude of the solar and the linear coefficients.

1. Introduction

There is compelling evidence that the Earth's climate is undergoing long-term change, and there is a strong consensus among scientists that this is largely due to anthropogenic influence [*IPCC*, 2007]. Model calculations have shown that increases in the level of carbon dioxide cause the lower atmosphere (troposphere) and middle atmosphere (stratosphere and mesosphere) to react differently: the lower atmosphere warms and the middle atmosphere cools. Furthermore, the rate of temperature change in the middle atmosphere is expected to be about ten times greater than that in the lower atmosphere [*Roble and Dickinson*, 1989; *Rind et al.*, 1990; *Fomichev et al.*, 2007]. Additionally, simulations by *Khosravi et al.* [2002] show that the middle atmosphere cooling rate and solar response are additive, which justifies independent linear and solar terms in an ordinary least squares model. Because of the expected higher cooling rate in the middle atmosphere many researchers have focused their efforts on trend detection in the stratosphere and mesosphere. Information about how atmospheric temperatures are evolving on decadal time scales, as well as seasonally, and to external influences such as solar variability, are often extracted using ordinary least squares regression (OLSR). However, OLS regression results can be affected by a variety of factors, many of which have already been considered. The effects of autocorrelation have been discussed in *Frederick* [1985], *Tiao et al.* [1990], *Krzyścin* [1997], and *Weatherhead et al.* [1998]. The effects of interventions such as volcanic eruptions and adjustments in instrumentation have been reviewed by *Weatherhead et al.* [1998]; and the effects of variable selection by *Kerzenmacher et al.* [2006] and *Mäder et al.* [2007]. *Mäder et al.* considers the possible effect of collinearity on the stepwise variable selection procedure.

But as yet no one has considered the possibility of collinearity on least squares model coefficients.

The OLSR technique provides the best linear unbiased estimator (BLUE), provided each measurement is unbiased and the errors are uncorrelated. Looking at it in a different way, a column of temperature data is projected onto a column space of independent variables so as to minimize the variance of the residuals. If the relevant independent variables are included in the model, then OLSR minimizes what the model cannot explain. However, even in the presence of collinearity the model coefficient estimates are still BLUE.

This consideration of collinearity was prompted by the analysis of 11-years of Rayleigh-lidar mesospheric temperatures from Utah State University [*Wickwar et al.*, 2001; *Herron and Wickwar*, 2010]. A simple OLSR analysis of the USU data from near the mesopause produced a cooling rate greater than 1 K/year (10 K/decade) and a questionable dependence on solar input. These results, which have not yet been published, did not seem right—we found a much greater linear cooling trend and a weaker solar response than those predicted and inferred from other data at slightly higher altitudes, e.g., *Offermann et al.* [2004].

In this paper a simplified regression model is employed, and the nature of the collinearity problem, as it applies to it, is discussed. The model, which contains a linear term and solar proxy only, is introduced in section 2. Section 3 explores the problem of collinearity between the time and solar regressors when the data set spans one solar cycle, and one half solar cycle. It was found in our simplified model that if the data set spanned one solar cycle or less, the model is susceptible to a strong collinearity problem,

depending on the phase of the solar proxy relative to the data. If the data set spans less than one solar cycle the problem can be severe. This is especially applicable to our USU Rayleigh-scatter lidar temperatures; our data set spans one solar cycle and the phase of the solar proxy is such that there exists a strong collinearity problem. In section 4 the special case of minimal collinearity is addressed. In section 5 we address the cause of the problem as it applies to the simplified model under consideration. Section 6 contains a discussion of a different situation, how a step function explanatory variable modeling an instrumentation change introduces a collinearity problem between regressors. In section 7 the collinearity problem as it applies to the USU temperature model is illustrated and briefly discussed. In section 8 the final conclusions are listed.

2. The Problem

The problems of collinearity (sometimes referred to by its redundant term multicollinearity) arise in the presence of near linear dependence among model variables. But near linear dependence is not a simple matter of regressors failing to be strictly linearly independent. Near linear dependence arises when $v_1\boldsymbol{A}_1 + \ldots + v_n\boldsymbol{A}_n = \boldsymbol{c}$, where $\boldsymbol{c}$ is a small number. In other words, collinearity arises when one or more regressors nearly lies in the space spanned by one or more of the other regressors. If such a relationship exists between two variables, or between groups of variables, the problems of collinearity arise. One might assume that if two variables are uncorrelated or orthogonal they have no near linear dependence. However, that is not necessarily the case. Two regressors can be perfectly correlated and orthogonal: For example, the two vectors $(-1, -1, -1, \sqrt{3})^T$ and $(1, 1, 1, \sqrt{3})^T$ are orthogonal and have a correlation of +1. Two regressors can be

uncorrelated and oblique: For example, the vectors $(2, -3, -2, 1)^T$ and $(0, -3, 3, -2)^T$ have zero correlation and inner product of +1 [*Rodgers et al.*, 1984]. And collinearity can arise when two regressors are uncorrelated and oblique, or correlated and orthogonal. Uncorrelated or orthogonal regressors make collinearity unlikely, but not impossible. Near linear dependence problems can still arise, and diagnosing the number of near dependences is not a simple matter of looking at correlations between regressor pairs. However, general collinearity is not the focus of this paper, but rather the correlations that can arise between estimated regression coefficients. *Belsley* [1991] has noted that correlated coefficients are a sufficient (but not a necessary) condition for collinearity. Our treatment of collinearity focuses on this sufficient condition.

If one or more regressors are strongly correlated with one or more other regressors then the problems of collinearity are likely to arise. In a simple model with two explanatory variables, if they are positively (negatively) correlated their estimated model coefficients will be negatively (positively) correlated. In general, the partial correlation between two regressors is the negative of the correlation between their regression coefficients. The correlation inflates coefficient standard errors (SEs), increasing the chance of determining a coefficient to be insignificant when in fact it is significant, or visa versa. In the absence of collinearity the regression coefficients indicate the effect of a one-unit change on such regressors. For example, if the coefficient of the time regressor (the cooling trend) is –0.3 K/year, then for every year the temperature decreases by 0.3 K. However, when regressors are correlated their coefficients are also likely to be correlated and, if so, a joint inference must be made. Because many models include time and solar regressors, understanding the nature of the collinearity between them, and being able to

identify when it occurs, is important.

OLSR on atmospheric temperatures generally includes the following explanatory variables: annual oscillation and semiannual oscillation, linear trend, and a solar proxy representing changes in solar input. It might also include information about the quasi-biennial oscillation, or short-term effects such as changes in atmospheric optical depth due to volcanic eruption. Consider the following model,

$$T_i = w + b \cdot t_i + s \cdot sp_i + A_1 \sin(2\pi \cdot t_i) + A_2 \cos(2\pi \cdot t_i) + B_1 \sin(4\pi \cdot t_i) + B_2 \cos(4\pi \cdot t_i) + \varepsilon_i, \quad (2.1)$$

where T_i is the temperature at time t_i, time is in years, w is the intercept term; b is the linear trend coefficient, s is the solar response coefficient, A_1 and A_2 give the amplitude and phase of the annual oscillation; B_1 and B_2 give the amplitude and phase of the semiannual oscillation; ε_i is the residual and sp is a solar proxy term. Because time may be measured from any given moment, the location of t = 0 is arbitrary. So we choose t = 0 to be the time center of the data set. Furthermore, several solar proxies have been employed as approximations to solar UV input: sunspot number, F10.7 solar flux, Mg II index, and the He I 1083-nm line. In place of a solar proxy *Remsberg* [2002] employs a sine function with a phase offset.

These six explanatory variables form a column space onto which the temperature column T is projected: (1) t, (2) sp, (3) $sin(2\pi \cdot t)$, (4) $cos(2\pi \cdot t)$, (5) $sin(4\pi \cdot t)$, (6) $cos(4\pi \cdot t)$, and (7) a column of 1s for the intercept. Under ideal conditions they would form an orthogonal column space, in which case collinearity is unlikely. When model variables are mean centered, orthogonality is equivalent to being uncorrelated, and visa versa.

Obviously (3), (4), (5), and (6) are orthogonal to each other. The sine and cosine functions are also nearly mean centered. And, as mentioned above, there is no special commitment to any particular solar proxy, so we take the liberty of mean centering (2). Since any mean-centered regressor is orthogonal to a constant vector, (1-6) are each orthogonal to (7). Also, correlation of (1) with (3-6), and (2) with (3-6) is minimal. So a simplified model may be considered,

$$T_i = 0 + b \cdot t_i + s \cdot sp_i + \varepsilon_i \,, \qquad (2.2)$$

where T, t, b, sp, and ε are as indicated above. The coefficients b and s are the linear trend and solar proxy coefficients respectively. The temperatures T are also mean centered, allowing the regression to be forced through zero, indicated by the zero on the right hand side of Equation (2.2). Because the other regressors are orthogonal and/or uncorrelated to the model variables in Equation (2.2) removing them from the model should not affect the other coefficient estimates. It is generally recommended that mean centering not be used when diagnosing collinearity problems [*Belsley*, 1991; *Draper and Smith*, 1998]. However, as the sines and cosines are nearly mean centered and the location of t = 0 is arbitrary and because we have no special commitment to any particular solar proxy, we employ mean centering here. Furthermore, mean centering does not affect coefficient estimates, hence coefficient correlations, which are the focus of this paper, will be unaffected. Mean centering can otherwise mask underlying general collinearity problems. To further simplify, a sine function with unit amplitude and 11-year period (approximately equivalent to one solar cycle) was used in place of a solar proxy. Gaussian noise was added to the sine solar proxy in such a way as to simulate the greater

random variability observed in solar proxies at solar maximum (see Figures 2.1a, 2.2a, 2.3a, and 2.4a for examples). The phase of the solar proxy is referenced to the time center ($t = 0$) of the data set; the phase will be measured from the point where the solar proxy is halfway between solar maximum and solar minimum.

The data for this analysis is drawn from data simulations. A time series of temperatures was generated having a cooling rate of -0.4 K/year and a 4 K solar response. Gaussian noise ($\sigma = 12.2$ K) was added to the temperatures. Both the cooling rate and magnitude of the solar response are for the upper mesosphere and were taken from *Keckhut et al.* [1995]; 12.2 K is the average standard deviation of the residuals of an OLSR on the USU temperatures from the MLT region. Equation (2.2) was then applied to the simulated data. The nature of the collinearity problem under investigation is independent of the magnitude of the estimated solar and linear trend coefficients.

3. Strong Collinearity

The equation for the correlation between any two estimated regression coefficients is $\rho_{mn} = s_{mn}/\sqrt{(s_{nn} \cdot s_{mm})}$, where s_{mn}, is the (m,n)th element of the variance covariance matrix $s_e^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. One way to visualize how the coefficient estimates are correlated is through a technique called bootstrapping. After an initial OLSR, coefficient estimates, residuals and predicted values are obtain. The residuals are then randomly sampled with replacement and added to the predicted values $E(T) = b \cdot t + s \cdot sp$. Replacement means that for n data points n residuals are selected in such a way that any given residual may be selected more than once or not at all. These sampled residuals are then added to the predicted values $E(T) = b \cdot t + s \cdot sp$, creating a new data set. An OLSR is done on this new data set from

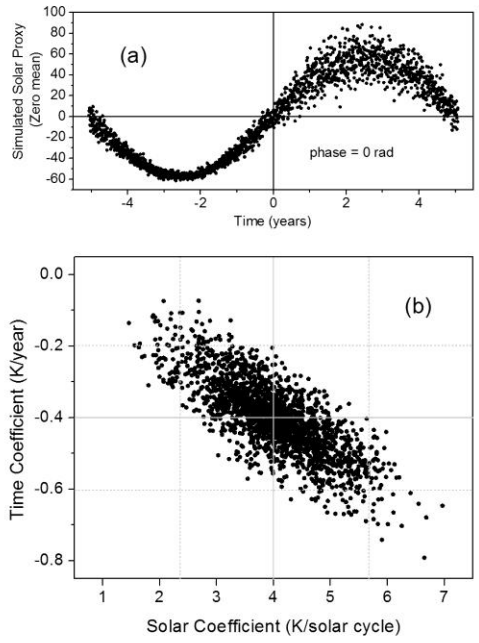


Figure 2.1. Simulated solar proxy and correlation scatter plots. Plot (a) shows a simulated solar proxy with zero mean. Plot (b) shows a pairs plot of linear trend and solar coefficient estimates from bootstrapped regressions when the solar proxy has 0^o phase and spans one solar cycle. The plot shows coefficient estimates from 1500 bootstrapped regressions of Eq. (2.2). The solar response coefficients were multiplied by 2 to put them on a scale of K (solar max – solar min).

which new (and slightly different) coefficient estimates obtain. This process is repeated approximately 500 to 2000 times, producing a set of coefficients from each bootstrapped regression. From these coefficient sets the distribution of each estimate may be inferred. (For more on bootstrap methods see *Shalizi* [2010] and *Hollander and Wolfe* [1999].) This technique has the advantage of avoiding assumptions about the underlying coefficient distribution. By plotting one set of coefficients against another, the effect of coefficient correlation becomes apparent; the pattern is similar to that in Figure 2.1b, which shows time and solar coefficients from 1500 bootstrapped regressions plotted against each other. The elliptical pattern in Figure 2.1b is that of a bivariate normal

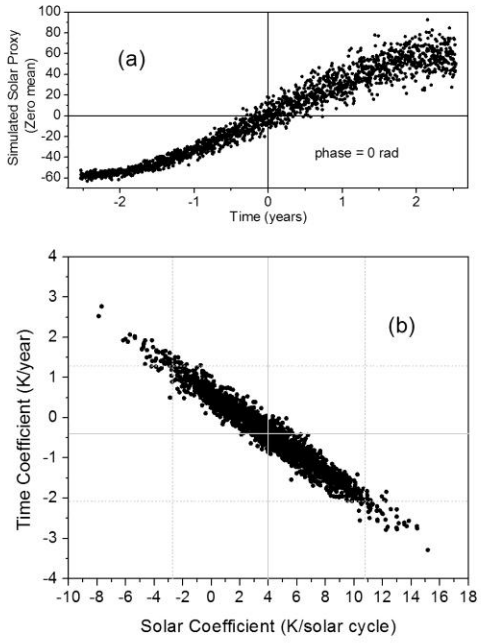


Figure 2.2: Same as Figure 2.1 except for a data set spanning half a solar cycle.

distribution and indicative of the kind of collinearity under consideration.

If there were no collinearity the elliptical pattern would be horizontal, vertical, or circular (e.g., in Figure 2.3b and 2.4b the coefficients are uncorrelated), and the confidence intervals would indicate, to a specified level of confidence, the region that presumably includes the true value of the estimator. Also, the possible values of one coefficient would say nothing about the possible values of the other. But when coefficients are correlated the elliptical pattern has a clear slant to it (e.g., Figures 2.1b and 2.2b). In such cases the region covered by the indicates possible coefficient values from any given regression. As mentioned previously, the data have a −4 K/year cooling rate and 4 K solar response. The phase of the solar proxy is indicated in Figures 2.1a and 2.2a. The dashed gray lines indicate the 2σ or 95% confidence intervals for each coefficient. In Figure 2.1 they indicate a likely cooling rate between

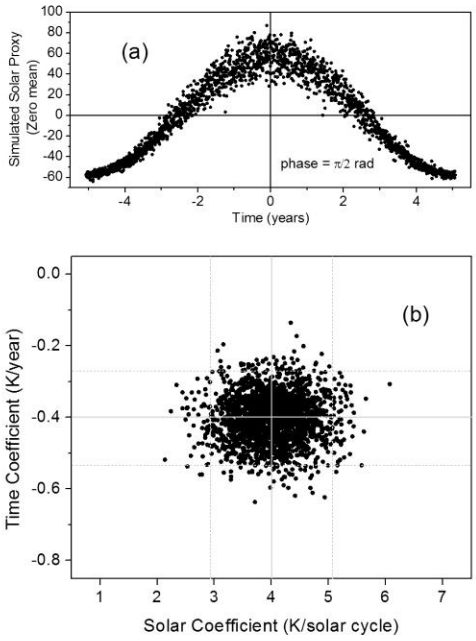


Figure 2.3: Linear and solar coefficient estimates from bootstrapped regressions when the solar proxy has a 90^{o} phase. (Solar max is at the time center of the data set.) Figures (a) and (b) are otherwise the same as in Figure 2.1. The coefficients are uncorrelated.

–0.2 and –0.6 K/year and a likely solar response between 2.2 and 5.7 K. But because the coefficients are correlated joint inferences cannot be made freely. For example, the regressions indicate it would be highly unlikely that the true value of the solar response be 2.5 K and the true value of the time coefficient be –0.6 K/year. Those two values when taken together are outside the elliptical region jointly covered by the bootstrapped time and solar coefficient estimates. Also, if an inference is made from an independent source that the true value of the solar response is 4.7 K, then at s = 4.7 K the horizontal and vertical width of the elliptical region is quite narrow and the cooling rate is restricted to approximately –0.54 to –0.40 K/year, which is much narrower than the overall spread. Similarly, if it is independently inferred that the cooling rate is −0.3 K/year, then the solar response is approximately between 2.7 and 3.9 K. If the value of one is specified then a

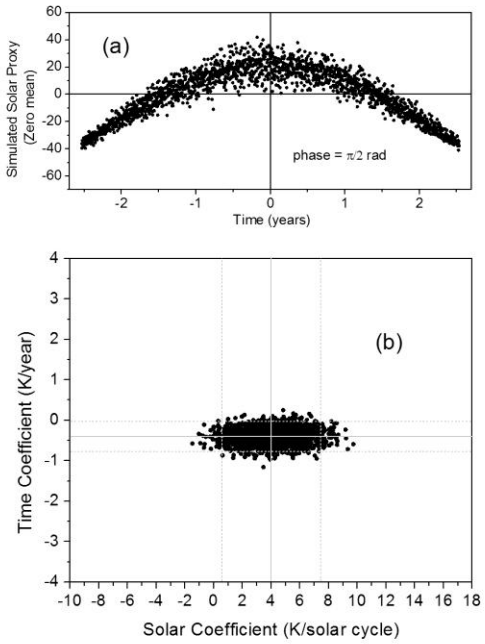


Figure 2.4: Same as Figure 2.2 except the solar proxy maximum is at the time center of the data set. (A solar proxy phase angle of $90^{o.}$)

narrow confidence interval for the other obtains. The important point is that when coefficient estimates are strongly correlated, inferences about one coefficient cannot be made independent of inferences about another to which it is correlated.

Each coefficient value in Figures 2.1b-2.4b can also be thought of as a possible mean value for solar and time coefficient estimates obtained from an OLSR on local measurements from 1500 different data collection sites. Assuming that the temperature data at every site has a –0.4 K/year cooling rate and a 4 K solar response, and assuming that the same model is fit to the data, then the distribution shown in Figure 2.1b represents possible OLSR time and solar coefficients from 1500 sites. Since the standard error estimates (SEs) are unaffected by mean coefficient values, they each have a bivariate distribution similar to the overall pattern but centered on their own mean

coefficient value. Using Figure 2.1b as an example, if the time coefficient estimate from a given site has a smaller magnitude (i.e. a greater cooling rate) than the actual cooling rate the solar coefficient estimate is likely to be too small. But if the cooling rate estimate is larger than the actual cooling rate then the solar coefficient estimate is likely to be greater than the true solar response. However, true values are typically unknown, making it difficult to know if the results from any given site have a high/low or low/high tendency.

Figure 2.2b illustrates another example of extreme coefficient correlation. It differs from the first case in that the simulated data extends one-half solar cycle. In this case the correlation between coefficients is extreme and the error is unacceptably high.

One might consider omitting the solar proxy altogether. However, if there is a significant solar response in the temperature data this could introduce significant bias to the linear trend coefficient.

4. No Collinearity

When the solar proxy phase angle is $\pm 90^{o}$, that is, when the solar proxy maximum or minimum is at the time center of the data set, there is no correlation between the time and solar coefficient estimates. Figures 2.3b and 2.4b show cases with no coefficient correlation for data sets spanning 1 and ½ solar cycles, respectively. Under these conditions, there is no coefficient correlation between the estimated time and solar response coefficients. The true value of the solar and time coefficients is within the elliptical region, but the value of one coefficient estimate says nothing about the value of the other; inferences about one may be made without reference to the other. Also, when compared to the spread in Figures 2.1b and 2.2b, the overall spread in Figures 2.3b and

2.4b is substantially narrower.

The phase of the solar proxy can have a significant impact on the estimated coefficient SEs. In going from Figure 2.3b to Figure 2.1b the standard deviation of the time and solar coefficient estimates each increased 59%. In going from Figure 2.4b to Figure 2.2b, the standard deviation of the solar coefficient estimates increased 96%, and that of the time coefficient 327%.

5. Effect of Collinearity on Standard Errors

High standard errors can indicate greater sample-to-sample variation; however, collinearity can also inflate coefficient standard errors, increasing the possibility that a given coefficient estimate is far from its true value. The reason behind larger standard errors can be seen more clearly from the equation for the SE of a regression coefficient

$$SE_{bk} = \frac{s_e}{\sqrt{(1 - R_k^2)TSS_k}}, \tag{2.3}$$

where SE_{bk} is the estimated standard error of the coefficient for the X_k regressor. s_e is the residual standard deviation, R_k^2 is the coefficient of determination obtained by regressing X_k on the other X variables, and TSS_k is the total sum of squares, , where

is the average of X_k. The factor $(1 - R_k^2)^{-1}$ is called the variance inflation factor (VIF). High VIFs can indicate a collinearity problem. (As with correlated coefficients, high VIFs are sufficient but not necessary for collinearity. Also, the number of near dependences cannot be determined from VIFs, and there is no objective standard to determine how high a VIF must be to indicate a collinearity problem.) Because there are

only two explanatory variables in the model under consideration, the coefficient of determination R_k^2 becomes the square of the correlation between the time and solar regressors, and the total sum of squares may be written as $TSS_k = \Sigma X_{ki}^2 = |X_k|^2 = X_{k,1}^2 + \ldots + X_{k,n}^2$.

Applying these changes to Equation (2.3) and writing the equations for the standard error estimates for the time and solar coefficients we get the following two equations:

$$SE_t = \frac{s_e}{\sqrt{(1-\rho_{s,t}^2)\left|t\right|^2}}, \tag{2.4}$$

$$SE_s = \frac{s_e}{\sqrt{(1-\rho_{s,t}^2)\left|sp\right|^2}}, \tag{2.5}$$

where $\rho_{s,t}$ is the coefficient of correlation between the solar and time regressors, and $|sp|^2$ and $|t|^2$ are the square of the magnitudes of the solar and time explanatory variables, respectively. That is,

$$\left|sp\right|^2 = \sum_i (sp_i)^2 ,$$

$$\left|t\right|^2 = \sum_i t_i^{\,2} .$$

With the SEs written in this form it is easier to see how the interaction of the two independent variables and length of the data set influence the SE estimates. A high correlation between the solar proxy and time regressors, i.e. $\rho_{s,t}$ approaching 1, increases the standard error. In contrast, a longer data set has a larger $|sp|^2$ and $|t|^2$, which lowers the standard error.

However, it was found from simulations that if the data set spans 1.3 solar cycles or more, then $|sp|^2$ and $|t|^2$ are sufficiently large and the correlation between the time and solar regressors is sufficiently small that the SEs are not greatly inflated and the coefficient correlation between their respective coefficients negligible.

6. Effect of Instrument Changes on Collinearity

If there is a large temperature perturbation near the beginning or end of a data set, there is likely correlation between the regressor modeling the perturbation and the time and/or solar regressor. A correlation problem would also apply to calibration step functions sometimes used in regressions on rocketsonde temperatures. Over time instrumentation changes may have introduced bias into the temperature measurements, and a step function is sometimes used to account for these changes. Such temperatures from various sites can span several decades (Ryori, Japan [*Keckhut and Kodera*, 1999]; US rocketsondes in North and South America [*Keckhut*, 1999]; and Volgograd, Russia [*Kubicki et al.*, 2006]).

As an illustration, temperatures from the Volgograd site shown in Figure 2.5a were taken from *Kubicki et al.* [2006]. These temperatures have an instrumentation change in the first half of the data set. The present authors applied an OLSR model to the data, containing a unit step function, solar proxy, and time regressor. From this initial regression, sets of coefficient estimates were calculated using the bootstrap method. Figure 2.5(b, c, and d) are pairs plots of the solar, time, and step function coefficient estimates obtained from bootstrapping. There is a clear correlation between the step function and time coefficient estimates (Figure 2.5d), if there were no correlation the

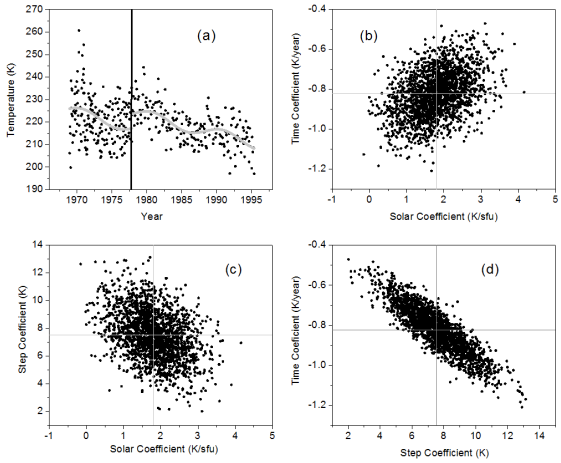


Figure 2.5: (a) Average rocketsonde temperatures above Volgograd Russia from 55-75 km from January 1969 to September 1995, taken from Figure 1 of *Kubicki et al.* [2006]. The vertical line shows a sensor change that occurred April 1978. The solid gray line shows the predicted values from an OLSR that include linear trend, solar proxy, and step function terms. (b) The bootstrapped coefficient estimates for the solar proxy and time explanatory variables. The coefficient correlation is small. (c) Same as (b) but for solar and step function coefficient estimates. (d) A pair's plot of the estimated step function coefficients and linear trend coefficients. The coefficient correlation is strong.

elliptical pattern would be horizontal, vertical, or circular. If the value of the step function coefficient is unimportant, the correlation between it and the linear trend coefficient need not be considered. Also, there is a weak correlation between the estimated solar and step function coefficients (Figure 2.5c). However, it appears small enough to be neglected.

7. The USU Lidar Data

The USU data used for this dissertation consists of 593 data points at 45 km spanning a time from September 1993 to August 2003. For a review of the number of data points and the seasonal distribution of the nightly observations see Chapter 5. The collinearity between the linear term and solar proxy from Equation (2.1) as applied to the USU temperatures at 45 and 90 km is shown in Figure 2.6, where the solar proxy is the MgII

data. The effects of collinearity was shown, as before, by the bootstrapping technique using the trend and solar terms found in the OLSR. The correlation between the linear terms and solar proxy terms is nearly the same at both altitudes because the correlation is based on the model and not the temperatures. (At 45 km the correlation between the sin ωt term and the linear term is −0.86. At 90 km it is −0.88.) The standard deviations determine how wide the elliptical pattern is. The slope of a line going through the principal axis of the elliptical pattern shown in Figure 2.6 is given by $a = \rho \cdot SE_b/SE_s$, where a is the slope, ρ is the correlation between the b and s coefficients (linear trend coefficient and solar proxy coefficient), and SE_b and SE_s are the standard errors of the b

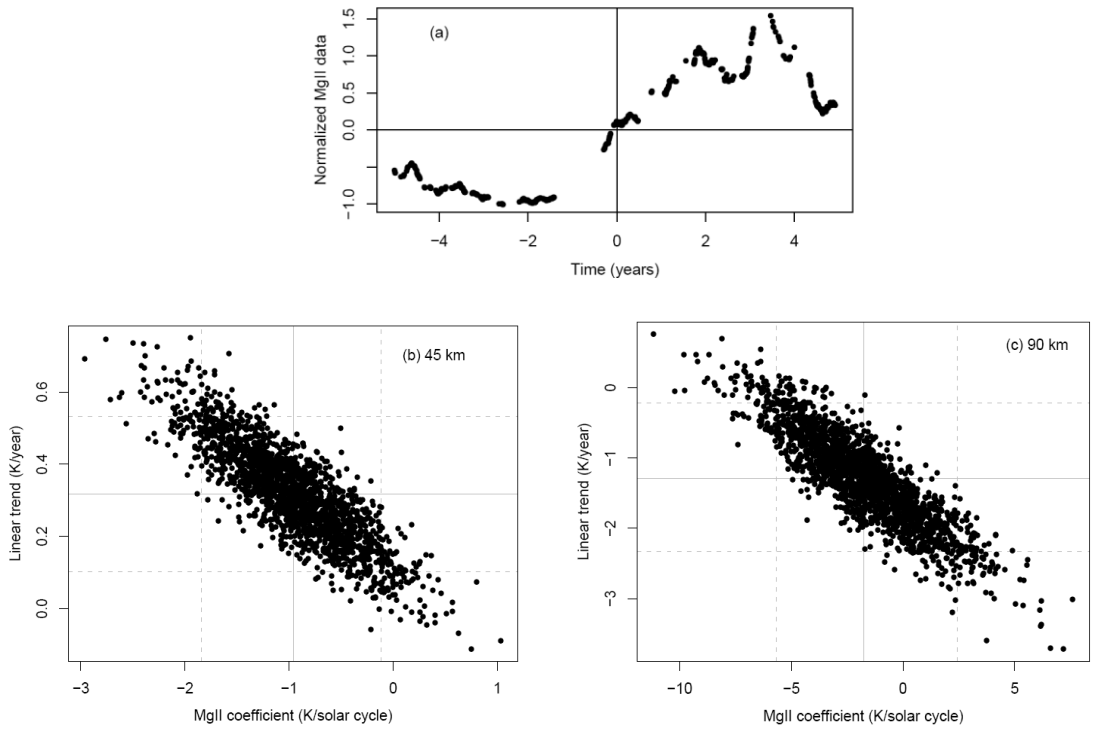


Figure 2.6: A realistic collinearity simulation. Plot (a) shows the MgII data used in the regression. Plots (b) and (c) illustrate the collinearity between the linear term and solar proxy term for Equation (2.1) applied to the USU temperatures at 45 and 90 km.

and *s* coefficients. The SE of the MgII coefficient at 90 km is nearly 5 times larger than its SE at 45 km. This is also true for the linear term: Its SE at 90 km is nearly 5 times larger than at 45 km. It is also noteworthy that at 90 km the atmospheric solar response is such that it is possible to get a change in sign. If the linear trend is −2 K/year, then the atmospheric solar response will go from negative to slightly positive. At 45 km if the linear trend is +0.1 K/year, then the solar response goes from −1 K/solar cycle to near zero.

8. Conclusions

In most cases not much can be done about collinearity. Even in cases where the time and solar regressors are collinear the coefficient estimates are still BLUE. However, for collinearity problems where coefficient estimates are correlated their possible values are restricted to an elliptical region shown by the bootstrapped coefficient estimates.

The simulations in this paper show that models giving questionable trend and solar responses can arise because of problems relating to collinearity.

(1) Collinearity between the time and solar proxy variables leads to their respective coefficient estimates being correlated if solar maximum or minimum does not occur near the time center of the data set. For a step function regressor, if the location of the step transition (i.e. where step transitions from low to high) is near the time center of the data set, then there will be a strong correlation between the time and step function regressors, thus their coefficient estimates are also likely to be correlated; the strongest correlation between the time and step function regressors occurs when the step is in the time center of the data set. More generally, any explanatory variable modeling a temperature

oscillation should be considered for possible coefficient correlation with the other coefficient estimates. For example, data modeling a temperature spike due to volcanic eruption could be correlated with the time and/or solar regressors. (2) Coefficients of collinear regressors are likely to be correlated. (3) The SEs of correlated regressors will be inflated. (4) The effects of coefficient correlation can be visualized using the bootstrap method. This is an easy method that does not make assumptions about the underlying coefficient distributions and allows a visual representation of the region of possible coefficient values. (5) In the case of collinearity between time and solar regressors, the effects of collinearity can be mitigated by extending the data set. Our simulations indicate that extending the data set to 1.3 solar cycles significantly reduces the coefficient correlation between the estimated time and solar coefficients. In the case of a step function regressor, the maximum coefficient correlation between the step function and time regressors occur when the step is near (or at) the time center of the data set. If the step is closer to the beginning (or end) of the data set, as would happen as data is added to the data set, the coefficient correlation is diminished. (6) If one regressor has a near linear dependence with another, a joint interpretation is unnecessary if the coefficient estimate of one of the regressors is unimportant. (7) Inflated SEs and the degree of coefficient correlation are independent of the magnitude of the coefficient estimates themselves.

This may also explain some of the results reported in Table 5 of *Beig et al.* [2003] in which temperature trends were derived using various methods such as OH emission layer, lidar, rocketsonde, etc., from the MLT region and ranged from +5 K/decade to −10.5 K/decade. Several of the data sets listed span approximately one solar cycle or less. As far as the USU data set is concerned, our data set spans one solar cycle, and the

addition of three to four years of additional data could effectively eliminate the collinearity problem.

References

Beig, G., et al. (2003), Review of mesospheric temperature trends, *Rev. Geophys.*, *41*(4), doi:10.1029/2002RG000121. [online] Available from: http://www.agu.org/pubs/crossref/2003/2002RG000121.shtml.

Belsley, D. A. (1991), *Conditioning Diagnostics*, Wiley, New York.

Draper, N. R., and H. Smith (1998), *Applied Regression Analysis*, Wiley, New York.

Fomichev, V. I., A. I. Jonsson, J. de Grandpré, S. R. Beagley, C. McLandress, K. Semeniuk, and T. G. Shepherd (2007), Response of the middle atmosphere to CO2 doubling: results from the canadian middle atmosphere model, *J. Climate*, *20*(7), 1121, doi:10.1175/JCLI4030.1.

Frederick, J. (1985), Measurement requirements for the detection of ozone trends, vol. 2362, pp. B1-B19, NASA, Goddard Space Flight Center. [online] Available from: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19850014801_1985014801.pdf

Herron, J., and V. Wickwar (2010), Mid-Latitude Mesospheric Temperature Climatology Obtained with the Rayleigh-Scatter Lidar at USU's Atmospheric Lidar Observatory (41.7ºN, 111.8ºW), manuscript.

Hollander, M., and D. A. Wolfe (1999), *Nonparametric Statistical Methods*, Wiley, New York.

IPCC (2007), *Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis.* Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge Univ. Press, Cambridge, UK and New York, NY, USA.

Keckhut, P. (1999), Stratospheric and mesospheric cooling trend estimates from U.S. rocketsondes at low latitude stations (8°S–34°N), taking into account instrumental changes and natural variability, *J. Atmos. Sol. Terr. Phys.*, *61*(6), 447-459, doi:10.1016/S1364-6826(98)00139-4.

Keckhut, P., and K. Kodera (1999), Long-term changes of the upper stratosphere as seen by Japanese rocketsondes at Ryori (39°N, 141°E), *Ann. Geophys.*, *17*(9), 1210-1217.

Keckhut, P., A. Hauchecorne, and M. L. Chanin (1995), Midlatitude long-term variability of the middle atmosphere: Trends and cyclic and episodic changes, *J. Geophys. Res.*, *100*(D9), 18887–18897.

Kerzenmacher, T. E., P. Keckhut, A. Hauchecorne, and M. Chanin (2006), Methodological uncertainties in multi-regression analyses of middle-atmospheric data series, *J. Environ. Monit.*, *8*(7), 682, doi:10.1039/b603750j.

Khosravi, R., G. Brasseur, A. Smith, D. Rusch, S. Walters, S. Chabrillat, and G. Kockarts (2002), Response of the mesosphere to human-induced perturbations and solar variability calculated by a 2-D model, *J. Geophys. Res.*, *107*(D18), 4358.

Krzyścin, J. (1997), Detection of a trend superposed on a serially correlated time series, *J. Atmos. Sol. Terr. Phys.*, *59*(1), 21-30, doi:10.1016/S1364-6826(96)00003-X.

Kubicki, A., P. Keckhut, M. Chanin, A. Hauchecorne, E. Lysenko, and G. Golitsyn (2006), Temperature trends in the middle atmosphere as seen by historical Russian rocket launches: Part 1, Volgograd (48.68°N, 44.35°E), *J. Atmos. Sol. Terr. Phys*, *68*(10), 1075-1086, doi:10.1016/j.jastp.2006.02.001.

Mäder, J. A., J. Staehelin, D. Brunner, W. A. Stahel, I. Wohltmann, and T. Peter (2007), Statistical modeling of total ozone: Selection of appropriate explanatory variables, *J. Geophys. Res.*, *112*(D11), doi:10.1029/2006JD007694. [online] Available from: http://www.agu.org/pubs/crossref/2007/2006JD007694.shtml.

Offermann, D., M. Donner, P. Knieling, and B. Naujokat (2004), Middle atmosphere temperature changes and the duration of summer, *J. Atmos. Sol. Terr. Phys*, *66*(6-9), 437-450, doi:doi: DOI: 10.1016/j.jastp.2004.01.028.

Remsberg, E. E. (2002), Seasonal and longer-term variations in middle atmosphere temperature from HALOE on UARS, *J. Geophys. Res.*, *107*(D19), doi:10.1029/2001JD001366. [online] Available from: http://www.agu.org/pubs/crossref/2002/2001JD001366.shtml.

Rind, D., R. Suozzo, N. K. Balachandran, and M. J. Prather (1990), Climate change and the middle atmosphere. Part I: The doubled CO2 climate, *Atmos. Sci.*, *47*(4), 475-494, doi:10.1175/1520-0469(1990)047<0475:CCATMA>2.0.CO;2.

Roble, R. G., and R. E. Dickinson (1989), How will changes in carbon dioxide and methane modify the mean structure of the mesosphere and thermosphere?, *Geophys. Res. Lett.*, *16*(12), 1441–1444.

Rodgers, J. L., W. A. Nicewander, and L. Toothaker (1984), Linearly independent, orthogonal, and uncorrelated variables, *The American Statistician*, *38*(2), 133-134.

Shalizi, C. (2010), The Bootstrap, *American Scientist*, *98*(3), 186, doi:10.1511/2010.84.186.

Tiao, G. C., G. C. Reinsel, D. Xu, J. H. Pedrick, X. Zhu, A. J. Miller, J. J. DeLuisi, C. L. Mateer, and D. J. Wuebbles (1990), Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation, *J. Geophys. Res.*, *95*(D12), 20507-20517, doi:10.1029/JD095iD12p20507.

Weatherhead, E. C., et al. (1998), Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res.*, *103*(D14), 17149–17161.

Wickwar, V., T. Wilkerson, M. Hammond, and J. Herron (2001), Mesospheric temperature observations at the USU/CASS Atmospheric Lidar Observatory (ALO), in *Proceedings of SPIE*, pp. 272-284, Sendai, Japan. [online] Available from: http://link.aip.org/link/?PSI/4153/272/1&Agg=doi.

CHAPTER 3

SOLAR CYCLE VARIABILITY:

AMPLITUDE AND PHASE ANGLE

Abstract. Determining the atmospheric temperature response to solar input is typically done by fitting a linear model to time-series temperature data using a least squares approximation. These models typically include a solar proxy that follows the 11-year solar UV intensity variation and a linear term to estimate the cooling rate. One difficulty encountered by researchers is separating the atmospheric solar response from solar-like variations of decadal timescale; the atmospheric solar-like response could be out of phase with the solar-like input. If so, and a fixed-phase solar proxy is employed, the phase difference between the solar input and the atmospheric solar response can significantly bias the linear regression coefficient and attenuate the solar proxy coefficient. The nature of the bias and attenuation is investigated. Also, sine and cosine solar-like terms were fitted to USU lidar data, as well as the stratopause ERA and CPC temperatures, to check for signals of a decadal timescale that are out of phase with the solar input. The ERA and CPC data are nominally from the location of the USU Rayleigh lidar (41.74ºN, 111.81ºW). The sine and cosine terms describing the atmospheric solar response were found to be statistically significant at some altitudes but not others. The phase difference between the solar input and the solar-like atmospheric response can vary from 0 to 2π rad. The magnitude of the solar-like response (max – min) varies from 3.5 K at 45 km to 0.5 K between 50 and 60 km and, then, from 60 to 90 km it steadily increases to around 4.5 K (max – min). Also, a solar-noise term was included in the least squares model and

was found to be of statistical significance in the lower and upper mesosphere (45-54 km and 75-87 km respectively), but not the middle mesosphere.

1. Introduction

The solar electromagnetic flux follows an approximate 11-year intensity variation. The solar ultraviolet output in the near and middle UV is of particular interest because of its significant impact on stratosphere and mesosphere temperature structure. While overall solar intensity varies less than 1%, the shorter UV spectrum varies from 5% at 205 nm increasing to 50% in the Lyman-α line [*Donnelly et al.*, 1982; *Donnelly*, 1991]. The shorter wavelengths less than 300 nm are nearly completely absorbed by the Earth's middle and upper atmosphere [*Rottman*, 1988]. This large variation in short wave radiation affects photochemical ozone production and can alter middle atmosphere thermal characteristics which can, in turn, alter the propagation of planetary waves and global circulation patterns affecting heat advection and chemical transport [*Calisesi and Matthes*, 2007]. Given the scale and influence of these mechanisms it is important to determine how the atmosphere responds to variations in solar input.

Various methods have been employed to determine how the atmosphere responds to variations in solar input. The most direct method involves looking for elevated temperatures at solar maximum and depressed temperatures at solar minimum [*McCormack and Hood*, 1996; *Chanin*, 2006]. This direct inspection method makes no assumptions about how the atmosphere is responding to solar input and does not have the problems that sometimes accompany the least squares modeling approach, model misspecification and collinearity. But this method lacks the precision of a least squares

analysis. Two other methods employ a solar UV proxy: (1) deseasonalized temperatures can be checked for correlation with a fixed-phase solar proxy, and (2) a least squares model containing a solar proxy can be fit to time series data. The first cannot determine the existence of a phase offset between solar input and any decadal scale solar-like atmospheric responses. If atmospheric, the solar-like response is 90^o out of phase with the solar input, the correlation between the deseasonalized temperatures, and the solar proxy will be approximately zero. The second method suffers from a similar problem. By including a fixed-phase solar proxy in a least squares model, an implicit assumption is made. The atmosphere responds only in phase or out of phase to the solar input. However, if the atmospheric response differs from solar input by a phase of 90^o, then the solar proxy coefficient will be approximately zero.

There are several choices of solar proxy: Sunspot number, F10.7 flux, He I 1083 nm line, Mg II core-to-wing ratio, and Plage index. (See *Lean* [1991] for a short list.) Based on the findings of *Keckhut et al.* [1995] and our own experience, we have employed the Mg II proxy in this analysis. The daily Mg II data we employ are from the NOAA Space Weather Prediction Center website. Missing data points were interpolated and the data were smoothed using an 81 day boxcar average to suppress oscillations of periods less than 81 days.

2. Section Summary

In section 3, the issues of model specification are addressed. It is shown that when a decadal scale variation is out of phase with the solar input and a fixed-phase solar proxy is used in a least squares model, the phase difference can lead to significant bias in the

linear trend coefficient and attenuate the amplitude of the fixed-phase proxy coefficient. However, in cases where a fixed-phase proxy is used, if the proxy is sine-like (as opposed to cosine-like) the solar proxy will not bias the linear terms, regardless of the phase of the atmospheric response. Other cases where the linear term is unbiased by the solar term is where the atmosphere is responding directly in phase or out of phase to the solar input. In cases where a fixed-phase proxy is used, this would most likely occur where the absolute value of the solar proxy coefficient is at a maximum value. In section 4, evidence is presented for the existence of a decadal scale atmospheric temperature perturbation with a variable phase to the solar input. It is then argued that an independent solar-noise term should be included in a least squares mesosphere temperature model. Various models are also applied to USU lidar data. In section 5, this is repeated for the stratopause CPC and ERA temperature data and the model coefficients are checked for statistical significance. A Mallow's Cp test is employed to determine if the solar-like terms and the solar-noise terms are present in any of the favored sub-models (models made from subsets of the full model terms). Section 6 contains a discussion of the results from section 4 and a short discussion of the amplitude and phase profiles found in the USU data, as well as in situ variations in the atmospheric solar response found by other researchers. Section 7 contains a short discussion of possible causes of a phase lag between atmospheric solar response and solar input. Section 8 contains a summary and conclusions.

3. Derivation of Bias and Attenuation

If a model is not correctly specified, then problems of coefficient bias and attenuation can arise, as well as collinearity. While both are important linear regression issues, the

collinearity problem will not be addressed here: It has been addressed more fully in Chapter 2. We instead focus on model specification. There are three kinds of specification applicable to linear models. A model may be over specified, under specified, or cross specified. If over specified the proposed model has all true model regressors, and additional terms. The correctly specified terms will, on average, have unbiased coefficients, and the additional terms will have zero coefficient values. If the model is under specified, that is, if the proposed model has some but not all true model regressors and no others, then the coefficients of the correctly specified terms will be biased. If the model is cross specified, that is, if the proposed model contains some but not all true model regressors and additional regressors, the true model terms will have biased coefficients and the extra terms will have coefficients that are functions of the true model terms and extra terms. Examples of this follow. For our purposes cross specification is of primary interest.

For an over specified model, if the true model is $E(\boldsymbol{Y}) = \alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2$ and the proposed model is $E(\boldsymbol{Y}) = \alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \alpha_3\boldsymbol{x}_3$, then the correctly specified terms have unbiased coefficients and the extra $\boldsymbol{x}_3$ term has a zero coefficient value, that is, the expected coefficient values are $E(a_1) = \alpha_1$, $E(a_2) = \alpha_2$, $E(a_3) = 0$. For an under specified model, if the true model is $E(\boldsymbol{Y}) = \alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \alpha_3\boldsymbol{x}_3$ and the proposed model is $E(\boldsymbol{Y}) = \alpha_1\boldsymbol{x}_1 + \alpha_3\boldsymbol{x}_3$ the expected coefficient values $E(a_1)$ and $E(a_3)$ will be biased. These results are well known and the derivations are given in Appendix B.

For cross specification, if the true model is $E(\boldsymbol{Y}) = \alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2$ and the proposed model is $E(\boldsymbol{Y}) = \alpha_1\boldsymbol{x}_1 + \alpha_3\boldsymbol{x}_3$, the expected coefficient values $E(a_1)$ and $E(a_3)$ are given by

$$E(a_1) = \alpha_1 + \alpha_2[(\boldsymbol{x}_1{}^T \boldsymbol{x}_2)(\boldsymbol{x}_3{}^T \boldsymbol{x}_3) - (\boldsymbol{x}_1{}^T \boldsymbol{x}_3)(\boldsymbol{x}_2{}^T \boldsymbol{x}_3)]/\Delta \ ,$$

$$E(a_3) = \alpha_2 [(\boldsymbol{x}_1{}^T \boldsymbol{x}_1)(\boldsymbol{x}_2{}^T \boldsymbol{x}_3) - (\boldsymbol{x}_1{}^T \boldsymbol{x}_2)(\boldsymbol{x}_1{}^T \boldsymbol{x}_3)]/\Delta \ ,$$

where Δ is the determinant of the $\boldsymbol{X}^T\boldsymbol{X}$ matrix, where $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_3]$. (See Appendix B for derivation.) Notice the coefficient value for a_3 is a function of the true model regressors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and the extra regressor $\boldsymbol{x}_3$. This result will now be applied to a standard least squares middle-atmosphere temperature model.

A least squares model for middle atmosphere temperatures typically looks something like this,

$$T_{ij} = I_j + \beta_j t_i + \gamma_{1j}\cos(2\pi \cdot t_i) + \gamma_{2j}\sin(2\pi \cdot t_i) + \eta_{1j}\cos(4\pi \cdot t_i) + \eta_{2j}\sin(4\pi \cdot t_i) + \alpha_j SP_i + \varepsilon_{ij}.$$

The subscripts i and j are the time and altitude indices, respectively. T is a time series of atmospheric temperatures, t is the time of each measurement, I is the intercept, SP is the fixed-phase solar proxy, and ε is the residual; β is the linear cooling rate and α is the magnitude of the atmospheric solar response. The other four terms comprise the annual and semiannual oscillation. Because the correlation between the annual and semiannual terms is quite small, omitting them will not significantly affect the other regression coefficients. Consequently, both terms may be omitted from the following analysis, permitting consideration of a simplified model,

$$\boldsymbol{T}_j = \beta_j \cdot \boldsymbol{t} + \alpha_j \cdot \boldsymbol{SP} + \boldsymbol{\varepsilon}_j \ . \qquad (3.1)$$

This model is mean centered so the intercept term is omitted—mean centering does not affect the coefficient values. Let Equation (3.1) represent the proposed model, or applied

model (the model that will be applied to the temperature data). From here the subscripts will be dropped for convenience. Now suppose the correct model (the model we should have applied to the temperatures) is

$$\boldsymbol{T} = \beta \cdot \boldsymbol{t} + \alpha \cdot sin(\omega \boldsymbol{t} + \varphi) + \boldsymbol{\varepsilon} \tag{3.2a}$$

which may also be written as

$$\mathrm{E}(\boldsymbol{T}) = \beta \cdot \boldsymbol{t} + \alpha \cdot \boldsymbol{SLR} \,, \tag{3.2b}$$

where φ is the phase of the atmospheric solar response, β is the linear trend coefficient, α is the amplitude of the atmospheric solar response, ε is the residual, $\boldsymbol{SLR} = sin(\omega \boldsymbol{t} + \varphi)$, and $\mathrm{E}(\boldsymbol{T})$ are the expected values for the model. This model contains a decadal scale solar-like $sin(\omega \boldsymbol{t} + \varphi)$ term with phase offset φ and frequency $\omega = 2\pi/(11 \text{ yr})$. (Though the solar cycle can vary from 10 to 12 years, for our purposes we assume an 11-year solar cycle. This will not impact the conclusions.) The $\boldsymbol{SLR} = sin(\omega \boldsymbol{t} + \varphi)$ is shorthand for $SLR_1 = sin(\omega t_1 + \varphi)$, $SLR_2 = sin(\omega t_2 + \varphi)$, etc., where ω and φ are not time varying. Since the phase offset can be measured from any time, the time center of the data set is selected for convenience. This also applies to the proposed model. Because data acquisition may begin at any point during the solar cycle the solar proxy has a phase which is also measured from the time center of the data set. Writing the solar proxy with a phase offset as we did with the $\boldsymbol{SLR}$ term we get $\boldsymbol{SP} = sin(\omega \boldsymbol{t} + \theta)$, where θ is the phase of the solar proxy, also measured from the time center of the data set. The applied model, represented by Equation (3.1), becomes

$$\boldsymbol{T} = \beta \cdot \boldsymbol{t} + \alpha \cdot sin(\omega \boldsymbol{t} + \theta) + \boldsymbol{\varepsilon}, \tag{3.3a}$$

which may be written as

$$\mathrm{E}(\boldsymbol{T}) = \beta \cdot \boldsymbol{t} + \alpha \cdot \boldsymbol{SP} \, . \tag{3.3b}$$

It needs to be pointed out that in both the solar-like ***SLR*** term and the solar proxy ***SP*** term ω, φ, and θ are not determined by the regression, their values can be thought of as predetermined; only the β and α terms are solved for in the regression. In standard normal form the expected values of the Model (3b) coefficients is given by

$$\mathrm{E}(\boldsymbol{g}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \mathrm{E}(\boldsymbol{T}),$$

where $\boldsymbol{X} = [\boldsymbol{t}, \boldsymbol{SP}]$ is the data space, E(***T***) are the expected values of the time series of atmospheric temperatures, and $\boldsymbol{g} = [b, a]^T$. E(***g***) are the expected values for the model coefficients and the $\boldsymbol{X}^T\boldsymbol{X}$ matrix is n×n. If there is no coefficient bias or attenuation then we expect that E(b) = β and E(a) = α, where b is the regression estimate for the true linear trend β and a is regression estimate for the true atmospheric solar response amplitude α. Substituting Equation (3.3b) into E(***T***) in the above equation and solving for E(b) and E(a) we get

$$\mathrm{E}(b) = \beta + \alpha\{ \, (\boldsymbol{SP}^T\boldsymbol{SP})(\boldsymbol{SLR}^T\boldsymbol{t}) - (\boldsymbol{SP}^T\boldsymbol{t})(\boldsymbol{SP}^T\boldsymbol{SLR})\} \, / \, \Delta, \tag{3.4}$$

$$\mathrm{E}(a) = \alpha\{ \, (\boldsymbol{t}^T\boldsymbol{t})(\boldsymbol{SP}^T\boldsymbol{SLR}) - (\boldsymbol{SP}^T\boldsymbol{t})(\boldsymbol{SLR}^T\boldsymbol{t})\} \, / \, \Delta, \tag{3.5}$$

where $\boldsymbol{t}^T\boldsymbol{t}$, $\boldsymbol{SP}^T\boldsymbol{SLR}$, $\boldsymbol{SP}^T\boldsymbol{t}$, etc, are inner products, and Δ is the determinant of $\boldsymbol{X}^T\boldsymbol{X}$. These

equations indicate a biased linear trend term and an attenuated solar response amplitude. For reasons that will be made clear later, the linear term is unbiased if the solar input is sine-like (as opposed to cosine-like). (Figure 3.1.)

The expected values for Equations (3.2a) and (3.3a) are

$$E(\boldsymbol{T}) = \beta \cdot \boldsymbol{t} + \alpha \cdot sin(\omega \boldsymbol{t} + \varphi), \tag{3.6}$$

$$E(\boldsymbol{T}) = \beta \cdot \boldsymbol{t} + \alpha \cdot sin(\omega \boldsymbol{t} + \theta). \tag{3.7}$$

Model (3.6) is the true model (the model we should have used) with a solar like response $sin(\omega \boldsymbol{t} + \varphi)$; Model (3.7) is the applied model with solar proxy $sin(\omega \boldsymbol{t} + \theta)$. The parameter θ is determined by the phase of the solar proxy and φ is determined by the way the atmosphere responds to solar input and can vary with altitude. The $\varphi - \theta$ is the phase difference between the atmospheric solar response φ and the solar input θ.

To further develop Equations (3.4) and (3.5) the inner products must be evaluated. The inner products are summations, which can be approximated as integrals. Because we are assuming mean centered time, $(t_{max} - t_{min})/2$, the integrals are evaluated from $-t_0$ to $+t_0$, where t_0 is the maximum time of the time-centered, time regressor; if t_{max} = 12 years and t_{min} = 2 years then t_0 = 5 years. For example, evaluating $\boldsymbol{t}^T\boldsymbol{t}$ gives $\delta t \cdot \sum t_i t_i \sim \int t^2\, dt$. Integrating from $-t_0$ to $+t_0$ we get $2/3\ t_0^3$, the δt term is an analogue to dt and in Equations (3.4) and (3.5) divides out. Evaluating the other inner-product terms in a similar manner gives the following expressions for E(b) and E(a).

$$E(b) = \beta + \alpha\{[t_0 - s_2 cos(2\theta)/2\omega - 2s_1^2 sin^2\theta/t_0\omega^2][2 \cdot cos(\varphi)(s_1 - \omega t_0 c_1)/\omega^2] -$$

$$[t_0 cos(\varphi - \theta) - s_2 cos(\varphi + \theta)/2\omega - 2s_1^2 sin(\varphi) sin(\theta)/t_0\omega^2]$$

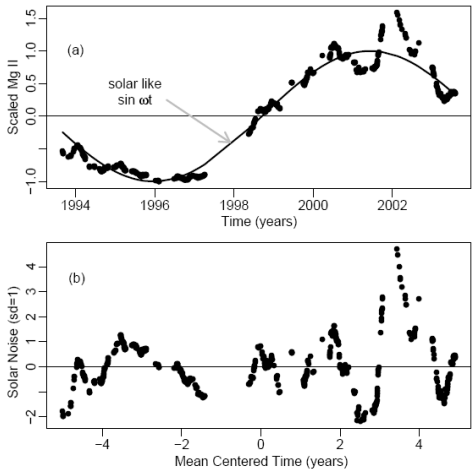


Figure 3.1. MgII proxy and solar noise: (a) shows the mean centered, smoothed, scaled MgII index covering the time span of the USU data set; (b) shows the solar-noise term obtained by subtracting sin ωt from the Mg II data. Only the points for the times for which we have data are shown. Plot (b) is scaled such that the standard deviation is equal to 1.

$$[2\cdot cos(\theta)(s_1 - \omega t_0 c_1)/\omega^2]\}/\delta \; . \tag{3.8}$$

$$E(a) = \alpha\{\, [2t_0^3/3][\, t_0 cos(\varphi - \theta) - s_2 cos(\varphi + \theta)/2\omega - 2s_1^2 sin(\varphi) sin(\theta)/t_0\omega^2] - cos(\theta)\cdot cos(\varphi)\cdot[2(s_1 - \omega t_0 c_1)/\omega^2]^2\}/\delta \quad , \tag{3.9}$$

where $\delta = (\boldsymbol{t}^T\boldsymbol{t})(\boldsymbol{S}\boldsymbol{P}^T\boldsymbol{S}\boldsymbol{P}) - (\boldsymbol{S}\boldsymbol{P}^T\boldsymbol{t})^2 = \{\, [2t_0^3/3]\cdot[t_0 - s_2 cos(2\theta)/2\omega - 2s_1^2 sin^2\theta/t_0\omega^2] - [2\cdot cos(\theta)(\, s_1 - \omega t_0 c_1)/\omega^2]^2 \,\}$, $s_1 = sin\ \omega t$, $c_1 = cos\ \omega t$, and $s_2 = sin\ 2\omega t$. E(a) is the expected value for the atmospheric solar response coefficient and E(b) the expected value for the linear trend coefficient. (A more detailed derivation is given in Appendix G.)

To test the accuracy of these equations, a least squares simulation was done with data generated from the true model given by (3.2a) with the regression model given by (3.3a). The quantities a/α and $b - \beta$ were compared to the attenuation and bias predicted by Equations (3.8) and (3.9). That is, regressions were done with known values of α and β with combinations of phase angles of φ and θ; the error of ($b - \beta$) and (a/α) was checked

against the bias and attenuation predicted by Equations (3.8) and (3.9). The angles θ and φ were varied from 0 to 2π in steps of 0.017 rad (~1º), and every combination of θ and φ were checked. The amplitude of the atmospheric solar response was α = 1 K and the linear trend was $\beta = -0.4$ K/year. It was found that the maximum absolute value difference of $(a - \alpha)$ and $(b - \beta)$ was on the order of 1×10^{-4} for both.

The next question is how well do Equations (3.8) and (3.9) predict the attenuation and bias for the solar response and linear trend coefficients when an annual oscillation (AO) and semiannual oscillation (SAO) are included in the model. Simulated temperature data was generated as mentioned above but with an AO and SAO included. The regression model also contained an AO and SAO. The angles θ and φ were varied from 0 to 2π in steps of 0.017 rad (~1º), and every combination of θ and φ were checked. The amplitude and linear trend were left at α = 1 K and $\beta = -0.4$ K/year, respectively. In this case the maximum absolute value difference of $(a - \alpha)$ and $(b - \beta)$ were 0.008 K and 0.003 K/year, respectively. These are larger than those in the previous paragraph because of the inclusion of the AO and SAO oscillations in the test signal. Because the AO and SAO oscillations cover several periods over the course of a solar cycle their amplitudes and phases will not affect the linear trend and solar response coefficients. Consequently we did not vary the amplitudes and phases of the AO and SAO.

To check for more realistic conditions the above simulation was repeated, this time replicating the number of data points and data gaps in the USU data set. The maximum absolute value difference between $(a - \alpha)$ and $(b - \beta)$ were 0.39 K and 0.075 K/year, respectively. As these results would apply to the USU data, the maximum error between the bias and attenuation predicted by Equations (3.8) and (3.9) and the fixed-phase proxy

amplitude and the linear trend is about 0.39 K and 0.075 K/year, respectively. These are larger than the results from the previous two paragraphs because of the presence of data gaps in the signal simulation.

Neither these results nor Equations (3.8) or (3.9) can be used to make corrections to any of the linear regression coefficients reported in the literature because the phase and amplitude of the atmospheric response are unknown when a fixed-phase solar proxy is employed in the regression model.

What follows is an explanation of some special cases of interest. If the phase of the solar proxy is $\theta = 0$ radians, that is, where the solar proxy is sine-like, then $E(b) = \beta$ and $E(a) = \alpha \cdot cos(\varphi)$. If $\theta = \pi$, then $E(b) = \beta$ and $E(a) = -\alpha \cdot cos(\varphi)$. Also, with the addition of more data the bias approaches zero but $E(a)$ approaches $\alpha \cdot cos(\varphi - \theta)$. Another interesting case is where $\theta = \varphi$, or $\theta = \varphi \pm \pi$, that is, where the atmospheric solar response is in phase or out of phase with the solar proxy. In the case of $\theta = \varphi$, $E(b) = \beta$ and $E(a) = \alpha$. In the case of $\theta = \varphi \pm \pi$, $E(b) = \beta$ and $E(a) = -\alpha$.

The attenuation of the true atmospheric solar response amplitude cannot be greater than one, so altitudes where the fixed-proxy coefficient is at a maximum or minimum value seem more likely to be regions where the atmosphere is directly in phase or out of phase to the solar input. Given this special case, in models where a fixed-proxy regressor is employed, altitudes where the fixed-proxy coefficient is at a maximum or minimum value are ideal for identifying the maximum of the atmospheric solar response and the magnitude of the linear trend coefficient. In these regions the bias is likely to be smaller and the proxy coefficient is more likely to represent the true atmospheric solar response amplitude. In between those levels the magnitude of the fixed-proxy coefficient is

determined principally by the amplitude and phase of the decadal scale solar-like atmospheric signal and the solar proxy phase. Naturally, one cannot be sure a maximum or minimum fixed-proxy coefficient value exactly coincides with an in phase or out of phase atmospheric solar response. But it seems unlikely the atmospheric solar response could affect a fixed proxy coefficient significantly more than the phase difference ($\varphi - \theta$) does because the proxy coefficient cannot go negative without a phase offset.

Figure 3.2 shows the maximum absolute-value of the bias as a function of the length of the data set as calculated from Equation (3.8) for the case where the atmospheric solar response amplitude α is 1 K. The linear trend bias drops off rapidly with the addition of more data and has a maximum absolute value of 0.96 K/year at 5 years (t_0 = 2.5 y), 0.24 K/year at 10 years (t_0 = 5 y), and 0.13 K/year at 12 years (t_0 = 6 y). (The parameter t_0 is the maximum value of the time centered time regressor. In other words if t_0 = 2.5 years then the time span of the data is from −2.5 to +2.5 years, or 5 years long.) Given that the expected middle atmosphere cooling is in the neighborhood of 0.1 K/year, this amount is not insignificant. (See section 3 in Chapter 1 for references.) To calculate the linear trend bias for different atmospheric solar responses, multiply the appropriate value from Figure 3.2 by α. For example, if the amplitude of the atmospheric solar response is 2 K then the maximum absolute value of the bias becomes 1.92 K/year at 5 years, 0.48 K/year at 10 years, and 0.26 K/year at 12 years. Also, the bias goes to zero at 15.6 years and again at 26.8 years (t_0 = 7.812 and 13.43 years, respectively). This occurs because for t_0 values of 7.812 and 13.43 years, s_1, defined as $sin\ \omega t$, equals $\omega t_0 c_1$, where $c_1 = cos\ \omega t$. That is, $sin(\omega t) = \omega t_0 cos(\omega t)$.

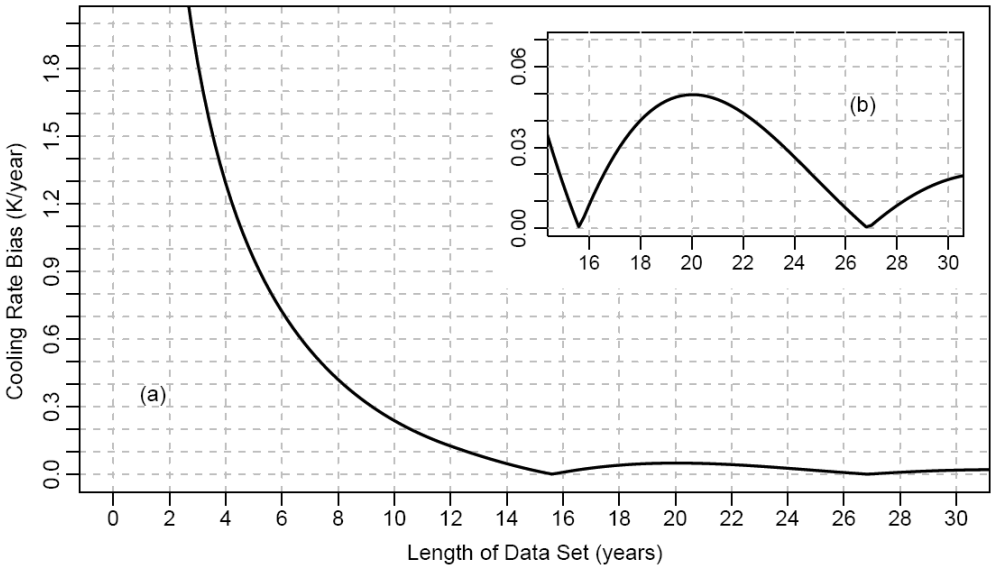


Figure 3.2. Maximum absolute value of the bias calculated from Equation (3.8); (b) shows a magnified view of the 15-30 year range. The amplitude of the atmospheric solar response α is 1 K. To get bias for different solar response amplitudes multiply by the amplitude of the atmospheric solar response. If α = 2 K then everything is doubled.

4. Applied to USU Data

In this section we look at the amplitude and phase of the atmospheric solar response found in the USU temperature data. The USU data consists of 593 nightly observations from September 1993 to August 2003 ranging from 45 to 90 km in altitude. The following model was fit to the USU temperature time series data,

$$T_i = IO_i + C_1 sin(\omega t_i) + C_2 cos(\omega t_i) + \varepsilon_i, \tag{3.10}$$

where IO_i is a placeholder for $I_i + \beta \cdot t_i + \gamma_1 sin(2\pi \cdot t_i) + \gamma_2 cos(2\pi \cdot t_i) + \eta_3 sin(4\pi \cdot t_i) + \eta_4 cos(4\pi \cdot t_i)$, I is the intercept, β is the linear trend coefficient, γ_1 and γ_2 the coefficients for the annual oscillation, η_3 and η_4 the coefficients for the semi-annual oscillation, and t is the time. The subscript i is the time index. From C_1 and C_2 one can calculate the

amplitude and phase of the solar-like response; $\omega = 2\pi/11\ y^{-1}$. It was found that at some altitudes only the sine term C_1 was statistically significant at greater than 95%. At others only the C_2 cosine term was significant at greater than 95%. And there were altitudes where both terms were significant at greater than 95%. The *sin* ωt and *cos* ωt are 90º out of phase and when this phase difference is expressed in terms of the 11-year solar cycle, they are about 2.75 years out of phase. Consistent with the USU findings, it has been found by several researchers that the solar proxy coefficient for a fixed solar proxy can be negative at some altitudes and positive at others, indicating a phase offset of 180º, or 5.5 years. This means that the phase difference between the solar input and atmospheric solar-like response can be 0º, 90º, 180º, 270º, or some other value.

To further test for the presence of an out-of-phase solar-like response, the following test models were applied to the USU data:

$$\mathrm{E}(\boldsymbol{T}) = \boldsymbol{IO} + \alpha_1 \cdot \boldsymbol{MgII}. \tag{3.11}$$

$$\mathrm{E}(\boldsymbol{T}) = \boldsymbol{IO} + \alpha_2 \cdot sin(\omega \cdot \boldsymbol{t}) \ . \tag{3.12}$$

$$\mathrm{E}(\boldsymbol{T}) = \boldsymbol{IO} + \alpha_3 \cdot sin(\omega \cdot \boldsymbol{t} + \varphi). \tag{3.13}$$

IO is as mentioned above; the Mg II data was scaled to fit *sin* ωt (Figure 3.1), ω is the frequency of the solar cycle. There was some difficulty getting a non-linear least squares technique to converge using Model (3.13). We worked around this problem by first fitting $\mathrm{E}(\boldsymbol{T}) = \boldsymbol{IO} + C_1 \cdot sin(\omega \boldsymbol{t}) + C_2 \cdot cos(\omega \boldsymbol{t})$ and calculating the phase $\varphi = atan(C_1/C_2)$ then refitting Model (3.13) with a fixed phase in $sin(\omega \boldsymbol{t} + \varphi)$ and solving for α_3 using ordinary least squares. Our main interest is in the magnitude of α_3 and its p-value, which is a measure of the statistical significance of α_3. We found the differences between α_3 and

$\sqrt{(C_1^2 + C_2^2)}$ to be minute. Model (3.11) is a typical model for middle atmosphere temperatures. Model (3.12) substitutes a sin ωt for MgII. These two equations assume the atmosphere response either in phase or out of phase to the solar input. Model (3.13) assumes the atmosphere can respond with any phase to the solar input. Models (3.11)-(3.13) were fit to the USU time series temperature for every altitude bin. The confidence levels of α_1, α_2, and α_3 are shown in Figure 3.3a. The α_3 term, corresponding to $sin(\omega t + \varphi)$, is statistically significant over a greater altitude range than the other solar related terms, and with the exception of a region between 50-53 km none of the other solar-like terms have greater statistical significance than α_3. One thing that stood out was the similarity between the confidence levels of α_1 and α_2 (corresponding to *MgII* and $sin\ \omega t$), between 65 and 75 km.

The only significant difference between the $sin\ \omega t$ and *MgII* terms is the short term noise variations, suggesting the solar-noise, or short term solar variations, might be of physical importance at some altitudes but not others. With this in mind, a separate solar-noise term was added to Model (3.13). This term is simply the residuals obtained from fitting $sin(\omega t + \theta)$ to the solar proxy (Figure 3.1a,b). The addition of a solar-noise term in the model did not alter the other coefficient values to any significant degree and SEs were only marginally affected. (See Chapter 4 for coefficient correlations.) As can be seen from Figure 3.3b, between 45 and 53 km, and between 76 and 82 km, the solar-noise term is statistically significant at or above the 95% confidence level. In fact, from 45 to 50 km the solar-noise term had a confidence level greater than 99%. Between 52 and 76 km the solar-noise term has very low significance. Between 65 and 73 km, where α_1 and α_2 have high statistical significance, the solar-noise term is statistically insignificant.

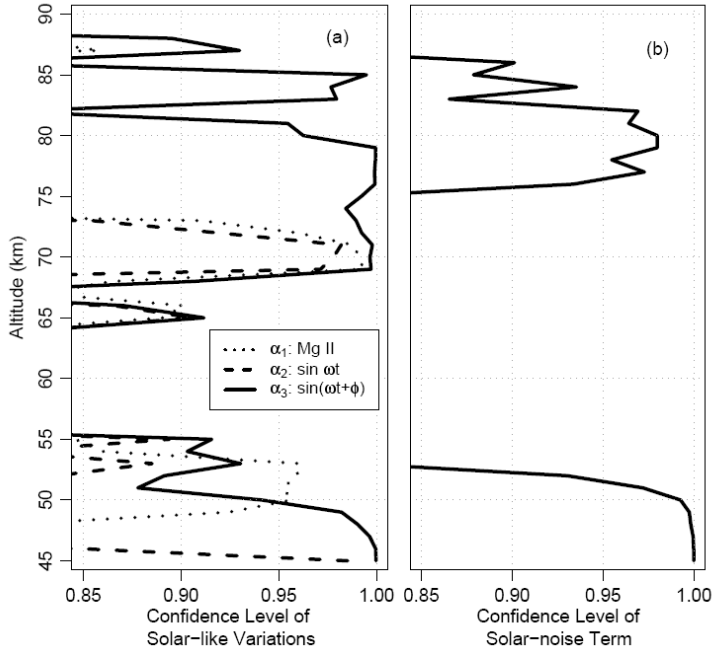


Figure 3.3. Confidence levels for regression terms: (a) shows the confidence levels for the *MgII*, *sin*(ωt) and *sin*($\omega t + \varphi$) terms. Only the 0.85 level and greater are shown; (b) shows confidence levels for the solar-noise term.

The differing significance of the solar-noise term and the *sin*($\omega t + \varphi$) term possibly points to differing physical mechanisms driving the middle atmosphere temperature response to solar input. The solar-noise could be taken as a proxy indicative of how the atmosphere responds to short-term solar variations and the solar-like term indicative of longer period, i.e. 11-year, dynamical effects.

5. Applied to CPC and ERA Data

To further test the applicability of both a solar-like atmospheric response and a solar-noise term, the following two models were applied to CPC and ERA temperatures. The CPC temperatures are from 2 and 1 hPa (45 and 48 km), and the ERA temperatures are from 3, 2, and 1 hPa (41, 45, and 48 km). These geopotential altitudes are close to stratopause and at the bottom of our altitude range. The ERA data is from the European

Centre for Medium-Range Weather Forecasts and the CPC data is from the NOAA Climate Prediction Center. The ERA and CPC data are complied and calculated from radiosonde, balloon, aircraft, and satellite measurements and the latitude/longitude data is calculated using forecasting computer models. Both the CPC and ERA data is nominally from the location of our USU lidar. The two models applied to the ERA and CPC data are,

$$\boldsymbol{T} = \boldsymbol{IO} + C_1 \cdot sin(\omega \boldsymbol{t}) + C_2 \cdot cos(\omega \boldsymbol{t}) + D \cdot \boldsymbol{solnoise} + \boldsymbol{\varepsilon} \, . \tag{3.14}$$

$$\boldsymbol{T} = \boldsymbol{IO} + C \cdot \boldsymbol{MgII} + \boldsymbol{\varepsilon} \, . \tag{3.15}$$

Model (3.14) captures the variable phase relationship where $\alpha_3 = \sqrt{(C_1^2 + C_2^2)}$ and the phase $\varphi = atan(C_1/C_2)$. Again, $sin(\omega \boldsymbol{t}) = \{sin(\omega t_1), ...\}$ and $cos(\omega \boldsymbol{t}) = \{cos(\omega t_1), ...\}$. The data covers almost exactly the same time period as our USU data. One goodness-of-fit statistic is the Adjusted R-squared values. This is similar to the R-squared value, but is adjusted for the number of variables included in the model. Based on the Adjusted R-squared values, Model (3.14) provides an overall better fit to the CPC data at 1 hPa and to the ERA data at 3 hPa than did Model (3.15). For the other altitude levels, the Adjusted R-squared differences were minute. The statistical significance of the coefficients and Adjusted R-squared values for each data set are given in Table 3.1. The Adjusted R-squared values indicate the quality of the fit. An Adjusted R-squared value near 1 indicates that the model is a near perfect fit to the data. Also listed in Table 3.1 are the significance levels of the model terms. A significance level near 1 indicates that the regressor has high explanatory value in the model. A significance level of 0 indicates that

it has no explanatory value and should probably be omitted. Often researchers will include regressors that have a significance level of 0.8 or greater, but there is no strict cutoff rule to determine this.

Though most terms from the two models have high statistical significance when applied to the CPC data, for Model (3.14) both the sine and cosine solar-like terms, as well as the solar-noise term, have high significance at 1 hPa (48 km). At 2 hPa (42 km) the sine and solar-noise terms have high statistical significance but the cosine solar-like term is significant at the 74% level. For the ERA data the results are mixed. For Model (3.14) at 3 hPa, both solar-like terms have high statistical significance and the solar-noise term has questionable significance at 77%, whereas for Model (3.15) at the same altitude the MgII term is not significant, being at the 31.5% level. At 2 hPa, the solar-like terms in (3/14) have high to moderate statistical significance, whereas for Model (3.15) at that same altitude the MgII coefficient is significant at only the 21% level. At 1 hPa, the confidence level of the MgII coefficient in (3.15) is high, but in Model (3.14) the confidence level of the solar-like sin ωt term is low, at only 11.9%. It should also be pointed out that the Mg II data for the time period of these data sets is nearly sine-like,

Table 3.1. The statistical significance of the coefficients for the terms in Model (3.14) and (3.15) applied to the CPC and ERA data sets.

	CPC Data				ERA Data					
	1 hPa	2 hPa	1 hPa	2 hPa	3 hPa	2 hPa	1 hPa	3 hPa	2 hPa	1 hPa
Model	(15)	(15)	(14)	(14)	(15)	(15)	(15)	(14)	(14)	(14)
Intercept	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Linear Term	1.000	1.000	1.000	1.000	1.000	0.981	0.613	1.000	0.126	1.000
sin 2πt	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
cos 2πt	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
sin 4πt	1.000	1.000	1.000	1.000	0.309	1.000	1.000	0.233	1.000	1.000
cos 4πt	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MgII	1.000	1.000	*	*	0.315	0.205	1.000	*	*	*
sin ωt	*	*	1.000	1.000	*	*	*	1.000	0.972	0.119
cos ωt	*	*	1.000	0.744	*	*	*	1.000	0.830	1.000
solar-noise	*	*	0.998	0.986	*	*	*	0.773	0.331	0.996
Adj. R-squared	0.665	0.769	0.680	0.768	0.706	0.655	0.599	0.721	0.656	0.6028

which means it is not likely to bias the linear trend coefficient.

Though the sine-like Mg II proxy and the solar-like sin ωt term are unlikely to significantly bias the linear trend regressor, there will be significant collinearity between the MgII term (and the sin ωt term) with the linear trend, indicating correlation between the coefficients of those two terms. (For collinearity see Chapter 2.) The change in statistical significance between the linear term and the solar proxy (and solar-like term) in Model (3.14) and (3.15) at both 44 and 49.6 km for the ERA data is symptomatic of this problem. At 2 hPa, the statistical significance of the linear term is 98.1% for Model (3.15) but drops to 12.6% when Model (3.14) is applied to the data. Whereas the sine-like MgII term in Model (3.15) is low at 20.5%, the sin ωt term jumps to 97.2% when Model (3.14) is applied. There is a similar reversal occurring at the 1 hPa level: The MgII term in (3.15) is significant at 100% and the linear term is at 61.3%, but for Model (3.14) the sin ωt term is significant at 11.9% and the linear term jumps to 100%. There is not much that can be done about these reversals, except to point them out. However, the cosine term is significant at 80% or higher at all three altitude levels. Because the cos ωt term is orthogonal to the sin ωt term, as well as the time regressor, its coefficient is uncorrelated to either regressor's coefficients. Consequently the significance of the cos ωt term is unlikely to be a collinearity artifact.

To further investigate which model terms should be included in an OLSR applied to the CPC and ERA data, a Mallows' C_p test was conducted on sub-models of (3.14). The Mallow's C_p statistic is often used in variable selection methods to find suitable sub models from a collection of proposed explanatory variables. The danger in putting too many variables in a model is collinearity. Selection methods, such as the stepwise

method, test various combinations of these explanatory variables to find combinations that provide a good fit to the data. One tries to find a model with a C_p value nearly equal to p, which is the number of regressors in the model including the intercept. These models provide the least amount of bias in the coefficient estimates. When doing this test, it is possible to get ambiguities when different sub-models produce C_p values equally close to p, at which point individual experience must be applied [*Mallows*, 1973]. This method is used here to build a case for the inclusion of the solar-noise term, and the sin ωt and cos ωt terms, in the least squares model. Their presence in the preferred sub models is taken as evidence that they should be included in similar middle atmosphere temperature models. The results from these tests are shown in Table 3.2, with the xs indicating terms included in the preferred sub-model. It was found that the solar-noise term was included in all sub-models for the CPC and ERA data. For the CPC data the *cos ωt* and *sin ωt* were included at 1 hPa, but only the *sin ωt* at 2 hPa. For the ERA data both the *sin ωt* and *cos ωt* terms are included at 3 hPa. At 2 hPa only the sine term was included; at 1 hPa only the cosine term was included. Similar results were obtained when applying the Mallow's C_p test to Model (3.14) applied to USU data. The solar-noise term was present in the preferred model in 19 out of 46 altitude bins, both the sine and cosine terms were included in 8 bins; the sine term without a cosine term occurred in 10 bins

Table 3.2. The results of a Mallow's Cp test of the coefficients for the terms in (3.14) and (3.15) as applied to the CPC and ERA data sets.

	Altitude	Int.	time	sin 2πt	cos 2πt	sin 4πt	cos 4πt	sin ωt	cos ωt	solar noise	p	Cp
CPC	2 hPa	x	x	x	x	x	x	x		x	8	8.3
	1 hPa	x	x	x	x	x	x	x	x	x	9	9
ERA	3 hPa	x	x	x	x		x	x	x	x	9	7.1
	2 hPa	x		x	x	x	x	x		x	7	6.9
	1 hPa	x	x	x	x	x	x		x	x	8	7.5

and the cosine term without a sine term occurred in 13. This emphasizes the value of the solar-noise in the linear regression model.

6. Discussion of Results

Atmospheric solar-response amplitudes and phases were obtained by fitting $E(T) = \beta \cdot t + \alpha_1 \cdot sin(\omega t + \varphi)$, $E(T) = \beta \cdot t + \alpha_2 \cdot sin\ \omega t$, and $E(T) = \beta \cdot t + \alpha_3 \cdot MgII$ to the USU data after removing the annual and semiannual variations from the temperatures. The phase angle φ was determined as previously described in section 4. The αs are subscripted for convenience. The coefficient profiles for the αs and for φ are shown in Figure 3.4. (See

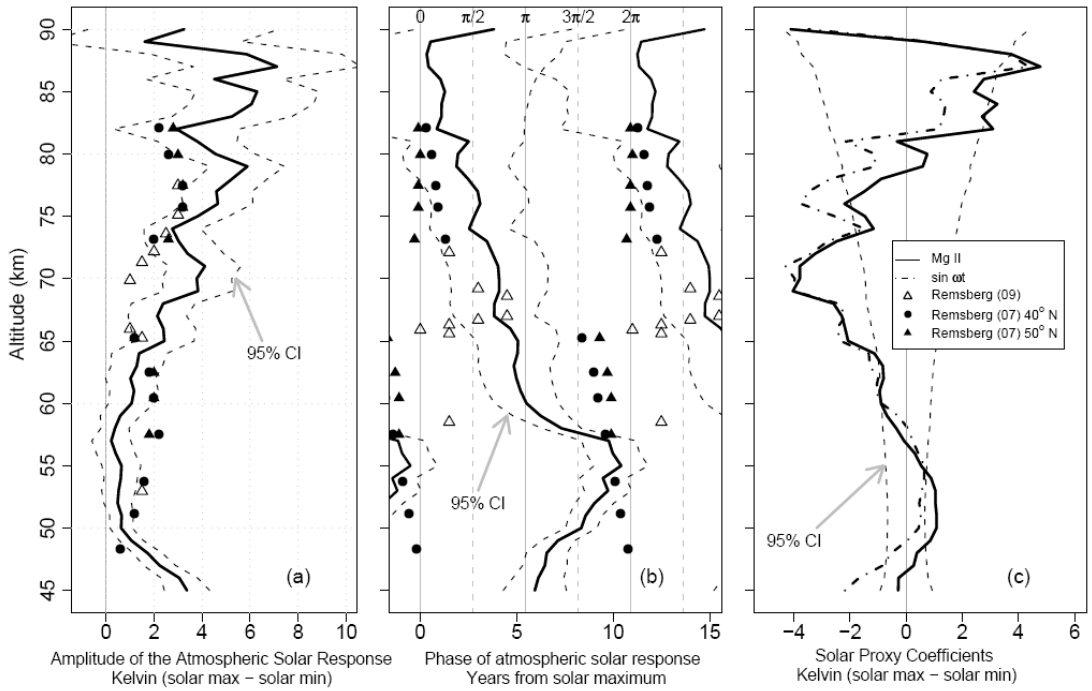


Figure 3.4. Solar amplitude and phase plots. Plot (a) is the magnitude (solar max – solar min) of the solar-like proxy α_2; (b) is the phase of the solar from solar maximum; (c) shows the magnitude (solar max – solar min) of the fixed-phase coefficients α_1 corresponding to *sin* ωt, and α_3 corresponding to the *MgII* proxy. The circles and triangles in (a) and (b) are amplitudes and phases from data 40ºN adapted from Table 10 and 11 in *Remsberg* [2007]; the solid triangles are the same but for 50ºN. The phase data is repeated every 11-years to aid in comparison.

Appendix C for the methodology on calculating the error limits.) One interesting feature is the rapid change in φ between 57 and 60 km. At 57 km it is in phase with the solar input and at 60 km it is nearly 180º out of phase. The fixed-proxy coefficients α_2 and α_3 also go from positive to negative at about 58 km. There are several ways zero α_2 and α_3 values could occur. The most likely are a near zero amplitude in α_1, an out of phase atmospheric response attenuating the amplitude of the true atmospheric solar response α, or a combination of both. The fixed-phase coefficient cannot go negative without a phase offset, which does exist at that altitude. The amplitude of the variable-phase solar proxy α_1 also approaches zero at 57 km, the magnitude being 0.4 K. So both causes are involved. A small solar amplitude in this region is consistent with findings from other researchers. *Kubicki et al.* [2008] reported an atmospheric temperature response transitioning from positive to negative at 59 km during winter and 52 km during summer. *Keckhut and Kodera* [1999] found a temperature change from positive to zero at 52 km for winter but a fairly uniform temperature response of 1 K from 30-55 km for summer. A similar sign reversal near 50 km was found by *Keckhut et al.* [1995], as well as *Cossart and Taubenheim* [1987]. *Chanin et al.* [1987] show deseasonalized temperatures from 1979 to 1985 from 40 to 65 km along with the F10.7 solar flux for that time period. At 40 km there is a clear negative response, at 50 km the temperature response is zero, and at 65 km it is positive. This result from *Chanin et al.* is noteworthy because the comparison does not rely on linear regression techniques, and thus the problem of coefficient attenuation is not an issue. Though the altitude of this zero-response feature varies geographically and with altitude, its existence as found by several researchers suggests a zero or near zero atmospheric temperature response to the solar

input in the upper stratosphere/lower mesosphere region.

In analyzing data from the Halogen Occultation Experiment (HALOE) *Remsberg* [2002] fit a variable-phase solar proxy to the data. They found a phase lag in the atmospheric temperatures of 2.3 years from solar maximum at 40ºN and 0.05 hPa (~ 70 km). They also report a lag of 1.9 and 1.5 years at 0.03 hPa and 0.02 hPa (~75 km and 77 km) respectively at the same latitude. At that time they were working with 9.5 years of data spanning from late 1991 to early 2001. In an updated paper, *Remsberg* [2007] analyzes 14 years of temperature data and reports phase lags of 1.3 years at 0.03 hPa and 0.9 years at 0.02 hPa, at 40ºN. These changes indicate the phase estimate is sensitive to the length of the data set. Significant phase lags are again reported in *Remsberg* [2009]. These are shown in Figure 3.5 for a range of latitudes. This most recent paper shows that the phase lags are confined mostly to the Northern Hemisphere middle atmosphere, which coincides with the location of the USU Rayleigh lidar. A comparison of our USU solar response amplitudes and phases with those taken from Table 10 of *Remsberg* [2007] are shown in Figures 3.4a and b. Above 65 km our solar amplitudes are about 1 K larger than those from Remsberg. Between 60 and 65 km their data matched to within our 95% error limits. There is also similarity around 50 km. However, between 50 and 60 km the data from Remsberg is warmer by about 1.5 K. Between 73 and 83 km there is agreement within our error limits, or nearly so. The difference in the height of the phase offset in our data may be due to local variations in atmospheric solar response, whereas the HALOE data is a zonal mean. Large zonal differences were produced in simulations by *Hampson et al.* [2005] who found zonal asymmetries in atmospheric solar response of up to 10 K at 49 km. There also exists the possibility that a Pinatubo effect could be influencing the

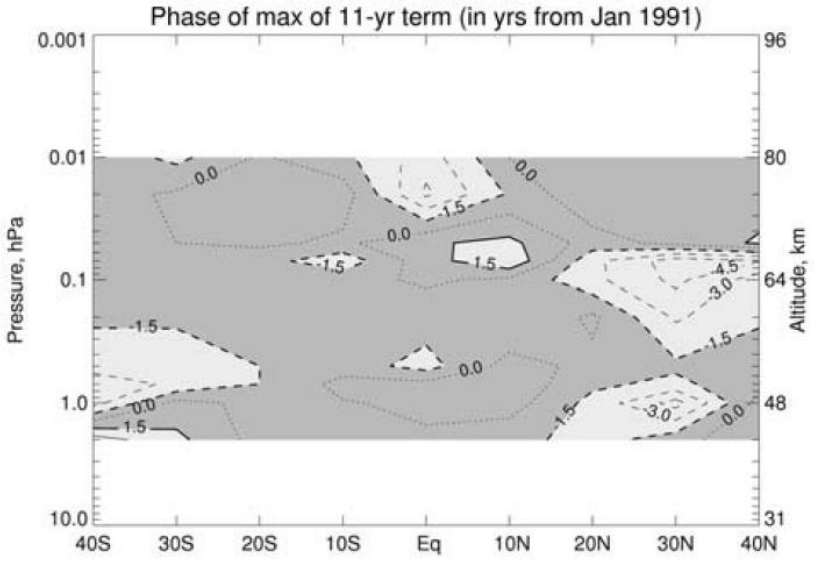


Figure 3.5. The phase of the solar-like term from the HALOE data. (Image is Figure 6 from *Remsberg* [2009]. Used with permission of American Geophysical Union.)

phase and amplitude of any decadal scale atmospheric signals in our data. According to *She et al.* [1998] a temperature perturbation on the order of 8 K occurred at 87 km in mid 1993, which is approximately when our data set begins, the temperature spike occurs during solar minimum. This, coupled with the shortness of our data set, decreases the reliability of our phase calculations. A similar concern was expressed by *Remsberg* [2002] and *Remsberg* [2007], where the results were considered exploratory. Owing to the shortness of our data set, the utility of our solar-like amplitude and phase results should likewise be considered exploratory. Also, above 80 km the error bars on our phase data are large enough that little can be said about the phases in that region.

As indicated in Figure 3.4c, there is a noticeable difference in the amplitude of the *MgII* coefficient and the solar-like *sin ωt* term in the upper and lower mesosphere. The Mg II data was scaled by fitting it to the sin ωt term, which makes the coefficients comparable. The regions of greatest difference coincide with the regions where the solar-

noise term has its greatest significance, from 45 to 53 km and from 75 to 87 km. In these regions the amplitudes of the solar-like *sin ωt* proxy is shifted to smaller values (to the left) by about 1 to 2 K. The Mg II data includes short-term solar variations, so one might be tempted to think that inclusion of a separate solar-noise term is redundant. So an additional test was employed to determine if this temperature shift was caused by the inclusion of the solar-noise term.

The models (a) $E(\boldsymbol{T}) = \beta \cdot \boldsymbol{t} + \alpha_2 \cdot sin(\omega \boldsymbol{t}) + SN \cdot \boldsymbol{solnoise}$, and (b) $E(\boldsymbol{T}) = \beta \cdot \boldsymbol{t} + \alpha_3 \cdot \boldsymbol{MgII} + SN \cdot \boldsymbol{solnoise}$ were fit to the data, the MgII proxy being scaled as already described. It was found that α_3 closely matched α_2. The average absolute value difference between α_3 and α_2 from 45 to 90 km was 0.15 K. The α_2 coefficient was little affected by the inclusion of the solar-noise term, having a maximum of difference of 0.1 K when the solar-noise was omitted. The fact that $\boldsymbol{y} = \beta \cdot \boldsymbol{t} + \alpha_2 \cdot sin(\omega \boldsymbol{t}) + SN \cdot \boldsymbol{solnoise}$ produced nearly identical αs as $\boldsymbol{y} = \beta \cdot \boldsymbol{t} + \alpha_3 \cdot \boldsymbol{MgII} + SN \cdot \boldsymbol{solnoise}$ suggests the solar-noise term is not redundant.

There are additional reasons for including the solar-noise as a separate term in the linear regression model. One reason is the decadal scale variation may be out of phase with the solar input, so one would want to include an out of phase solar-like variation in the model. The other reason is more compelling. The idea behind least squares is to project a vector of data onto a column space of explanatory variables in such a way as to minimize the residual sum of squares (RSS). A traditional solar proxy has both a noisy signal and a decadal signal. Which of the two reduces the residual sum of squares to a greater degree? To test this question a Monte Carlo simulation was done to see how influential the noise was in determining the confidence levels on a solar proxy coefficient. A times series of temperatures was constructed having the following

structure: $\boldsymbol{y} = 5 \cdot sin(\omega \boldsymbol{t}) + \boldsymbol{sn} + \boldsymbol{n}$, where ω is the frequency of the solar cycle, ***sn*** is generated from Gaussian noise with $\sigma = 1$ K, which is then autocorrelated with a correlation coefficient of 0.5, ***n*** is an additional Gaussian noise term added on top of this with $\sigma = 1$ K. To this data the following model was fit: $\boldsymbol{y} \sim \boldsymbol{sol}$, where ***y*** is the temperatures just mentioned and $\boldsymbol{sol} = cos(\omega \boldsymbol{t}) + \boldsymbol{sn}$. Note that ***sol*** has completely the wrong shape to it, but it does have the correct solar-noise (Figure 3.6a). It was found that the statistical significance of ***sol*** was on average 95.3%, even though it had completely the wrong shape. It should also be pointed out that the correlation between **y** and ***sol*** was 0.094 on average. This was also true when ***sol*** was changed to $cos(3\omega \boldsymbol{t}) + \boldsymbol{sn}$ (Figure 3.6b). The average correlation was about 0.095 and the statistical significance of ***sol*** was about 95.4%. It appears that if a signal has the wrong shape and the right noise there is a good chance it will appear to be statistically significant. The noise seems to have a much greater RSS reducing effect than longer time scale signals. When a solar proxy is included in a least squares model the coefficient of interest is the amplitude. If the noise in the solar proxy follows the noise in the atmospheric temperatures and the solar proxy has completely the wrong shape to it, there is a risk that the amplitude coefficient could be considered significant when, in fact, it is not. As it applies to atmospheric temperatures it seems safer to separate the solar-noise from the decadal-scale solar signal.

7. Cause of the Phase Lag

The cause of the lag between the sine-like solar input and sine-like atmospheric solar response is unknown. *Remsberg* [2002] suggested the effects of wave forcing on mesospheric circulation induced by stratospheric solar forcing effects. *Hampson et al.*

[2005] suggest in situ wintertime wave activity over a given lidar site may be responsible for variations in lidar temperature profiles with the zonally average HALOE responses. There is also the possibility of a solar cycle interaction with the QBO [*Soukharev and Hood*, 2001]. There are several positive and negative correlations between solar outputs such as the 27-day solar rotation, the 11-year solar cycle, and the 22-year Hale cycle with weather and climate measurements. Though it remains uncertain how these physical phenomena are coupled to correlations between solar indices and state variables they do provide impetus for continued investigation [*Tsiropoula*, 2003].

8. Conclusions

A positive solar response coefficient at one altitude and a negative coefficient at another means the coefficient value must go through zero somewhere in between, which could mask decadal-scale signals that are out of phase with the solar input. This, along with issues of coefficient bias and attenuation, emphasizes the importance of including a variable phase solar-like term in a linear regression model of middle atmosphere

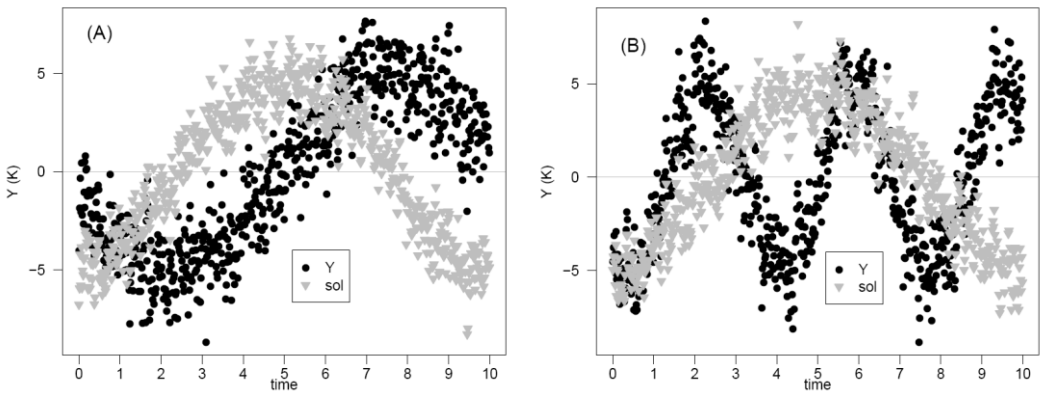


Figure 3.6. Simulated time series temperatures: (a) shows a simulated temperature time series indicated by the black circles and the solar-noise regressor, ***sol***, indicated by the gray inverted triangles; (b) shows the same for a different ***sol*** regressor.

temperatures.

The general conclusions are as follows. (1) If a fixed phase solar proxy is employed in a OLS temperature model and the atmosphere is responding out of phase to the solar input, the solar response coefficient can be attenuated and the cooling rate severely biased. Consequently, a variable phase solar proxy is preferable to a fixed solar proxy. (2) For cases where a fixed solar proxy is used, if the solar-proxy phase angle is 0 or π radians the bias on the linear trend coefficient will be reduced. At altitudes where the atmospheric solar response is in phase or out of phase with the solar input (where $\varphi = \theta$ or $\varphi = \theta \pm \pi$), the bias on the linear term is likewise reduced. In cases where a fixed proxy is employed, where the proxy coefficient has a maximum or minimum value the magnitude of the proxy coefficient at those altitudes seems more likely to better approximate the magnitude of the true atmospheric solar response at those altitudes. (3) Our data show evidence of near zero amplitude atmospheric solar response between 50 and 60 km, consistent with findings from other researchers. (4) A significant phase lag exists between the solar input and the atmospheric solar response at some altitudes. Our phases differ generally from *Remsberg* [2007], which might be due to zonal asymmetries and uncertainties related to the shortness of our data set. (5) The solar-noise term is statistically significant in the lower and upper mesosphere, but not in the middle mesosphere. Altitudes where the solar-noise term is statistically significant coincide with altitudes with the greatest difference between the *MgII* coefficient and the $sin\ \omega t$ coefficient values. This difference can be as great as 2 K. Based on analyzing the CPC and ERA data, as well as the USU data, a Malows' Cp test indicates that the inclusion of the solar-noise term is justifiable. Based on Monte Carlo simulations, including an *MgII*

term in a least squares model does not reveal information about the short term variations or out of phase solar responses. A two-proxy model is needed to bring out any such variations. (6) Researchers should be aware of the danger of fitting a regressor to temperature data if the noise of the regressor closely matches the noise in the temperature data. This can occur when the noise in the fixed-phase solar proxy is correlated to the noise in the temperatures. Such a correlation increases the risk of a false positive. The solar-proxy regressor might be considered statistically significant even if the large-scale sine-like variation is 90º out of phase to the atmospheric solar response.

While both our measurements and those from the HALOE instrument have shown that a significant variable phase proxy exists, and an analysis of our data and the CPC and ERA data shows that the solar-noise term is significant, it is now up to the theoreticians and modelers to explain the why these terms are found in middle atmospheric temperatures.

References

Calisesi, Y., and K. Matthes (2007), The middle atmospheric ozone response to the 11-year solar cycle, in *Solar Variability and Planetary Climates*, vol. 23, pp. 273-286. [online] Available from: http://dx.doi.org/10.1007/978-0-387-48341-2_22 (Accessed 4 February 2010).

Chanin, M. (2006), Signature of the 11-year cycle in the upper atmosphere, *Space Science Reviews*, *125*(1), 261-272, doi:10.1007/s11214-006-9062-5.

Chanin, M. L., N. Smirès, and A. Hauchecorne (1987), Long-term variation of the temperature of the middle atmosphere at mid-latitude: dynamical and radiative causes, *J. Geophys. Res.*, *92*(D9), 10933-10941, doi:10.1029/JD092iD09p10933.

Cossart, G. V., and J. Taubenheim (1987), Solar cycle and long-period variations of mesospheric temperatures, *J. Atmos. Terr. Phys.*, *49*(4), 303-307, doi:10.1016/0021-9169(87)90026-2.

Donnelly, R. F. (1991), Solar UV spectral irradiance variations, *J. Geomagn. Geoelectr., 43 Suppl.*, 835-842.

Donnelly, R. F., D. F. Heath, and J. L. Lean (1982), Active-region evolution and solar rotation variations in solar UV irradiance, total solar irradiance, and soft X Rays, *J. Geophys. Res.*, *87*(A12), 10318-10324, doi:10.1029/JA087iA12p10318.

Hampson, J., P. Keckhut, A. Hauchecorne, and M. Chanin (2005), The effect of the 11-year solar-cycle on the temperature in the upper-stratosphere and mesosphere: Part II numerical simulations and the role of planetary waves, *J. Atmos. Sol. Terr. Phys.*, *67*(11), 948-958, doi:10.1016/j.jastp.2005.03.005.

Keckhut, P., and K. Kodera (1999), Long-term changes of the upper stratosphere as seen by Japanese rocketsondes at Ryori (39°N, 141°E), *Ann. Geophys.*, *17*(9), 1210-1217.

Keckhut, P., A. Hauchecorne, and M. L. Chanin (1995), Midlatitude long-term variability of the middle atmosphere: Trends and cyclic and episodic changes, *J. Geophys. Res.*, *100*(D9), 18887–18897.

Kubicki, A., P. Keckhut, M. Chanin, G. Golitsyn, and E. Lysenko (2008), Temperature trends in the middle atmosphere as seen by historical Russian rocket launches: Part 2, Heiss Island (80.6°N, 58°E), *J. Atmos. Sol. Terr. Phys.*, *70*(1), 145-155, doi:10.1016/j.jastp.2007.09.010.

Lean, J. (1991), Variations in the sun's radiative output, *Rev. Geophys.*, *29*, 505-535.

Mallows, C. L. (1973), Some comments on Cp, *Technometrics*, *15*(4), 661-675.

McCormack, J. P., and L. L. Hood (1996), Apparent solar cycle variations of upper stratospheric ozone and temperature: Latitude and seasonal dependences, *J. Geophys. Res.*, *101*(D15), 20933-20944, doi:10.1029/96JD01817.

Remsberg, E. E. (2002), Seasonal and longer-term variations in middle atmosphere temperature from HALOE on UARS, *J. Geophys. Res.*, *107*(D19), doi:10.1029/2001JD001366. [online] Available from: http://www.agu.org/pubs/crossref/2002/2001JD001366.shtml.

Remsberg, E. E. (2007), A reanalysis for the seasonal and longer-period cycles and the trends in middle-atmosphere temperature from the Halogen Occultation Experiment, *J. Geophys. Res.*, *112*(D9), doi:10.1029/2006JD007489. [online] Available from: http://www.agu.org/pubs/crossref/2007/2006JD007489.shtml.

Remsberg, E. E. (2009), Trends and solar cycle effects in temperature versus altitude from the Halogen Occultation Experiment for the mesosphere and upper

stratosphere, *J. Geophys. Res.*, *114*(D12), doi:10.1029/2009JD011897. [online] Available from: http://www.agu.org/pubs/crossref/2009/2009JD011897.shtml.

Rottman, G. (1988), Observations of solar UV and EUV variability, *Adv. Space Res.*, *8*(7), 53-66, doi:10.1016/0273-1177(88)90172-X.

She, C., S. W. Thiel, and D. A. Krueger (1998), Observed episodic warming at 86 and 100 km between 1990 and 1997: Effects of Mount Pinatubo eruption, *Geophys. Res. Lett.*, *25*(4), 497, doi:10.1029/98GL00178.

Soukharev, B. E., and L. L. Hood (2001), Possible solar modulation of the equatorial quasi-biennial oscillation: Additional statistical evidence, *J. Geophys. Res.*, *106*(D14), 14855-14868.

Tsiropoula, G. (2003), Signatures of solar activity variability in meteorological parameters, *J. Atmos. Sol. Terr. Phys.*, *65*(4), 469-482, doi:10.1016/S1364-6826(02)00295-X.

CHAPTER 4

SUMMARY AND COMPARISON OF THE TEMPERATURE DATA FROM THE USU RAYLEIGH LIDAR

Abstract. This chapter contains a study of the temperature time series from the USU Rayleigh Lidar located at 41.74ºN, 111.81ºW. The database covers a time span from September 1993 to August 2003 and an altitude range of 45 to 90 km, and contains 593 nightly profiles. Cooling trend profiles, annual and semiannual amplitude and phase profiles, solar response amplitudes and phases are calculated. Collinearity and coefficient bias are considered as possible influences that could affect the regression results. The possibility that the Mt. Pinatubo eruption increased our early mesopause temperatures resulting in large mesopause linear cooling trends is considered. The linear trends are compared to those in the review by *Beig et al.* [2003]. These results are compared to others, in particular the semiannual oscillation from USU temperatures and the Halogen Occultation Experiment (HALOE) on the UARS satellite. There is also a brief description and comparison of the USU annual and semiannual amplitudes and phases with others found in the literature. The USU solar response amplitudes and phases are compared to those from the HALOE data.

1. Introduction

A theoretical connection between elevated atmospheric CO_2 levels and increased global temperatures has existed for over 100 years [*Callendar*, 1938; *Held and Soden*, 2000]. In the past, there was some debate as to whether or not industrialization would produce global warming or global cooling. However, over the past several decades the

literature has increasingly, heavily favored the global warming thesis, which states that significant increases in the quantity of atmospheric CO_2 results in elevated global temperatures. According to calculations made by the author, the atmosphere could experience a doubling of CO_2 from preindustrial levels between the years 2070 and 2100. Atmospheric models predict that doubling the amount of atmospheric CO_2 will increase heat retention in the troposphere and increase heat loss in the stratosphere and mesosphere. Surface temperatures are expected to increase by about 1.5 to 4 C, while the middle atmosphere is expected to cool between 8 to 10 C, depending on the model simulation, location and altitude [*Roble and Dickinson*, 1989; *Rind et al.*, 1990, 1998; *Held and Soden*, 2000; *Fomichev et al.*, 2007]. Hence, because of the larger temperature change many scientists are looking for evidence of global warming in middle atmosphere secular temperature trends.

Ordinary least squares (OLS) models are frequently employed as a way to extract useful information from temperature time series data about atmospheric parameters of interest, such as the amplitude and phase of the annual and semiannual oscillations, the atmospheric solar response, the linear cooling rate, the effects of interventions, such as the Mt. Pinatubo eruption [*She et al.*, 1998], the quasi-biennial oscillation, and the turnaround time and recovery for ozone levels [*Reinsel et al.*, 2002, 2005]. Least squares has many advantages. It minimizes what the model cannot explain, it offers the best linear unbiased estimator (the BLUE assumption) when certain conditions are satisfied, it is simple and the results are typically easy to interpret. Two difficulties sometimes associated with OLS are coefficient correlation (collinearity) and model specification. Serial autocorrelation also occurs in the residuals and can affect the regression results.

Collinearity and model specification are more specifically addressed in Chapters 2 and 3. Serial correlation is discussed in Chapter 5 and Appendix E.

This paper presents an analysis of the USU mesosphere time series of temperatures from the USU lidar database. Of interest here are the mesosphere cooling trends, the amplitudes and phases of the annual oscillation (AO) and semiannual oscillation (SAO), as well as the amplitudes and phases of the atmospheric solar response.

2. Summary of Sections

Section 3 contains a short description of the principal regression model used in this dissertation. Several reasons for separating the solar-noise from the solar proxy and introducing it as a separate model regressor are given. Section 4 contains a discussion of the linear trend coefficient, its time evolution, and the effects of collinearity including coefficient correlation and inflated standard errors (SEs). The effects of model specification on the linear trend coefficient and the solar proxy coefficients are discussed. Comparisons of linear trends from the USU temperatures are made with linear trends from the literature. Section 5 presents evidence for the existence of a Pinatubo effect in mesopause temperatures and the effects it may have had on our linear trend estimates are discussed. In section 6 the discussion from section 5 is continued. In particular, the collinearity and model specification issues from section 4 are expanded upon and summarized. In section 7 the atmospheric solar response is discussed. Solar response amplitudes and phases of the USU temperatures are discussed and compared with those from the HALOE instrument reported in *Remsberg* [2007, 2009]. For comparison, amplitude profiles for a fixed-proxy model are included and compared with model

simulations, as well as results reported by other researchers. The summer and winter profiles are briefly examined. Section 8 contains a short discussion of the annual and semiannual oscillations. Our findings are compared to those from other researchers. An SAO climatology of the USU temperatures is discussed and compared to an HALOE SAO climatology. Section 9 contains the final conclusions.

In this paper, least squares is used quite heavily. It is therefore convenient to use a short hand notation to refer to different models. For example, $y \sim x1 + x2$ indicates a column of **y** data projected onto the column space $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{x1}, \boldsymbol{x2})$, or $\boldsymbol{y} = I\cdot\boldsymbol{1} + a\cdot\boldsymbol{x_1} + b\cdot\boldsymbol{x_2} + \boldsymbol{\varepsilon}$. In the shorthand notation the intercept $\boldsymbol{I}$, the noise $\boldsymbol{\varepsilon}$, and the coefficients are implied. This follows the convention of the R programming language. Another shorthand convention is $sin(\omega\boldsymbol{t})$, which should be taken to indicate that $sin(\omega\boldsymbol{t})$ is evaluated at each time value: $sin(\omega\boldsymbol{t}) = \{sin\ \omega t_1, sin\ \omega t_2, sin\ \omega t_3, \ldots, sin\ \omega t_n\}$.

3. The Model

Unless otherwise stated the regression model for this chapter is

$$\boldsymbol{T}(z) = \alpha(z)\boldsymbol{\imath} + \beta(z)\cdot\boldsymbol{t} + A_1(z)cos(2\pi\cdot\boldsymbol{t}) + A_2(z)sin(2\pi\cdot\boldsymbol{t}) + B_1(z)cos(4\pi\cdot\boldsymbol{t}) + B_2(z)sin(4\pi\cdot\boldsymbol{t}) + C_1(z)sin(\omega\boldsymbol{t}) + C_2(z)cos(\omega\boldsymbol{t}) + D(z)\cdot\boldsymbol{solnoise} + \boldsymbol{\varepsilon}(z), \quad (4.1)$$

where z is the altitude, α is the intercept, β is the linear trend, A_1 and A_2 yield the amplitude and phase of the annual oscillation, B_1 and B_2 are the same for the semiannual oscillation, C_1 and C_2 yield the amplitude and phase of the solar-like atmospheric response with $\omega \sim 2\pi/11\text{-year}^{-1}$, the frequency of the solar cycle. The ***solnoise*** term is obtained by fitting $MgII \sim sin\omega t + cos\omega t$, where the solar-noise are the model residuals.

The daily Mg II data were obtained from the NOAA website. A handful of missing Mg II data points were interpolated and the time series was smoothed using an 81-day boxcar average. For convenience, this will hereafter be referred to as Model (4.1).

One justification for separating the solar-like oscillation from the solar-noise is the possibility of a phase lag between the solar input and the atmospheric solar response. There is good reason to believe that the atmospheric response to the solar input can be significantly out of phase with the solar input. *Remsberg* [2002] found a phase lag of 2.3 years at 40ºN and 0.05 hPa (~69 km), a lag of 1.9 years at 0.03 hPa (~73 km), and 1.5 years at 0.02 hPa (~76 km), at the same latitude. (See Table 7 in that paper.) In an updated paper *Remsberg and Deaver* [2005] analyzed HALOE data from 1991-2004 and reported a phase lag of 3.8 years at 0.05 hPa and 2.2 years at 0.03 hPa. The existence of an altitude-dependent phase lag is confirmed again in *Remsberg* [2008], which reports a phase lag of 4.5 years at 69 km and a phase lag of about −1.5 years between 58 and 63 km.

Also, if the solar noise is omitted from the model, spurious results can arise at altitudes where the solar-noise is correlated with the model residuals. The least-squares technique minimizes a quantity called the residual sum of squares (RSS), and in several exemplifying simulations the solar-noise was found to reduce the RSS much more than a sine-like solar oscillation did. Consequently, if the two are not separated, the reduction of the RSS by the solar-noise could lead to a false positive: The solar proxy coefficient could be considered statistically significant when the atmospheric response is significantly out of phase with the solar proxy. By separating them, the problem is avoided. Additionally, the magnitude of the solar-noise coefficient might contain

information about how the atmosphere is responding to short-term solar input. This information cannot be retrieved without separating the solar-noise from the solar proxy. For more information about how the solar-noise can affect regression results see Chapter 5 and Appendix D.

4. Linear Trend Coefficient

The magnitude of the linear trend coefficient is commonly used as an indicator of the strength of middle atmosphere cooling. One challenge in interpreting the linear trend coefficient is the evolution it undergoes as more data is added to the data set. The following Monte Carlo simulations illustrate this point. A simulated temperature time series was generated containing a linear trend of −0.4 K/year, a 4 K solar response with an 11-year period, and Gaussian noise with zero mean and 9 K standard deviation, all of which are realistic parameters for upper mesosphere temperatures. Each simulation starts with 10 years of data and continues until 13.5 years are covered. A least squares regression is done on the simulated data and, as more data are added, new coefficients obtained, giving an idea of how the linear coefficient can evolve. The results from four different Monte Carlo runs are shown in Figure 4.1. The first simulation starts with a cooling trend of −0.4 K/year. The magnitude of the trend estimate then decreases to −0.3 K/year over the period of a year. Then, within a half-year, the magnitude increases to −0.5 K/year and quickly decreases again to −0.4 K/year. The second simulation starts with an estimated cooling of −0.15 K/year. The magnitude of the trend estimate steadily increases over the next two years to −0.4 K/year before decreasing only slightly. The fourth simulation starts with a trend estimate of −0.4 K/year, which increases in

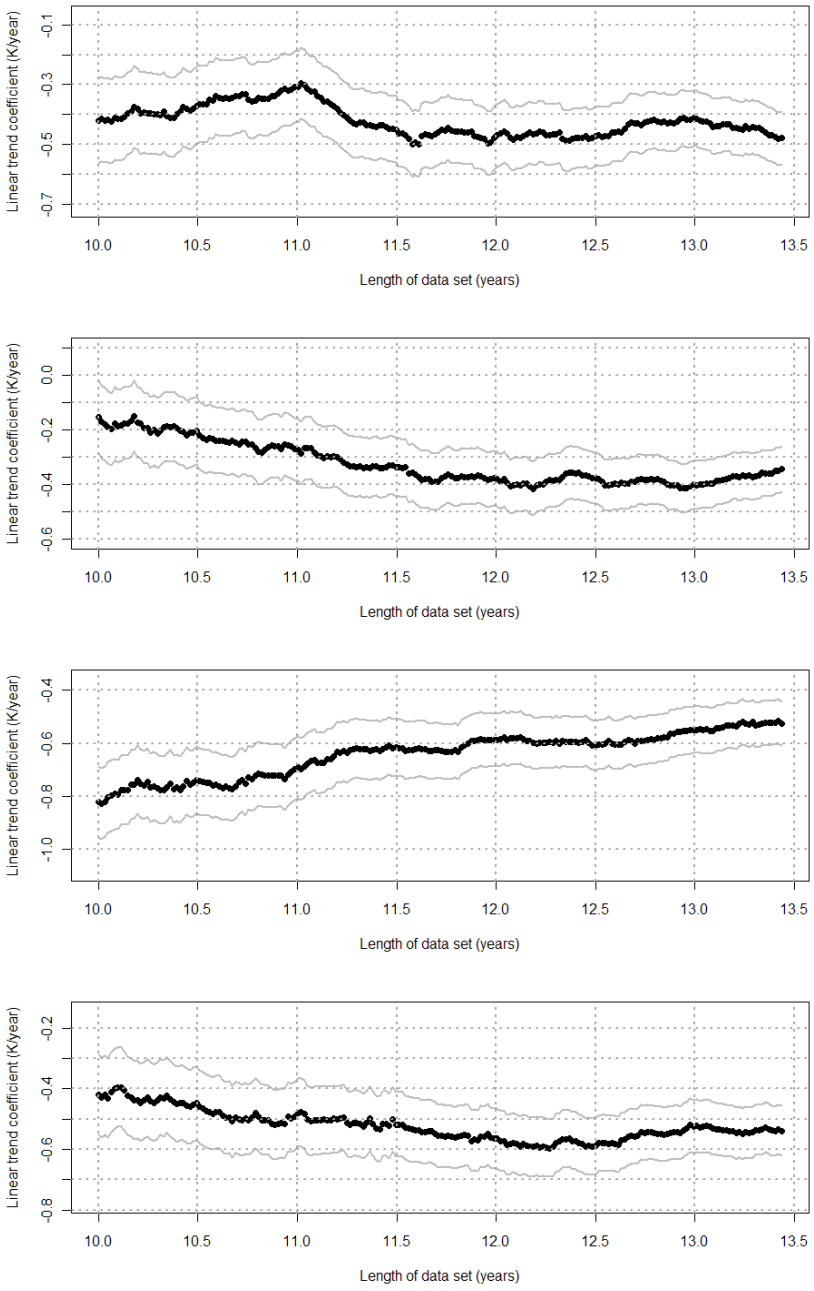


Figure 4.1. The time evolution of the linear trend coefficient from four Monte Carlo simulations. The simulations start with 10 years of data and continue to 13.5 years. These four simulations were selected to demonstrate the possible variations that can occur over a period of 3.5 years. The two middle simulations are extreme examples.

magnitude steadily over a period of two years to −0.6 K/year and then decreases to −0.5 K/year.

These results indicate that the linear trend coefficient not only has an inherent variance that depends on the model specification and residual standard deviation, but that it also undergoes significant temporal variability as the length of the data set increases. Given enough time the linear trend coefficient will approach its true value, but convergence might not be immediate or initially in the right direction.

Likewise, linear estimates from various sites can have different starting values and exhibit very different time evolutions. For example, Figure 4.2 shows the linear trend profiles from the combined data series from the French CEL (Centre d'Essais des Landes, 44ºN) and OHP (Observatoire de Haute Provence, 44ºN) lidars from 1979-1994 [*Keckhut et al.*, 1995] and an updated profile based on data from 1979-1998 [*Ramaswamy et al.*, 2001]. The addition of four years of data noticeably altered the vertical profile. At 64 km the magnitude of the difference is 0.3 K/year, which is significant considering the linear trend profile itself varies from –0.1 to −0.4 K/year. One way to deal with this variation is

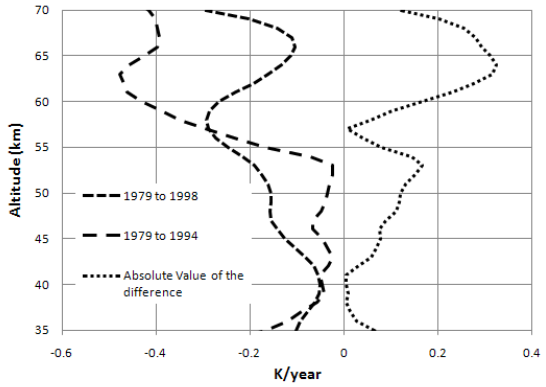


Figure 4.2. Linear trend profiles from the OHP and CEL lidars. The profile for 1979-1994 is from *Keckhut et al.* [1995]; the 1979-1998 profile is from *Ramaswamy et al.* [2001].

to compare linear trend estimates from different sites. But individually there is not much that can be done about the time evolution of the linear trend coefficient, except to bring out the fact that significant variation can occur and that different linear trend profiles obtained from observations can look very different from each other. This demonstrates the importance of a very long data set. It should be added that despite this variation, none of the Monte Carlo simulations showed a zero or positive cooling trend.

Another difficulty in interpreting the linear trends is the problem of coefficient correlation (collinearity). This problem arises from the model itself and is unrelated to the temperature data. Depending on the degree of linear dependence between regressors, their coefficients can be correlated. A more in depth analysis of this is given in Chapter 2. In short, if two regressors are highly correlated then their coefficients are also likely to be correlated.

The coefficient correlations for Model (4.1) are given in Table 4.1. The highest correlation is −0.83 and is between the linear trend coefficient β and the solar-like coefficient C_1 belonging to $sin\ \omega t$. The next strongest correlation is 0.52, between β and the other solar-like term C_2 belonging to $cos\ \omega t$. A negative correlation between β and C_1

Table 4.1. The coefficient correlations for Model (4.1). The strongest correlation is −0.83 between the solar-like sine term and the linear trend. The next strongest correlation is 0.56 between the linear trend and the solar-like cosine term. The intercept is of no interest and was omitted. Some of the sine and cosine terms are not completely orthogonal because of gaps in the data.

	β	A1	A2	B1	B2	C1	C2	D
β	1.000	-0.173	0.191	0.078	-0.045	-0.829	0.516	-0.011
A1	-0.173	1.000	-0.222	-0.01	-0.199	0.215	0.139	0.077
A2	0.191	-0.222	1.000	0.107	0.233	-0.106	-0.044	0.048
B1	0.078	-0.010	0.107	1.000	-0.001	0.016	0.119	-0.065
B2	-0.045	-0.199	0.233	-0.001	1.000	-0.014	-0.041	-0.103
C1	-0.829	0.215	-0.106	0.016	-0.014	1.000	-0.413	0.024
C2	0.516	0.139	-0.044	0.119	-0.041	-0.413	1.000	-0.013
D	-0.011	0.077	0.048	-0.065	-0.103	0.024	-0.013	1.000

indicates an inverse relationship. If β is higher than its true value then C_1 will be lower than its true value; conversely, if β is low then C_1 will be high. Two other effects of collinearity are inflated standard errors and sensitivity to model specification. When two regressors are correlated their SEs are also inflated; and, if regressors are added to or removed from the model, the coefficient values undergo large changes.

To check for coefficient sensitivity to model specification, regressions on the USU temperatures were done using models with the sine and cosine terms omitted, with the *MgII* term in place of the sine and cosine terms, with both the sine and cosine terms included, and with only the $sin\ \omega t$ term included. With the exception of the regressions containing the *MgII* term, all the models included a solar-noise term. It was found that, with the exception of the model omitting the sine and cosine terms, the linear coefficient estimates were quite similar between 50 and 70 km. However, all the profiles were confined to the 95% confidence intervals of the linear estimate for Model (4.1), indicating that the linear term is not highly sensitive to model specification. The linear trend estimates may be confined to the region of the error limits shown in Figure 4.3a and b.

There is a large cooling rate in the upper mesosphere, about −1.9 K/year (−19 K/decade) at 88 km. Only a few researchers report similarly large cooling trends. In a review of mesosphere temperature trends, *Beig et al.* [2003] lists several of those reported in the literature. Histograms for these mesopause and mesosphere trends are shown in Figure 4.4. If the trends are interpreted as a distribution, the median mesopause trend is not different from zero, but the median mesosphere trend is clearly different from zero. The median mesopause trend is approximately −0.05 K/year and the median

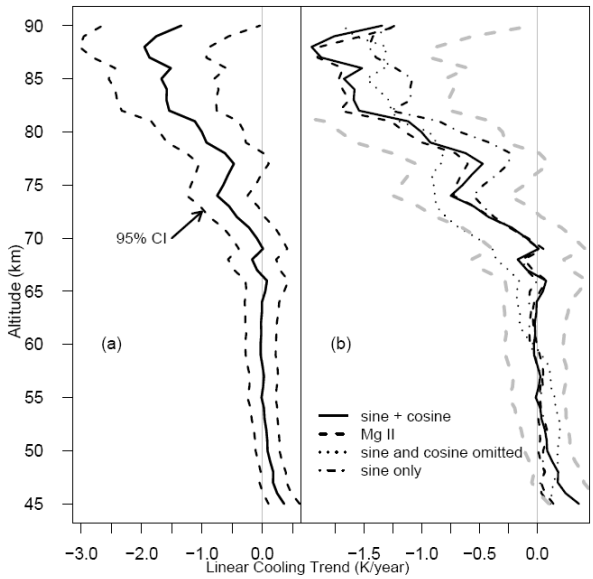


Figure 4.3. Two linear trend profiles. Part (a) is the linear trend profile for Model (4.1). Part (b) contains the linear trend profile for Model (4.1) and the model variations indicated in the legend. The error limits from (a) are included for comparison. Note the scale change.

mesosphere trend is approximately −0.35 K/year.

There are four cases reported in *Beig et al.* where the mesosphere and mesopause trends are around −10 K/decade. A mesopause trend from OH φairglow intensities over Argentina was reported by *Resin and Scheer* [2002] at −10.5 K/decade; and a wintertime −9 K/decade trend from OH* rotational bands was reported by *Semenov et al.* [2002]. Both are from mid-latitudes. The data used in *Semenov et al.* is from several data collection sites spanning from North America to Russia. Two large mesosphere trends were found in rocketsonde temperatures reported in *Golitsyn et al.* [1996] who found a −10 K/decade trend at high latitudes and a −8.8 K/decade trend at mid-latitudes.

5. Pinatubo Eruption

The Mt. Pinatubo eruption occurred during June 9-17 1991 and was, according to the

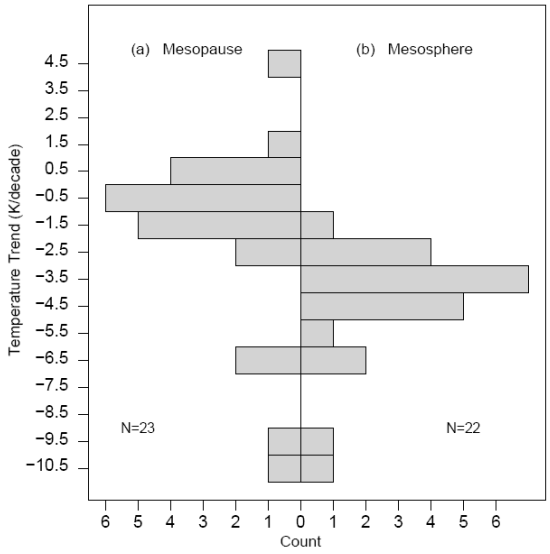


Figure 4.4. Histograms of temperature trends. Histograms of the temperature trends from *Beig et al.* [2003]. Histogram (a) is based on mesopause trends (80-100 km) from Table 5. Histogram (b) is based on mesosphere trends (50-79 km) from Table 4. For cases where the temperatures were reported as, for example, −1.4 to −2.1 K/decade, both the upper and lower limits were treated as data points.

U.S. Geological Survey, the second largest volcanic eruption of the 20th century. This eruption produced 20 to 30 megatons of new aerosol sulfate particles, mainly from chemical reactions with sulfur dioxide [*McCormick and Veiga*, 1992]. These particles scatter light at the visible wavelengths but absorb radiation in the IR and near IR spectral regions. The net effect is heating [*Thomas et al.*, 2009].

She et al. [1998] found a 9 K and 12.9 K warming at 86 and 100 km, respectively. The maximum of these warmings occurred in mid 1993 and early 1993, respectively, approximately two years after the eruption. *Keckhut et al.* [1995] reported a temperature increase of 2 to 3 K between 30 and 40 km in their temperature residuals occurring from 1992 to the summer of 1993, which they attributed to the Pinatubo eruption. They also included an optical depth parameter in their regression model and found it to be

statistically significant between 30 and 35 km and between 60 and 74 km. *Bittner et al.* [2002] also found evidence of a possible Pinatubo effect in temperatures spanning from 1987 to 1997 in the OH layer (~ 87 km) above Wuppertal, Germany (51ºN). They found a phase shift in the AO and SAO temperatures occurring 1 year after the Pinatubo eruption and a large increase in the amplitude of a ter-annual signal occurring approximately two years after the eruption. These AO and SAO phase and amplitude shifts were not found in the USU temperature data and the time of their occurrence is a little earlier than the effects reported in *Keckhut et al.*, *She et al*., and *Bittner et al.* attributed these differences in timing to zonal asymmetries.

The difficulty in determining the presence of a Pinatubo effect in the USU data is that our temperatures begin late 1993, which is when the temperature peak is likely to have already reached its maximum. So our early temperatures are expected to be perturbed higher. But this presents a problem when trying to detect a secular trend. Any initial temperature perturbation will exert leverage on the β coefficient increasing its magnitude.

One way to test for the presence of leverage is to divide the data into subsets and perform linear regressions on each. A one-year gap occurs four years from the beginning of our data set, so that seemed like a good place to divide the data. Let S_1 indicate data from September 1993 to April 1997 and S_2 indicate data from May 1998 to August 2003. S_1 consists of 251 nightly profiles and S_2 consists of 333, at 45 km. The solar terms were excluded from the model because, owing to the shortness of S_1 and S_2, the collinearity problem was extreme, to say the least. With both the *sin ωt* and *cos ωt* terms included, the linear trend for S_1 was −14 K/year! This was clearly a case where the collinearity problem was so extreme it would be better to eliminate some variables and risk

coefficient bias.

First the USU data was deseasonalized by removing the annual and semiannual oscillations from the temperatures. This is justifiable because the annual and semiannual sine and cosine terms are not correlated with the other model terms. (See Table 3.1.) The model fit to data subsets S_1 and S_2 was $T \sim t$, where t is time and T the deseasonalized temperatures. The regression profiles for both S_1 and S_2 are shown in Figure 4.5a. Below 50 km there is a maximum difference of 1 K/year between the linear trends for S_1 and S_2. S_1 has a warming of 0.39 K/year and S_2 has a cooling of −0.45 K/year. From 50 to 71 km the difference is less than 0.2 K/year and averages to about 0.2 K/year. Above 71 km they sharply diverge. The linear trend for S_1 is on average 2.1 K/year less than the linear trend for S_2. The error limits are not shown because our main interest is the difference in the mean β coefficient values.

A linear trend profile was also calculated with the first year omitted from the data set. The new data set S_3 consists of data from September 1994 to August 2003 and 531 nightly profiles. The linear trend profiles are shown in Figure 4.5b. From 50 to 80 km they are nearly identical to the OLS linear trends from the full data set. An exception is at 65 km where the S_3 linear trend is slightly smaller in magnitude than the OLS trend from the full data set. Above 80 km the linear trends for S_3 are smaller in magnitude than the OLS trends from the full data set by about 0.2 K/year.

This suggests that between 80 and 90 km the temperatures at the beginning of the data set are greater than those from remainder of the data set, suggesting a temperature perturbation consistent with a temperature increase from the Pinatubo eruption.

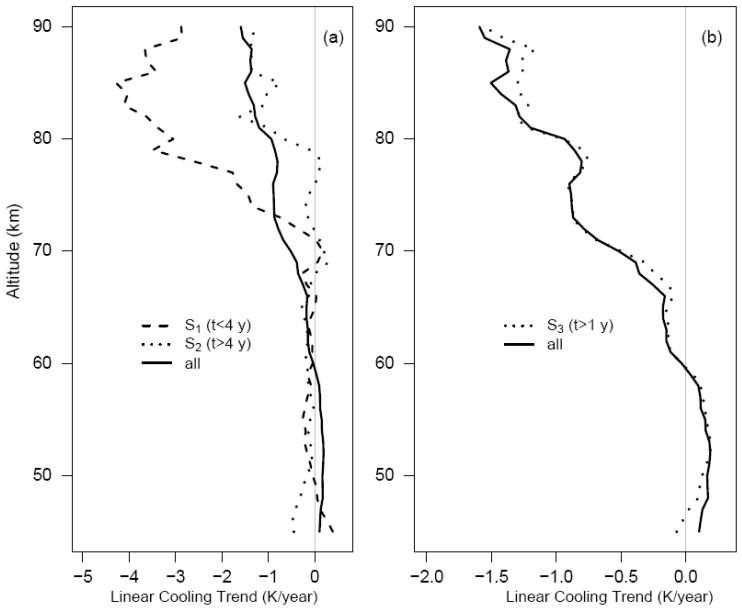


Figure 4.5. Comparison of the linear trends from an OLS model applied to S_1, S_2, and the entire data set. Figure (a) shows linear trends from the first half S_1 and second half S_2 of the data set, as well as the linear trends for the entire data set. Figure (b) shows the linear trends for the entire data set, as well as those with the first year omitted from the data. Note the different horizontal scale.

There is an additional difficulty in determining the presence of a Pinatubo effect. The atmospheric solar response can be out of phase with the solar input. When a model includes a traditional fixed-phase solar proxy, a sine-like signal can remain in the residuals. An example of this is shown in Figure 4.6 where the residuals for the model $y \sim t + AO + SAO + sin\omega t + solnoise$ are plotted, where y are the temperature time series from 45 km. A fourth order polynomial is added to the figure to bring out the structure. A clear sine-like variation remains in the residuals, possibly suggesting an early Pinatubo effect or an unaccounted for solar-like atmospheric response. When this is done with temperatures from 85 km an underlying signal also remains. The residuals are shown in Figure 4.7 where it is evident that the earlier residuals drop off quickly and then level out.

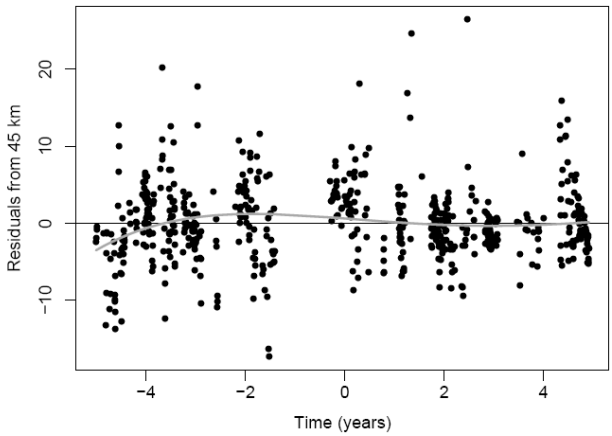


Figure 4.6: Some model residuals from 45 km. The residuals from the model $y \sim t + AO + SAO + C1 + solnoise$, where *C1* is the solar-like $sin\ \omega t$ term and y are the temperatures from 45 km. A fourth order polynomial was fitted to make the underlying sine-like structure in the residuals more visible.

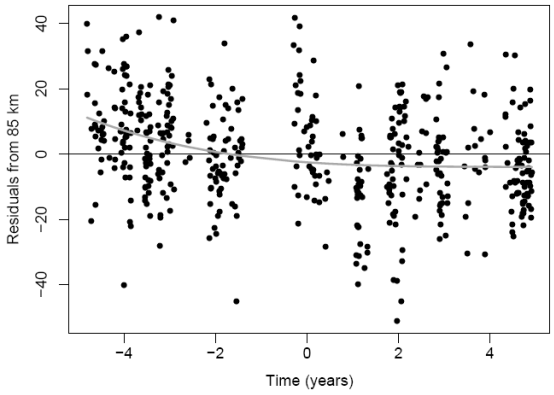


Figure 4.7. Some model residuals from 85 km. The residuals from the model $y \sim t + AO + SAO + C1 + solnoise$, where *C1* is the solar-like $sin\ \omega t$ term and y are the temperatures from 85 km. A fourth order polynomial is fitted to make the underlying structure in the residuals more visible.

This is indicative of a strong temperature perturbation at the beginning of the data set and could indicate a Pinatubo effect, an unaccounted for solar-like response, or both.

6. The Effects of Collinearity on Error Limits

One symptom of collinearity is a large change in the coefficient values when a model variable is inserted or omitted. As shown in section 4, this is not the case with the linear trend coefficient, which does not exhibit large changes when various solar-proxies are included or omitted from the model. Collinearity also inflates the standard errors which has a direct bearing on the confidence limits. If one model variable is highly correlated with another, the standard errors of both are inflated. This affects the confidence placed in the results. In linear regression problems, a typical null hypothesis is something like H_0: $\theta = 0$, where θ is a regression coefficient. The p-value gives evidence on whether to assert or reject H_0 given the data. If the evidence strongly indicates that H_0 is false, then H_0 is usually rejected and some confidence is placed in the regression results. Higher SEs increase the error limits, consequently increasing the chance of accepting H_0 when it is false.

To illustrate the point, two models were fit to the data. Model (4.1) is the full model and Model (4.2) is simply Model (4.1) with the solar-like terms omitted. Figure 4.8 shows the 95% linear trend error limits for Models (4.1) and (4.2) along with the linear trend profiles from each model. On average, the error limits from Model (4.2) are 48% smaller than those from Model (4.1). With the error limits increased, the chances of determining the linear trend coefficient to be indistinguishable from zero increases. That is, H_0: $\beta = 0$ is more likely to be accepted. The large changes in the error limits shown in

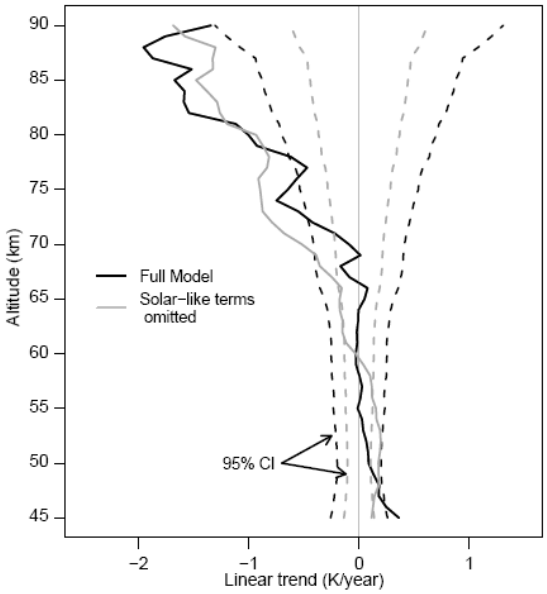


Figure 4.8. Two linear trend profiles. The linear trend profiles from Model (4.1) and Model (4.2) with 95% confidence levels given.

Figure 4.8 are principally due to collinearity between the linear term and solar-like *sin ωt* term.

At 45, there is a warming trend of about 0.36 K/year for the full model, and about 0.24 K/year for Model (4.2). The trends for Model (4.2) indicate a warming of about 0.17 K/year between 46 and 56 km. For the full model, there is no statistically significant warming or cooling between 46 and 73 km. The cooling rate between 65 and 90 km for Model (4.2) ranges from −0.2 to −1.7 K/year. For the full model there are no significant cooling trends between 47 and 72 km. The full model indicates a statistically significant cooling between 73 and 90 km, but with two exceptions at 77 and 78 km that are just below the 95% level.

The linear trends for models (4.1) and (4.2) differ by a maximum of 0.6 K/year. On average they differ by 0.2 K/year. The maximum linear trends for models (4.1) and (4.2)

are −1.9 K/year at 88 km and −1.7 K/year at 90 km, respectively. It should also be pointed out that the upper mesosphere is where the linear trend has its greatest uncertainty. For Model (4.1) the uncertainty is ±1.25 K/year to the linear trend, and for Model (4.2) it is ±0.6 K/year. The larger mesopause error limits arise partly from increased noise levels above 80 km, but also because there are fewer data points from the upper mesosphere than from 45 km.

There is no way to get around the fact that our mesopause cooling rates are larger than what other researchers have found. The natural thing to do is point to the influence of the Pinatubo eruption. Unfortunately because we do not have data before the eruption, a clear assessment cannot be made. It also appears that in the upper mesosphere, the first year's temperatures (Figure 4.5b and Figure 4.7) exert some leverage on the linear trend coefficient, though this amounts to only 0.2 K/year which means the cooling rate remains quite high. This also suggests that the Pinatubo effect, if it is present in the USU temperatures, is greatest between 70 and 90 km.

Naturally the results from Model (4.1) are preferred because this is the full model. Because a possible Pinatubo effect casts some doubts on upper mesopause cooling trends our greatest confidence can be place in the results between 73 and 80 km where a statistically significant linear trend from −0.5 to −1 K/year was found for the full model. Below that, down to 47 km, the cooling trends vary from −0.5 to +0.3 K/year and are not statistically significant. At 45 and 46 km there is a warming of about 0.35 K/year.

7. Atmospheric Solar Response

The approximate 11-year variation in solar output is known to affect middle

atmosphere temperatures and chemistry. In the middle atmosphere, the solar ultraviolet output is of particular interest because of its significant impact on stratosphere and mesosphere temperature structure. While overall solar intensity varies less than 1% over the 11-year solar cycle, the shorter UV spectrum varies from 5% at 205 nm, increasing to 50% in the Lyman-α line [*Donnelly et al.*, 1982; *Donnelly*, 1991]. This large variation in short-wave radiation affects photochemical ozone production and can alter middle atmosphere thermal characteristics, which can, in turn, alter the propagation of atmospheric waves and global circulation patterns affecting heat advection and chemical transport [*Calisesi and Matthes*, 2007].

In the Mesosphere Lower Thermosphere (MLT) region, species such as O and H become long lived and can be transported to other parts of the atmosphere where they loose their energy as heat [*Smith*, 2004]. This can delay the atmospheric response for several months but cannot account for a delay of several years. As argued in Chapter 2, the temperature response to solar input can vary from ±5 years. This delay is not a time delay as such; that is, the atmosphere does not wait five years to respond to the solar input, but rather, the driving of the atmosphere by the 11-year solar variation causes an out-of-phase atmospheric response. An example is the inverse solar response found at some altitudes. When the solar input is at a maximum, the atmospheric temperatures are at a minimum, which amounts to a 180º phase change which can be expressed as ±5 years. A similar argument could apply to any phase offset between the solar input and the atmospheric temperature response.

At the upper mesosphere, the principal driver of temperature change on solar cycle timescales is from chemical heating [*Aikin et al.*, 1991; *Huang and Brasseur*, 1993]. The

bulk of which is provided by odd oxygen (O_x) reactions such as $O + O + M$ and $O + O_2 + M$, along with a contribution by direct solar heating [*Huang and Brasseur*, 1993]. From solar maximum to minimum, the number density of odd oxygen constituents decreases by more than 40%. This decrease precipitates a decrease in ozone production and increases the levels of the OH catalytic agent, which enhances ozone loss and further depresses ozone levels during solar minimum [*Huang and Brasseur*, 1993; *Khosravi et al.*, 2002; *Smith*, 2004; *Brasseur and Solomon*, 2005].

Below 80 km, the temperature response to solar input is smaller than that at the mesopause. This is largely due to a smaller change in ozone mixing ratio over the period of a solar cycle. In this region atmospheric models predict a negative or near zero change in ozone mixing ratio from solar maximum to solar minimum. The $O + O_2 + M \rightarrow O_3 + M$ reaction which dominates odd oxygen partitioning also has a negative temperature dependence [*Froidevaux et al.*, 1989; *Jonsson et al.*, 2004; *Marsh et al.*, 2007]. Consequently, ozone production slightly increases with decreasing temperature, which offsets some of the temperature change over the solar cycle.

An early simulation by *Garcia et al.* [1984] modeled the effects of the 11-year solar cycle on middle atmosphere temperatures. This was a zonally averaged two-dimensional model with coupled photochemistry and dynamics, and covered and altitude range from 16 to 116 km. Their max-min temperature response was 2 K at 46 km and 40ºN, and about 1 K between 50 and 76 km. Above 76 km it steadily increased to 6 K at 90 km. *Huang and Brasseur* [1993] using a two dimensional interactive mesosphere-stratosphere model found a max-min response of 1.5 K at the stratopause that increased to around 10 K near the mesopause, at 40ºN. Using the SOCRATES 2-D model *Khosravi et al.*

[2002] found a max-min temperature response of less than 1 K at 50 km and 40ºN, increasing to 5 K at 90 km. A similar temperature max-min profile was found by *Schmidt et al.* [2006] using the HAMMONIA model, which simulates interactive dynamics, chemistry and radiation up to 250 km. A somewhat smaller response was found by *Marsh et al.* [2007] in a simulation using the NCAR Whole Atmosphere Community Climate Model (WACCM), a general circulation model with interactive chemistry. They found a 0.75 to 2 K atmospheric solar response at 40ºN and between 45 and 80 km. A different simulation run on the MSDOL model (developed from the NCAR ROSE model) was done by *Hampson et al.* [2005] who found a max − min atmospheric solar response of < 1 K in the upper stratosphere and lower mesosphere region. They also found the solar response to be notably stronger during winter (~1.8 K) than during summer (~0.5 K) in the same atmospheric regions.

7.1. Analysis and Comparison of USU Data

As described in Chapter 3, three models were fit to the deseasonalized USU temperatures: $E(T) = \beta \cdot t + \alpha_1 \cdot sin(\omega t + \varphi)$, $E(T) = \beta \cdot t + \alpha_2 \cdot sin\ \omega t$, and $E(T) = \beta \cdot t + \alpha_3 \cdot MgII$, where $E(T)$ are the expected values for the temperature models. The profiles for α_1, α_2, and α_3 are shown in Figure 4.9a,b,c. Figure 4.9a shows the amplitudes (max – min) of the atmospheric solar responses in the USU temperatures along with similar solar response amplitudes found in the HALOE temperatures reported in *Remsberg* [2007, 2009]. Between 50 and 64 km, our amplitudes are on average 1 K smaller than Remsberg's, with a maximum difference of nearly 2 K. In the upper half of the mesosphere Remsberg's amplitudes are about 0.1 to 2 K smaller than ours, but most are within or nearly within

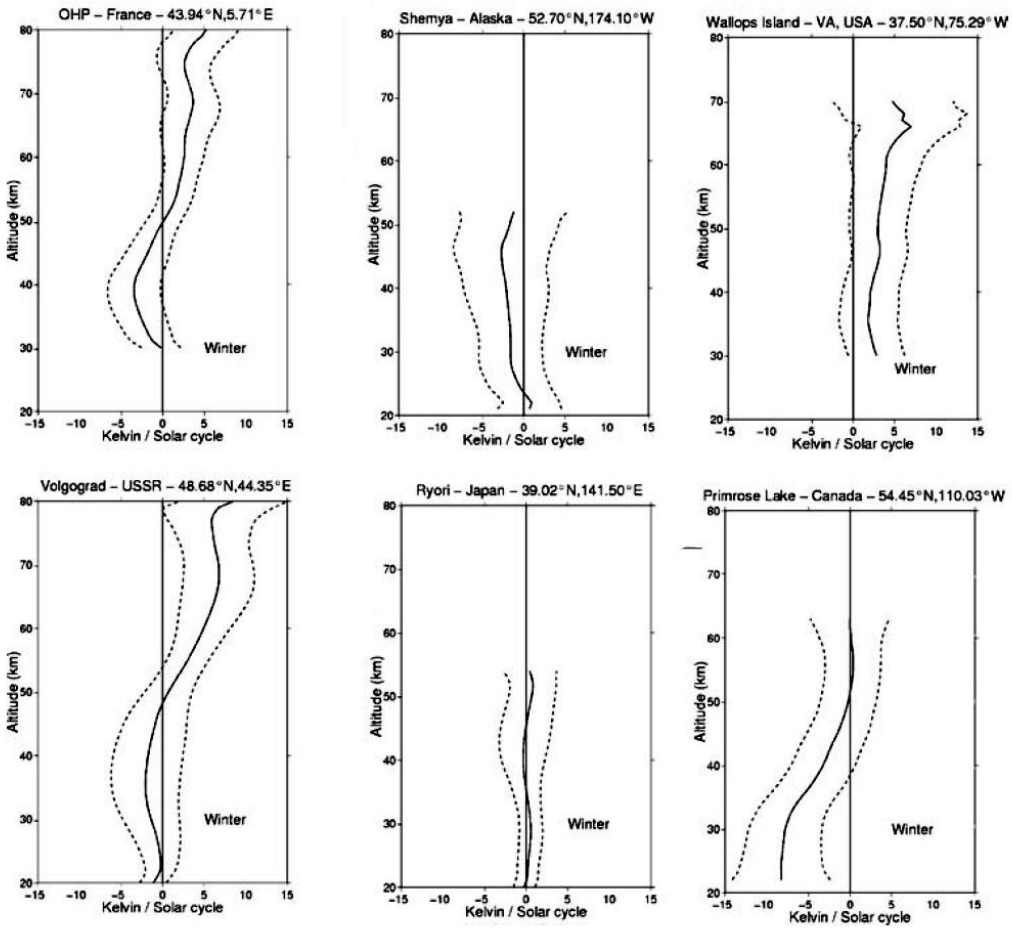


Figure 4.9. Various linear trend profiles. The solar response profiles from using a fixed-phase model on in situ rocket-sonde and lidar temperatures from various sites. (Figures are from Figure 8 in *Chanin* [2006]. Used with permission of Springer Publishing Company.)

our 95% confidence limits.

Figure 4.9b shows the phase differences between the atmospheric solar response and solar maximum found in the USU temperatures and similar phase differences reported in *Remsberg* [2007] and *Remsberg* [2009]. To make our data more comparable with Remsberg's our phases were measured as years from 1/1/2002, which is where Remsberg began measuring his phases. Remsberg's phases are from 40 and 50ºN. The USU solar

response is out of phase with the solar input at 45 km and nearly in phase with it at 55 km. Above 55 km the phase difference slowly decreases with increasing altitude until it reaches near zero at 88 km, above which it becomes out of phase with the solar proxy. Between 66 and 80 km the 40ºN Remsberg phases are within or nearly within our 95% error limits, which span nearly three years in the upper mesosphere. Below 66 km both our phases and Remsberg's show little similarity except for two points at 54 and 58 km where the phase offset is about −1.8 years.

Figure 4.9c shows the amplitude profiles of the sin ωt and MgII proxies obtained by fitting $E(T) = \beta \cdot t + \alpha_2 \cdot sin\ \omega t$ and $E(T) = \beta \cdot t + \alpha_3 \cdot MgII$ to the USU temperature data. The Mg II data was scaled to fit the sin ωt proxy, which puts the amplitudes on a similar scale. The motivation for fitting both the *MgII* and *sin ωt* regressors was to illustrate the degree to which the solar-noise affects the amplitudes. With the exception of the middle and very top of the mesosphere where the differences are essentially zero, the differences between the two coefficients are at most 2 K. In regions where there is a significant difference between the *sin ωt* and *MgII* coefficients, the sine amplitudes are systematically lower, shifted toward the left. This holds whether the amplitude is positive or negative and it points to a possible model specification problem. Regions where there is a significant difference between the *MgII* and *sin ωt* coefficients coincide with regions where the solar-noise has its greatest significance in Model (4.1). This suggests that if the solar-noise term is omitted from a temperature model, the omission could result in a biased solar proxy coefficient. When the solar-noise is included in a model containing an *MgII* proxy, the *MgII* coefficients closely match the *sin ωt* coefficients. That is, when the *MgII* coefficients from (A) *y ~ time + AO + SAO + MgII + solnoise* are compared to the

sin ωt coefficients from (B) *y ~ time + AO + SAO + sinωt + solnoise*, they are very similar. The maximum absolute-value difference between them is 0.4 K, and the average absolute value difference is 0.15 K. Even though the *MgII* regressor already contains the solar-noise, some of the effects of the solar-noise are not captured in the regression results if it is not separated from the solar proxy.

Also notice that at most altitudes where the phase difference is ±π/2, the sine proxy goes through zero. This is expected. The attenuation of the solar proxy term is zero where the phase difference is ±π/2. (See Chapter 3 and Appendix G.) For a fixed-phase model the amplitude of the *sin ωt* proxy converges to $\alpha \cdot cos(\rho)$, where ρ is the phase difference between solar input and the atmospheric solar response, and α is the amplitude of the atmospheric solar response.

7.2. Fixed-proxy Comparisons

Comparing our fixed-proxy amplitudes in Figure 4.9(c) to those from the other sites shown in Figure 4.10, one sees obvious differences in the shapes of the vertical profiles. The atmospheric solar response at Ryori Japan (39ºN) shows a near zero amplitude between 20 and 50 km. At Wallops Island, Virginia (37ºN) there is a positive response of 2.5 K to 5 K between 40 and 70 km which is nearly statistically significant at 95%. Temperatures above Shemya, Alaska (52ºN) shows a solar response of −2 K between 40 and 50 km but not at the 95% level. At Vologograd, USSR (48ºN) there is a slightly negative −2 K response at 35 km but it is not statistically significant. Above 55 km there is a 2.7 to 7 K atmospheric solar response at greater than 95% confidence. Primrose Lake, Canada (54ºN) does not show a statistically significant atmospheric solar response above

40 km. The USU temperatures have a sin ωt amplitude of 1 K at 52 km, –4 K at 70 km, and 4 K at 87 km. The maximum magnitude solar responses in Figure 4.10 are from about 2.5 to 7 K and are roughly comparable to the USU maxima and minima and roughly consistent with model predictions. However, the vertical profiles are very dissimilar.

Several researchers have found that the middle atmosphere solar response is stronger during winter than it is during summer [*Schwentek*, 1971; *Hood*, 1987; *Keckhut et al.*, 1995; *Pertsev and Perminov*, 2008]. This has been repeated in simulations by *Hamson et*

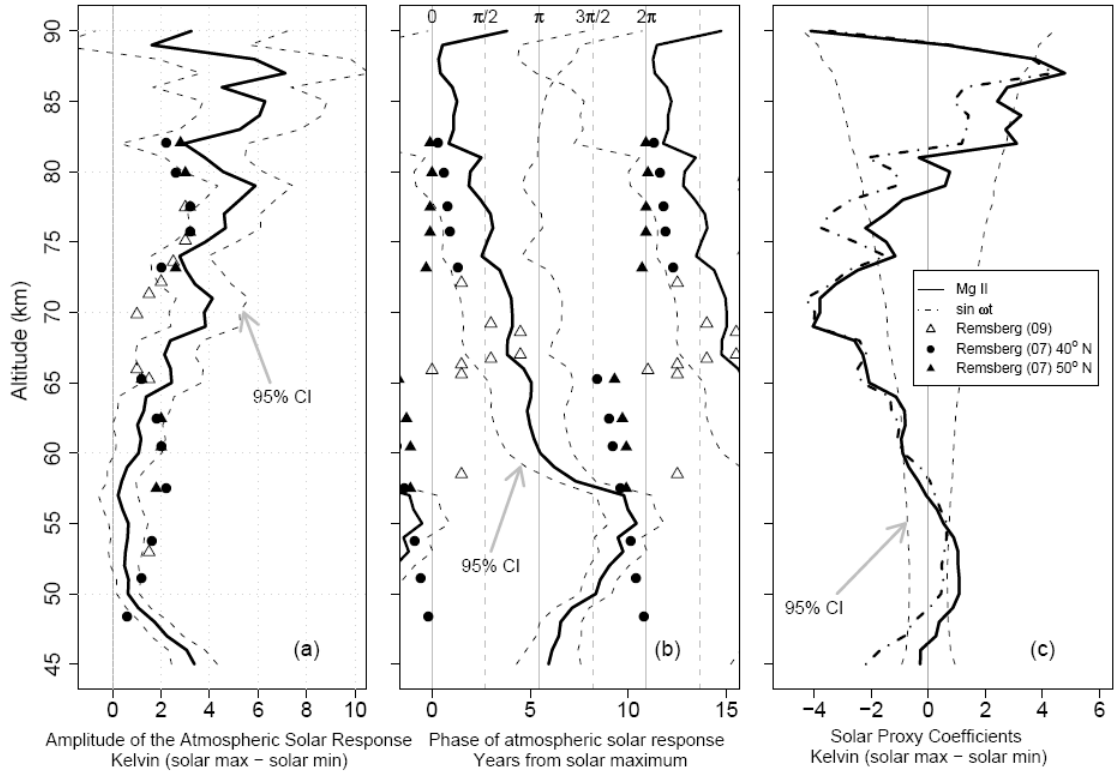


Figure 4.10. Solar response amplitudes and phases. Part (a) shows the amplitude (max – min) of the solar-like proxy α_1. Part (b) shows the phase φ of the solar proxy in years from 1/1/2002. Part (c) shows the amplitude (max – min) of the fixed-phase coefficients α_2 corresponding to *sin ωt*, and α_3 corresponding to the MgII proxy. The circles and open triangles in (a) and (b) are amplitudes and phases from 40ºN adapted from *Remsberg* [2007] and *Remsberg* [2009]; the solid triangles are the same but for 50ºN. The phase data is repeated every 11 years to aid with comparison.

al. [2005] who found a middle atmosphere winter response that is slightly greater than the summer response by about 1 K at 52.5ºN. A stronger winter response was also seen in the French temperatures reported by *Keckhut et al.* who found a maximum solar response during summer of 3 K and a 7 K response during winter.

Figure 4.11 shows the summer and winter amplitude profiles of the atmospheric solar response for $T \sim t + sin\omega t + cos\omega t + solarnoise$ shown in Figure 4.11a, and for the two fixed-proxy models $T \sim t + MgII$ shown in Figure 4.11b, and $T \sim t + sin\omega t + solarnoise$ shown in Figure 4.11c. For the atmospheric solar response from Model (4.1), between 60 and 74 km the summer response is stronger than the winter response, the maximum difference is about 6 K. From 74 to 90 km the winter response is stronger than the summer response, the maximum difference is about 6.5 K. Below 60 km there is at most a 1 K difference between the winter and summer amplitudes for Model (4.1).

Our summer/winter results shown in Figures 4.11b and c show a more complex structure than those from other groups shown in Figure 4.10. It is possible that these differences are due simply to zonal and latitudinal differences in the atmospheric solar response, but because the other groups used fixed proxies, their amplitudes could have been exaggerated by an unaccounted-for atmospheric solar response. The differences between most of the profile shapes of the solar responses shown in Figure 4.10 are big. Only the OHP and Volograd profiles are similar. Our fixed-phase profiles are also considerably different from the variable-phase profiles. Comparing ours to the OHP and Volgograd profiles there are apparent phase differences manifested in the different signs of the fixed-phase proxy coefficient. From 46 to 55 km our maximal amplitude is about 1 K whereas the OHP and Volgograd show a maximal amplitude of −2 to −3 K. In the

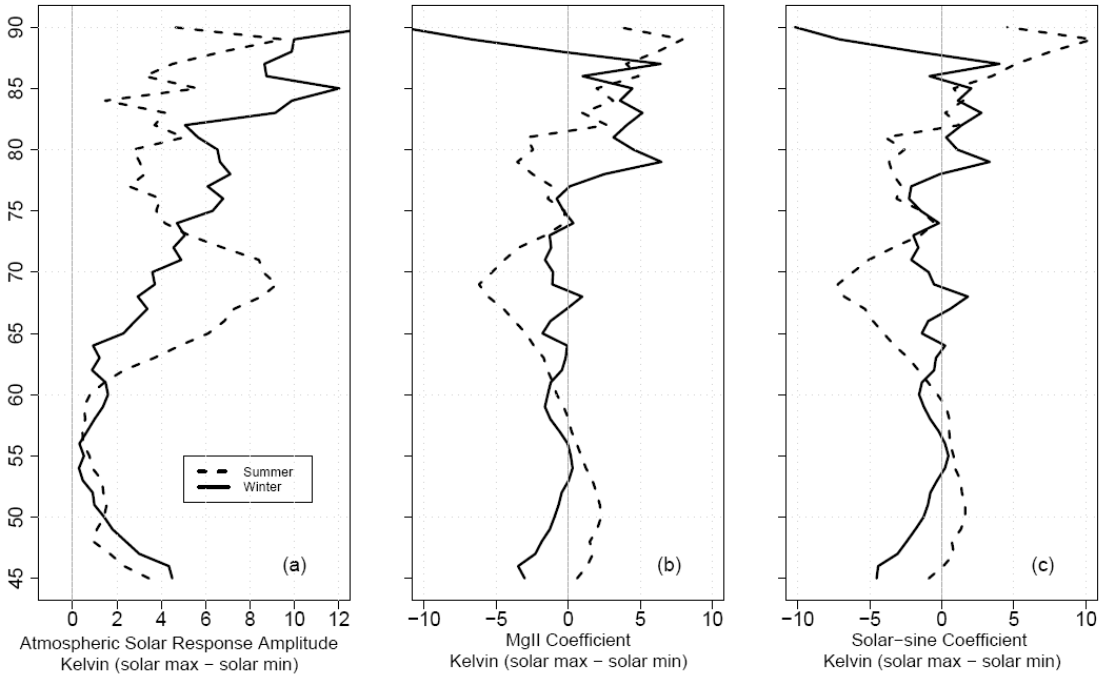


Figure 4.11. Seasonal amplitudes and phases of solar response. The solar response profiles for summer and winter. Figure (a) shows the magnitude (max – min) of the solar-like proxy α_2. Figure (b) shows the magnitude (max – min) of α_3 corresponding to the *MgII* proxy. Figure (c) shows the magnitude (max – min) of the coefficient α_1 corresponding to *sin ωt*. Summer is taken to be from March 21 to September 21 and winter is from September 22 to March 20. Four-month winters and summers were not attempted because of the limited number of data points.

upper half of the mesosphere both the OHP and Volgograd profiles show a positive amplitude of about 4 to 5 K. However, the USU lidar shows a negative response between 55 and 80 km of about −4 K and a positive response between 81 and 88 km of about 4 K. At 90 km it becomes negative again at −4 K.

8. Annual and Semiannual Oscillation

Amplitudes and phases of the annual and semiannual temperatures for the USU data, the French OHP and CEL lidars, the HALOE instrument, and semiannual amplitudes and phases from the SABER instrument are shown in Figure 4.12. The French data are taken

from *Leblanc et al.* [1998]. The HALOE temperatures are from the http://haloe.gats-inc.com. The HALOE data is nominally from the location of the USU lidar. The SABER amplitudes and phases were taken from Figure 4.10 in *Huang et al.* [2006]. The SABER (Sounding of the Atmosphere using Broadband Emission Radiometry) instrument is aboard the TIMED satellite and currently provides middle atmosphere temperature measurements.

8.1. Annual Oscillation

The annual oscillation is well known. Below 65 km the familiar warm summer and cold winter seasons in the lower mesosphere are caused by radiative effects related to the tilt of the Earth's axis as it orbits the sun. Above 65 km the annual oscillation dramatically shifts phase causing the upper mesosphere to experience warm winters and cold summers. This phase shift is caused by the effects of gravity wave activity and the Coriolis effect, which creates a summer to winter meridional flow. Because of mass continuity, the flow expands and cools adiabatically as it rises through the summer mesosphere. In the winter hemisphere the downwelling flow undergoes adiabatic compression and heating as it descends through the mesosphere. The upwelling air also has higher concentrations of CO_2 than the air near and above the mesopause level, which further enhances radiative cooling at the summer mesopause, and the downwelling air in the winter mesopause is depleted in CO_2 which depresses the cooling rate [*Garcia and Solomon*, 1985; *Smith*, 2004].

Below 80 km the USU annual oscillation amplitudes follow approximately those from the HALOE temperatures. Between 64 and 70 km our AO is slightly greater than

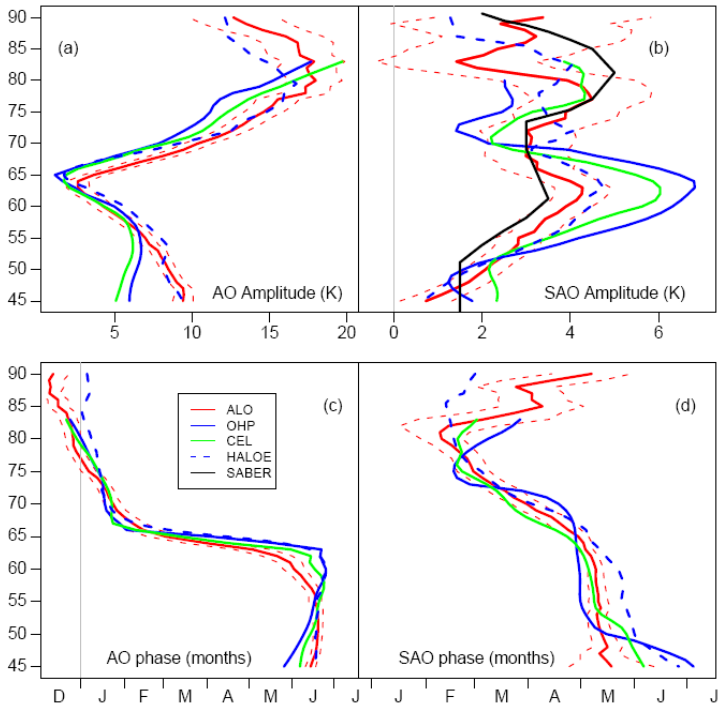


Figure 4.12. AO and SAO amplitudes and phases. The annual and semiannual amplitudes and phases from the USU, OHP, CEL, and HALOE temperatures. SABER SAO amplitudes and phases are also shown.

the HALOE AO, with a maximum difference of about 1 K at 65 km. From 65 to 80 km they follow each other very closely. Above 80 km our AO is ~2 K greater than the HALOE AO. From 45 to 50 km they follow each other very closely. From 50 to 65 km our AO is smaller by about 1 K.

From 55 to 80 km the French AO profiles follow the same profile shape as the USU profile, though the French amplitudes are about 2 K less than ours between 65 and 77 km and 2 K greater between 73 and 78 km. The French amplitudes between 45 and 55 km vary from 5 to 6 K. This is in contrast to both the USU and the HALOE AO, which have an amplitude of 10 K at 45 km, decreasing to 6 K at 55 km. Although it is not shown, the MSISe00 model closely follows the USU temperatures between 45 and 55 km.

The AO phase profiles follow a similar overall pattern. Below 60 km the temperature maximum is around June. From 60 to 65 km it transitions rapidly to January and then slowly moves toward December with increasing altitude. The USU phase shift is more gradual than the OHP, CEL, and HALOE phase shifts. Overall, the phases of the AOs shown in Figure 4.12c are similar, as a whole they vary at most by one month between 45 and 60 km and between 65 to 90 km. Between 60 and 65 km the phase shifts dramatically. In this region the phases can differ by several months but this is due to the flatness of the slope, so the variation is probably best understood as local altitude variations rather than phase variations.

8.2. Semiannual Oscillation

The semiannual oscillation (SAO) is a seasonal, twice-yearly oscillation in middle atmosphere temperatures and winds. It was previously thought that the origin of the SAO was the semiannual variation in solar irradiance: The solar zenith angle reaches a maximum twice every year at the equator. However, according to *Reed* [1966] this could not account for the westerly acceleration of the semiannual zonal wind. (See also *Dunkerton* [1979].) It is now believed that the stratosphere westerly accelerations in the wind field are caused by Kelvin waves and the easterly accelerations are caused by momentum advection across the equator. The mesopause SAO is believed to be caused by gravity wave breaking between 80 and 90 km associated with the easterly and westerly wind accelerations; it is also believed that the amplitude of the SAO is modulated by the quasi-biennial oscillation [*Dunkerton*, 1979; *Garcia et al.*, 1997; *Brasseur and Solomon*, 2005; *Watanabe and Takahashi*, 2005].

There is much larger variability between the SAO amplitude profiles than there are in the AO amplitude profiles. For the USU, SABER, CEL and OHP amplitudes, there is a maximum near 60-65 km and a second maximum between 70 and 85 km. The minima are located at 45-50 km, 65-70 km, and 80-90 km. The HALOE profile appears to be a few kilometers higher than the others. The biggest differences are with the French amplitudes at the lower-altitude maxima (60-65 km) where the USU, SABER, and HALOE amplitudes are smaller by about 2 to 3 K. This altitude is also where the SAO amplitude is greater than the amplitude of the AO by a few Kelvin. It is also where the AO amplitude is at a minimum and where the AO phase makes a rapid π phase change.

The SAO phase profiles also follow similar overall patterns. Because of the way the phases were calculated, both the earlier phase (April/May at 45 km) and the later phase (December/January at 45 km) are identical with the exception that the earlier phase is about six months earlier (or later) than the first. We therefore discuss only the first phase of the SAO. At 45 km the temperature maximum is about one month before solstice. It is gradually pushed back to February at 80 km, which puts the upper mesosphere SAO nearly out of phase with the lower mesosphere SAO. Above this it starts to move later in the year until at 90 km it occurs around April-May. There is fairly good agreement among the five profiles between 55 and 80 km. Below 55 km our SAO phase occurs around mid May, whereas the others are closer to solstice. Above 80 km, our phase occurs later in the year than the SABER and HALOE phases do. The HALOE semiannual phase is about two months earlier, and the SABER phase is about 15 days to one month earlier.

8.3. Semiannual Climatology

The climatologies of the semiannual oscillation for the HALOE and USU temperatures are made by removing the intercept, linear trend, annual oscillation, solar-like sin ωt and cos ωt variations, and the solar-noise terms from the data, leaving only the semiannual and residual components. These were made into a composite year and smoothed with a 61 day moving average. The USU SAO climatology is shown in Figure 4.13. The HALOE climatology is shown in Figure 4.14.

At 45 km for both the USU and HALOE SAO climatologies the temperature maxima occur at April-May and December and move toward earlier in the year with increasing altitude, up to 80 km. Above 80 km the USU phase moves toward later in the year with increasing altitude, but the HALOE climatology does not clearly show this reversal. Below 80 km the second-phase amplitudes in the USU climatology are stronger than the first by about 2-4 K. This is also true for the HALOE climatology. Above 80 km, in the USU climatology, the first phase is stronger than the second by about 2 K. In the HALOE climatology the second-phase amplitudes are stronger than the first throughout the altitude range and above 80 km the temperature differences between the first and second phases are about 2-4 K. This is greater than the difference in the USU climatology.

There is also a weakening of the SAO amplitude at 65 km that occurs in April and again at 73 km during August-September. This weakening is also found in the HALOE climatology at 67 km in April and at 73 km in August-September. Both climatologies also show a 1 K warming that starts at 45 km during September, two to three months before the second phase, and penetrates up to 60 km during October. This September-October warming and the difference in temperature between the two maxima are likely

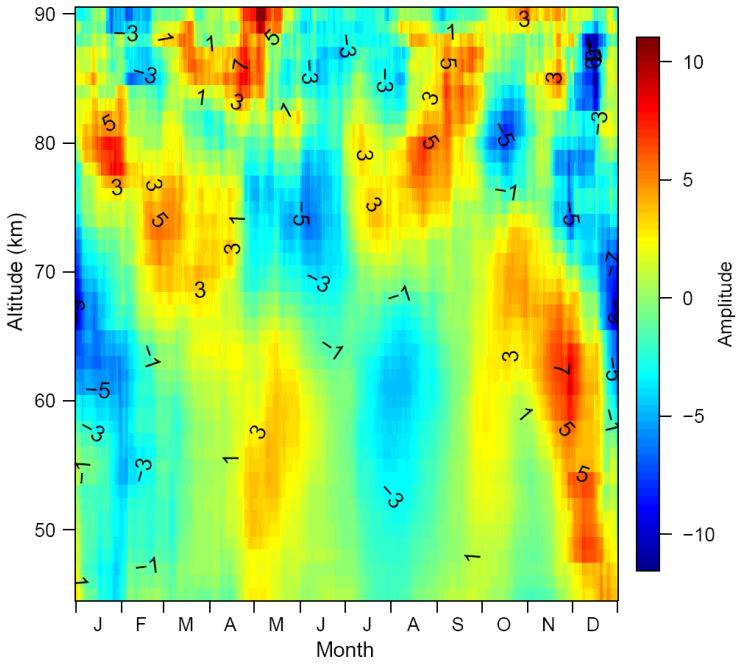


Figure 4.13. USU SAO climatology as seen over Logan, Utah.

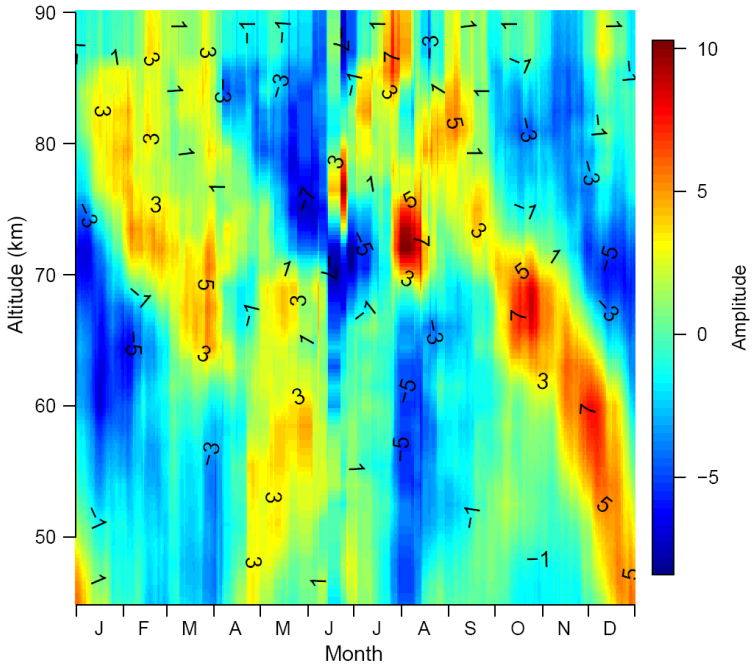


Figure 4.14. A HALOE SAO climatology. Derived from the HALOE data from the approximate location of the USU Rayleigh lidar.

the effect of higher order terms that were not included in our model.

9. Final Conclusions

While there are several factors that may contribute to the large upper mesosphere cooling rate found in our temperatures, it does not appear to be an artifact of the model specification or coefficient bias. One possibility for the large cooling trend is the influence that the Mt. Pinatubo eruption may have had on the mesopause temperatures. There is independent evidence that the eruption altered the mesopause temperature structure, and because the USU lidar came online at about the time the temperature perturbation was at its peak, this seems like a plausible explanation. The temperatures from the first years exert some leverage on the linear coefficient which is suggestive of elevated temperatures which lead to a large cooling rate. There also appears to be elevated temperatures in the residuals at 85 km, but this could be an unaccounted for solar-like response. Because we do not have data before the Pinatubo eruption, it is difficult to assess if these elevated temperatures are a result of a Pinatubo effect, an unaccounted-for solar-like oscillation, or both. At 45 km there are depressed temperatures in our early data, which might be related to the Pinatubo eruption; but owing to the fact that we do not have prior data this, too, cannot be assessed with any certainty. Another possibility is collinearity, or more specifically, coefficient correlation, but the presence of this is also difficult to determine.

All the linear trend profiles from the various model perturbations fell within the error limits for the Model (4.1) linear term, suggesting the value of the linear trend coefficient is not strongly affected by the solar-like sine and cosine terms. Inclusion of

the solar-like sine and cosine terms nearly doubles the linear trend error limits, but omitting them can bias the linear trend estimate.

When the solar-like terms were omitted from Model (4.1) the cooling rate was −0.3 to −1.0 K/year between 67 and 80 km at the 95% level. When the solar-like terms are included the cooling rate was −0.5 to −1.0 K/year between 73 and 80 km. On average the mesosphere cooling trends tabulated in *Beig et al.* [2003] are about −0.35 K/year, but four of the trends were in the neighborhood of −1 K/year. Our most reliable linear trend estimates are between 70 and 80 km.

Additional data would allow us to avoid the Pinatubo years without the detrimental effects that arise from shortening the data set. Also, collinearity could be affecting our linear trend estimate such that our linear trend values are larger or smaller than the true value. Given that our linear trend estimate is much larger than what other researchers have found, it seems more likely that if coefficient correlation is affecting our linear trend estimate, it is larger than the true value. This problem would be diminished by additional data.

The amplitude of the atmospheric solar response ranges from 3 K at 45 km to 5 K between 85 and 90 km. *Remsberg's* mesosphere phases between 66 and 82 km are within or nearly within our error limits. In the lower mesosphere there is little correlation between *Remsberg's* phases and ours. To my knowledge, *Remsberg* is the only other group to have allowed for a phase offset to the atmospheric solar response. Our amplitudes are within ±2 K to *Remsberg's* amplitudes.

When a fixed-phase model is applied to our temperatures the sin ωt coefficient is systematically lower than the MgII coefficient by about 1 K in the lower and upper

mesosphere. These differences coincide with the areas where the solar-noise term is statistically significant. When the solar-noise is added to a model containing an MgII proxy, the proxy coefficient closely matches the sin ωt coefficient. It is possible that omitting the solar-noise term from the model biases the MgII coefficient estimate.

Other groups that employed a fixed-phase proxy have found a maximum solar response magnitude ranging from 2.5 to 7 K, which is roughly consistent with model predictions. The fixed-phase models applied to the USU data exhibits a seasonal response. It shows amplitude magnitudes of 1 and ±4 K. The USU data shows that the middle mesosphere (60-74 km) solar response during summer is stronger than the winter response by about 6 K. In the upper mesosphere (74-88 km) this was reversed, but the maximum difference was still about 6.5 K. Other groups found a stronger atmospheric solar response only during winter.

Between 45 and 80 km our AO amplitudes and phases agree with the AO amplitudes and phases from the HALOE data to within a few Kelvin. Above 55 km our amplitudes follow the OHP and CEL amplitudes, but are larger by about 2 K. Below 55 km the OHP and CEL lidars show an amplitude between 5-6 K, whereas our amplitude starts at 10 K at 45 km and decreases to 6 K at 55 km. Our AO phases closely follow the phases found for the OHP, CEL, and HALOE, with the minor exception that our AO phase has a more gradual transition at 65 km.

Our SAO amplitudes show a minimum at 45 km, a maximum near 62 km, another minimum near 70 km, a second maximum near 78 km, and a minimum at 90 km. All the SAO amplitude profiles have this double peak profile. The HALOE semiannual amplitude profile appears to be shifted a few kilometers higher than the others.

The SAO phases also have a similar profile. At 45 km the USU SAO phase is about a month before solstice, around mid May. The phase moves earlier in the year with increasing altitude until 80 km, where it is around February. Above this the phase moves toward later in the year.

A climatology of the USU SAO shows that the amplitude of the second phase of the SAO is stronger than the first by about 2-4 K. A similar stronger second phase is also seen in the HALOE climatology. There is also a weakening of the first phase of the SAO at 65 km that occurs April-May, and again during the second phase at 75 km around September-October. This is also seen in the HALOE climatology. The USU climatology shows that near the mesopause the amplitude of the first SAO phase is stronger than the second phase; for the HALOE climatology the second phase is stronger than the first. There is also a temperature anomaly that starts at 45 km during September and persists to 65 km in October. This feature is about 1 K and is found in both the HALOE and USU climatologies. It, and the amplitude difference between the first and second phases, are a likely indication of the existence of higher order terms in the temperature time series.

References

Aikin, A. C., M. L. Chanin, J. Nash, and D. J. Kendig (1991), Temperature trends in the lower mesosphere, *Geophys. Res. Lett.*, *18*(3), 416–419, doi:10.1029/91GL00233.

Beig, G., et al. (2003), Review of mesospheric temperature trends, *Rev. Geophys.*, *41*(4), doi:10.1029/2002RG000121. [online] Available from: http://www.agu.org/pubs/crossref/2003/2002RG000121.shtml

Bittner, M., D. Offermann, H. Graef, M. Donner, and K. Hamilton (2002), An 18-year time series of OH rotational temperatures and middle atmosphere decadal variations, *J. Atmos. Sol. Terr. Phys.*, *64*(8-11), 1147-1166, doi:10.1016/S1364-6826(02)00065-2.

Brasseur, G., and S. Solomon (2005), *Aeronomy of the Middle Atmosphere*, 3rd ed., Springer, Dordrecht, The Netherlands.

Calisesi, Y., and K. Matthes (2007), The middle atmospheric ozone response to the 11-year solar cycle, in *Solar Variability and Planetary Climates*, vol. 23, pp. 273-286. [online] Available from: http://dx.doi.org/10.1007/978-0-387-48341-2_22 (Accessed 4 February 2010).

Callendar, G. S. (1938), The artificial production of carbon dioxide and its influence on temperature, *Quarterly J. Royal Meterol. Soc.*, *64*(275), 223-240, doi:10.1002/qj.49706427503.

Donnelly, R. F. (1991), Solar UV spectral irradiance variations, *J. Geomagn. Geoelectr.*, *43 Suppl.*, 835-842.

Donnelly, R. F., D. F. Heath, and J. L. Lean (1982), Active-region evolution and solar rotation variations in solar UV irradiance, total solar irradiance, and soft X rays, *J. Geophys. Res.*, *87*(A12), 10318-10324, doi:10.1029/JA087iA12p10318.

Dunkerton, T. (1979), On the role of the kelvin wave in the westerly phase of the semiannual zonal wind oscillation, *J. Atmos. Sci.*, *36*(1), 32-41.

Fomichev, V. I., A. I. Jonsson, J. de Grandpré, S. R. Beagley, C. McLandress, K. Semeniuk, and T. G. Shepherd (2007), Response of the middle atmosphere to CO2 doubling: results from the canadian middle atmosphere model, *J. Climate*, *20*(7), 1121, doi:10.1175/JCLI4030.1.

Froidevaux, L., M. Allen, S. Berman, and A. Daughton (1989), The mean ozone profile and its temperature sensitivity in the upper stratosphere and lower mesosphere: an analysis of LIMS observations, *J. Geophys. Res.*, *94*(D5), 6389–6417.

Garcia, R. R., and S. Solomon (1985), The effect of breaking gravity waves on the dynamics and chemical composition of the mesosphere and lower thermosphere, *J. Geophys. Res.*, *90*(D2), 3850–3868.

Garcia, R., S. Solomon, R. Roble, and D. Rusch (1984), A numerical response of the middle atmosphere to the 11-year solar cycle, *Planet. Space Sci.*, *32*(4), 411-423, doi:10.1016/0032-0633(84)90121-1.

Garcia, R. R., T. J. Dunkerton, R. S. Lieberman, and R. A. Vincent (1997), Climatology of the semiannual oscillation of the tropical middle atmosphere, *J. Geophys. Res.*, *102*(D22), 19-26.

Golitsyn, G. S., A. I. Semenov, N. N. Shefov, L. M. Fishkova, E. V. Lysenko, and S. P. Perov (1996), Long-term temperature trends in the middle and upper atmosphere,

Geophys. Res. Lett., *23*(14), 1741, doi:10.1029/96GL01592.

Hampson, J., P. Keckhut, A. Hauchecorne, and M. Chanin (2005), The effect of the 11-year solar-cycle on the temperature in the upper-stratosphere and mesosphere: Part II numerical simulations and the role of planetary waves, *J. Atmos. Sol. Terr. Phys.*, *67*(11), 948-958, doi:10.1016/j.jastp.2005.03.005.

Held, I. M., and B. J. Soden (2000), Water vapor feedback and global warming, *Annu. Rev. Energy. Environ.*, *25*(1), 441-475, doi:10.1146/annurev.energy.25.1.441.

Hood, L. L. (1987), Solar ultraviolet radiation induced variations in the stratosphere and mesosphere, *J. Geophys. Res.*, *92*(D1), 876-888, doi:10.1029/JD092iD01p00876.

Huang, T., and G. Brasseur (1993), Effect of long-term solar variability in a two-dimensional interactive model of the middle atmosphere, *J. Geophys. Res.*, *98*(D11), 20413–20427.

Huang, F., H. Mayr, C. Reber, J. Russell, M. Mlynczak, and J. Mengel (2006), Stratospheric and mesospheric temperature variations for the quasi-biennial and semiannual (QBO and SAO) oscillations based on measurements from SABER (TIMED) and MLS (UARS), *Ann. Geophys.*, *24*(8), 2131-2149, doi:10.5194/angeo-24-2131-2006.

Jonsson, A. I., J. D. Grandpré, V. I. Fomichev, J. C. McConnell, and S. R. Beagley (2004), Doubled CO2-induced cooling in the middle atmosphere: Photochemical analysis of the ozone radiative feedback, *J. Geophys. Res.*, *109*(D24103), doi:10.1029/2004JD005093. [online] Available from: http://www.agu.org/pubs/crossref/2004.../2004JD005093.shtml (Accessed 5 February 2010).

Keckhut, P., A. Hauchecorne, and M. L. Chanin (1995), Midlatitude long-term variability of the middle atmosphere: Trends and cyclic and episodic changes, *J. Geophys. Res.*, *100*(D9), 18887–18897.

Khosravi, R., G. Brasseur, A. Smith, D. Rusch, S. Walters, S. Chabrillat, and G. Kockarts (2002), Response of the mesosphere to human-induced perturbations and solar variability calculated by a 2-D model, *J. Geophys. Res.*, *107*(D18), 4358.

Leblanc, T., I. S. McDermid, P. Keckhut, A. Hauchecorne, C. Y. She, and D. A. Krueger (1998), Temperature climatology of the middle atmosphere from long-term lidar measurements at middle and low latitudes, *J. Geophys. Res.*, *103*(D14), 17191–17204.

Marsh, D. R., R. R. Garcia, D. E. Kinnison, B. A. Boville, F. Sassi, S. Solomon, and K. Matthes (2007), Modeling the whole atmosphere response to solar cycle changes

in radiative and geomagnetic forcing, *J. Geophys. Res.*, *112*(D23), doi:10.1029/2006JD008306. [online] Available from: http://www.agu.org/pubs/crossref/2007/2006JD008306.shtml.

McCormick, M. P., and R. E. Veiga (1992), SAGE II measurements of early Pinatubo aerosols, *Geophys. Res. Lett.*, *19*(2), 155–158.

Pertsev, N., and V. Perminov (2008), Response of the mesopause airglow to solar activity inferred from measurements at Zvenigorod, Russia, *Ann. Geophys.*, *26*(5), 1049-1056, doi:10.5194/angeo-26-1049-2008.

Ramaswamy, V., et al. (2001), Stratospheric temperature trends: Observations and model simulations, *Rev. Geophys.*, *39(1)*, 71-122.

Reed, R. J. (1966), Zonal wind behavior in the equatorial stratosphere and lower mesosphere, *J. Geophys. Res.*, *71*(18), 4223-4233, doi:10.1029/JZ071i018p04223.

Reinsel, G. C., A. J. Miller, E. C. Weatherhead, L. E. Flynn, R. M. Nagatani, G. C. Tiao, and D. J. Wuebbles (2005), Trend analysis of total ozone data for turnaround and dynamical contributions, *J. Geophys. Res.*, *110*(D16), doi:10.1029/2004JD004662. [online] Available from: http://www.agu.org/pubs/crossref/2005/2004JD004662.shtml.

Reinsel, G. C., E. C. Weatherhead, G. C. Tiao, A. J. Miller, R. M. Nagatani, D. J. Wuebbles, and L. E. Flynn (2002), On detection of turnaround and recovery in trend for ozone, *J. Geophys. Res.*, *107*(D10), doi:10.1029/2001JD000500. [online] Available from: http://www.agu.org/pubs/crossref/2002/2001JD000500.shtml.

Resin, E., and J. Scheer (2002), Searching for trends in mesopause region airglow intensities and temperatures at El Leoncito, *Physics and Chemistry of the Earth, Parts A/B/C*, *27*(6-8), 563-569, doi:10.1016/S1474-7065(02)00038-4.

Remsberg, E. E. (2002), Seasonal and longer-term variations in middle atmosphere temperature from HALOE on UARS, *J. Geophys. Res.*, *107*(D19), doi:10.1029/2001JD001366. [online] Available from: http://www.agu.org/pubs/crossref/2002/2001JD001366.shtml.

Remsberg, E. E. (2007), A reanalysis for the seasonal and longer-period cycles and the trends in middle-atmosphere temperature from the Halogen Occultation Experiment, *J. Geophys. Res.*, *112*(D9), doi:10.1029/2006JD007489. [online] Available from: http://www.agu.org/pubs/crossref/2007/2006JD007489.shtml.

Remsberg, E. E. (2008), On the observed changes in upper stratospheric and mesospheric temperatures from UARS HALOE, *Ann. Geophys.*, *26*(5), 1287-1297.

Remsberg, E. E. (2009), Trends and solar cycle effects in temperature versus altitude from the Halogen Occultation Experiment for the mesosphere and upper stratosphere, *J. Geophys. Res.*, *114*(D12), doi:10.1029/2009JD011897. [online] Available from: http://www.agu.org/pubs/crossref/2009/2009JD011897.shtml.

Remsberg, E. E., and L. Deaver (2005), Interannual, solar cycle, and trend terms in middle atmospheric temperature time series from HALOE, *J. Geophys. Res.*, *110*(D6), doi:10.1029/2004JD004905. [online] Available from: http://www.agu.org/pubs/crossref/2005/2004JD004905.shtml.

Rind, D., D. Shindell, P. Lonergan, and N. K. Balachandran (1998), Climate change and the middle atmosphere. Part III: The doubled CO2 climate revisited, *J. Climate*, *11*(5), 876–894.

Rind, D., R. Suozzo, N. K. Balachandran, and M. J. Prather (1990), Climate change and the middle atmosphere. Part I: The doubled CO2 climate, *J. Atmos. Sci.*, *47*(4), 475-494, doi:10.1175/1520-0469(1990)047<0475:CCATMA>2.0.CO;2.

Robel, R. G., R. E. Dickinson (1989), How will changes in carbon dioxide and methane modify the mean structure of the mesosphere and troposphere?, *Geophys. Res. Lett.*, *16*(*12*), 1441-1444.

Schmidt, H., G. Brasseur, M. Charron, E. Manzini, M. A. Giorgetta, T. Diehl, V. I. Fomichev, D. Kinnison, D. Marsh, and S. Walters (2006), The HAMMONIA chemistry climate model: sensitivity of the mesopause region to the 11-year solar cycle and CO2 doubling, *J. Climate*, *19*(16), 3903-3931.

Schwentek, H. (1971), The sunspot cycle 1958/70 in ionospheric absorption and stratospheric temperature, *J. Atmos. Terr. Phys.*, *33*(12), 1839-1852, doi:10.1016/0021-9169(71)90163-2.

Semenov, A. I., N. N. Shefov, E. V. Lysenko, G. V. Givishvili, and A. V. Tikhonov (2002), The season peculiarities of behaviour of the long-term temperature trends in the middle atmosphere on the mid-latitudes, *Phys. Chem. Earth, Parts A/B/C*, *27*(6-8), 529-534, doi:10.1016/S1474-7065(02)00034-7.

She, C., S. W. Thiel, and D. A. Krueger (1998), Observed episodic warming at 86 and 100 km between 1990 and 1997: Effects of Mount Pinatubo eruption, *Geophys. Res. Lett.*, *25*(4), 497, doi:10.1029/98GL00178.

Smith, A. (2004), Physics and chemistry of the mesopause region, *J. Atmos. Sol. Terr. Phys.*, *66*(10), 839-857, doi:10.1016/j.jastp.2004.01.032.

Thomas, M. A., C. Timmreck, M. A. Giorgetta, H. F. Graf, and G. Stenchikov (2009), Simulation of the climate impact of Mt. Pinatubo eruption using ECHAM5 – Part

1: Sensitivity to the modes of atmospheric circulation and boundary conditions, *Atmos. Chem. Phys.*, *9*, 757-769.

Watanabe, S., and M. Takahashi (2005), Kelvin waves and ozone Kelvin waves in the quasi-biennial oscillation and semiannual oscillation: A simulation by a high-resolution chemistry-coupled general circulation model, *J. Geophys. Res.*, *110*, 9 PP., doi:200510.1029/2004JD005424.

CHAPTER 5

EXAMINATION OF THE MODEL RESIDUALS

Abstract. When doing ordinary least squares regression (OLSR), one common assumption is constant residual standard deviation. But for atmospheric temperatures, this assumption is rarely true. In this chapter we examine the residuals and some of their variations. We examine some of the variations in the residual standard deviation, the residual autocorrelation, the 11-year variation in the residual standard deviation, and seasonal variations of the solar-noise coefficient. As part of this, we examine climatologies of the autocorrelation coefficient and solar-noise coefficient. We attribute the seasonal variation in the autocorrelation coefficient to seasonal temperature perturbations that evolve over a period of a day or more. We briefly examine the downward control principle described in *Bittner et al.* [2002]. We did not find evidence for this principle in our data, but owing to the shortness of our data set this was inconclusive.

1. Introduction

If the standard deviation of the model residuals is constant over time they are said to be homoscedastic. If they change over time they are said to be heteroscedastic. Coefficient standard errors (SEs) are based on the assumption of constant residual standard deviation. Ideally the residuals are independent and identically distributed, the i.i.d. assumption. Identically distributed means every residual is taken from the same distribution and consequently has the same standard deviation; independent means there is no correlation between residuals. The difficulty as it applies here is that with

atmospheric temperatures that assumption is rarely true. The residuals are strongly heteroscedastic. The standard deviations are greater during winter and evolve with the solar cycle; the residuals are autocorrelated and the correlation follows a seasonal cycle.

2. Section Summary

Section 3 presents the model from which the residuals are taken. Section 4 contains a brief description of the number of nights of data and how they vary with altitude and season. In section 5 we discuss the seasonal variation in residual standard deviation and present a plot of the seasonal standard deviation and discuss it in connection with gravity wave activity. In section 6 we look at the seasonal change of the solar-noise coefficient. In section 7 we look at the residual autocorrelation and present an image plot of the seasonal variation in the autocorrelation. In section 8 we look at the 11-year variation in the residual standard deviation in conjunction with the downward control principle.

3. The Model

The model from which the residuals are drawn is

$$T(z,t) = \alpha(z) + \beta(z)\cdot t + A_1(z)cos(2\pi\cdot t) + A_2(z)sin(2\pi\cdot t) + B_1(z)cos(4\pi\cdot t) + B_2(z)sin(4\pi\cdot t) + C_1(z)sin(\omega t) + C_2(z)cos(\omega t) + D(z)\cdot solnoise + \varepsilon(z,t), \quad (5.1)$$

where α is the intercept, β is the linear trend coefficient, A_1 and A_2 give the amplitude and phase of the annual oscillation, B_1 and B_2 give the same for the semiannual oscillation, C_1 and C_2 give the amplitude and phase of the atmospheric solar response, D is the solar-noise coefficient, and ε is the residual.

4. Number of Data Points

The number of nights is one of the factors that determine the coefficient SEs. More data generally mean smaller SEs, which permit a more accurate appraisal of the significance of the coefficient values. Shown in Figure 5.1 are the number of nights of data by season and altitude, as well as the total number of nights. For each three-month season the number of data points is fairly constant up to about 80 km where it starts to drop off. The drop off is due to observing conditions and lower power. When conditions are very good the lidar can detect signals up to 100 km, which permits a higher starting altitude in the temperature reduction. But observing conditions are rarely that good and on many nights the temperature integration must start from a lower altitude. The seasonal differences are set by a combination of operator availability and weather conditions. We have twice as many observations during summer as during spring and winter.

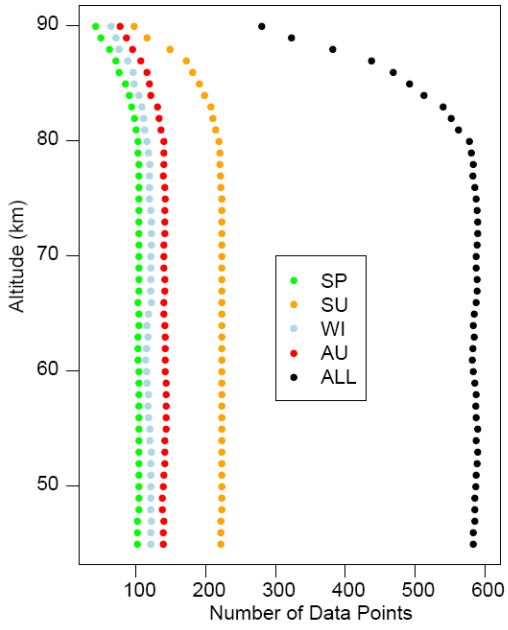


Figure 5.1. The total number of observed nights according to altitude and season.

For the analysis in this dissertation, all temperatures greater than 3σ from the mean for a given altitude and time series were considered outliers and omitted from the data set. Nine data points were omitted at 45 km leaving 584 data points. Only two data points were omitted at 80 km and one data point was omitted at 90 km.

5. Seasonal Changes in Temperature Variability

Between 45 and 75 km, there is the characteristic high standard deviation during winter and low standard deviation during summer. The likely cause of this seasonal difference is variable gravity and planetary wave activity, which follows an annual variation, being strongest in winter. According to theory, gravity wave amplitudes increase proportionally to $p_0^{-1/2}$, where p_0 is the background pressure. As the wave propagates upward, the wave amplitude grows until it reaches a region where wave instabilities cause it to break and deposit its energy. This condition can be characterized by a low Brunt-Väisälä frequency and/or high vertical wind shear. These conditions can be combined in what is sometimes called the gradient Richardson number or the local Richardson number, defined as

$$Ri = N^2/S^2 = N^2/ \left((\partial u/\partial z)^2 + (\partial v/\partial z)^2 \right),$$

where N^2 is the Brunt-Väisälä frequency, u and v the are zonal and meridional winds, and S is the total vertical wind shear. This number is a ratio of the convective instability represented by N^2 and the dynamic instability (shear) represented by S^2. When the Richardson is $0 > Ri < 1/4$ (when N^2 is small and/or S^2 is large) the atmosphere is considered to be dynamically unstable and causes gravity wave breaking [*Hines*, 1971].

The Richardson number determines to what altitude the waves will propagate. When they break, they deposit their energy affecting local temperature, composition, and winds.

Though the cause of the seasonal variation is not well understood, there are three mechanisms that are believed to play a role. (1) A seasonal variation in the gravity wave sources such as orography, convective activity such as weather fronts and thunder storms, and wind shear. (2) The initial properties of gravity waves may vary seasonally according to phase speed, direction of propagation, and background flow. And (3), the atmosphere could act as a selective filter blocking the vertical propagation of lower altitude waves differently in different seasons [*Wilson et al.*, 1991]. The cause of the variation is an open question. *Rauthe et al.* [2008] did not find a seasonal variation in gravity wave wavelengths or a correlation between the strength of gravity wave activity and wind direction or wind speed, nor did they find a correlation between a change in wind direction and the strength of gravity wave activity, item (3). However, as *Rauthe et al.* pointed out, the tilted propagation of gravity waves means that local vertical lidar measurements might not be influenced by the local wind field. *Kafle* [2009] did not find a seasonal variation in vertical wavelength, but was able to infer the presence a seasonal variation in horizontal wavelength. Though the cause of the seasonal variation is not yet settled, the vital role of gravity waves in determining atmospheric circulation, variability, and composition is well established.

The USU mean standard deviations were calculated by forming the Model (5.1) residuals into a composite. The standard deviation was then calculated using a 30-day window. This result includes variations from geophysical variability and photon counting. The USU standard deviation plot is shown in Figure 5.2a. Figure 5.2b is a

similar looking standard deviation plot of Figure 5.2b from *Hauchecorne et al.* [1991] and is on the same color scale as Figure 5.2a. Figure 5.3 is a more detailed contour plot from *Hauchecorne et al*. of Figure 5.2b. The French standard deviations are taken from the temperatures themselves which were smoothed with a 30-day Hanning window. However, for the level of detail that we are interested in, and, given that both we and the French were looking at a 30-day window, the differences in technique are minimal.

Overall, the USU standard deviations are similar to those in Figure 5.2b and Figure 5.3. The first thing to notice is below 75 km there is the characteristic high standard deviation during winter and low standard deviation during summer. In both the USU and French climatologies the summer lower-mesosphere standard deviations are 2 to 4 K and the winter standard deviations are 5 to 8 K. The middle mesosphere (60 to 75 km) summer standard deviations are about 4 to 7 K and the winter standard deviations are about from 6 to 12 K. Above 75 km there is little seasonal variation. A seasonal variation in standard deviation was also found by *Bittner et al.* [2002] at 1 hPa (~47 km; 50ºN/5ºE), who found a winter maximum of 11 K and a summer minimum of 1.5 K in SSU temperatures. They found this pattern was repeated in radiosonde (48ºN/11ºE) measurements from 31 km where the winter standard deviation is 8 K and the summer standard deviation is about 2 K. At 87 km they found a constant standard deviation in the OH bands (51ºN/7ºE) of 5.5 K, which showed no seasonal variation.

Another short-term standard deviation plot from *Rauthe et al.* [2008] is shown in Figure 5.4 and shows the seasonal temperature fluctuations obtained from two Leibniz-Institute of Atmospheric Physics lidars (54ºN). Their climatology shows the characteristic seasonal variation below 75 km. The lower mesosphere standard deviations are about 2 K

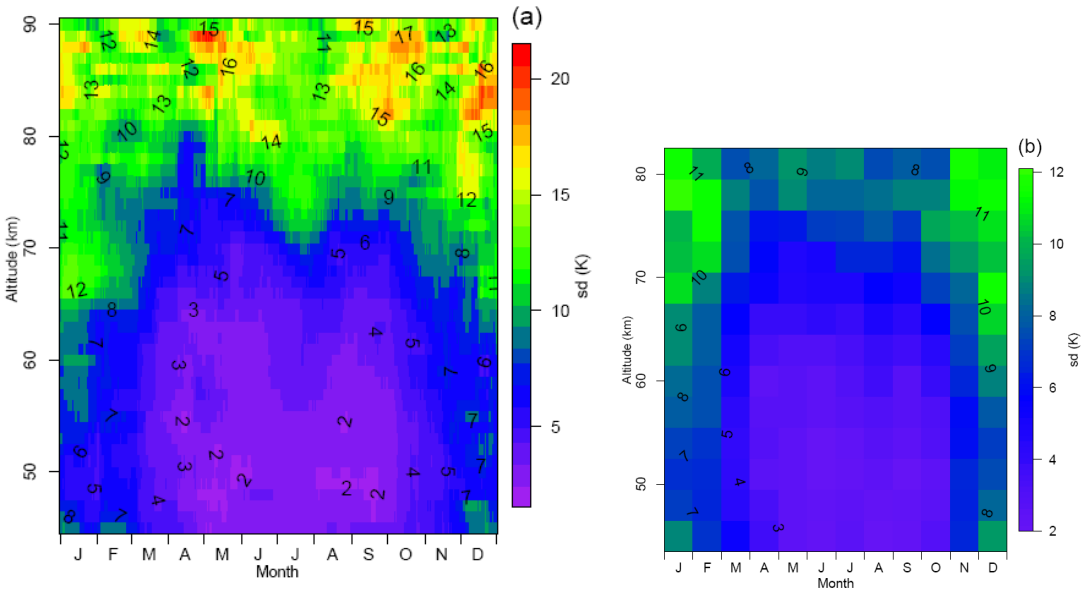


Figure 5.2. Seasonal variation of residual standard deviation. Plot (a) shows the climatology of the standard deviation of the residuals from Model (5.1). Plot (b) shows the climatology of the standard deviation from Table 4 of *Hauchecorne et al.* [1991] covering a period from 1984 to 1989. Both (a) and (b) are on the same color scale.

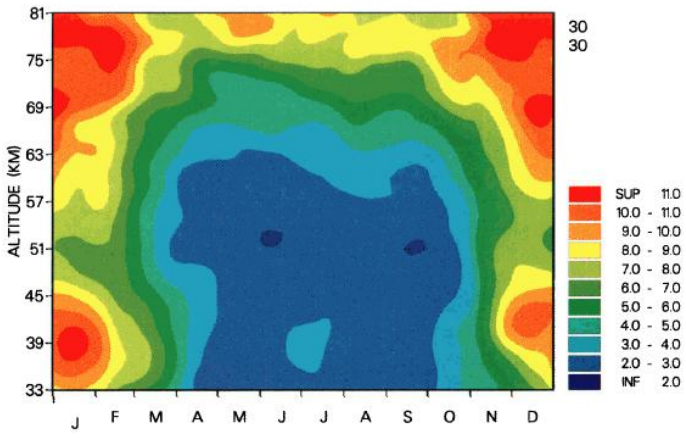


Figure 5.3. Seasonal variation in temperature standard deviation. A more detailed plot of Figure 2b, provided for comparison with Figure 2a and 2b. (This figure is Plate 3 from *Hauchecorne et al.* [1991]. Used with permission of the American Geophysical Union.)

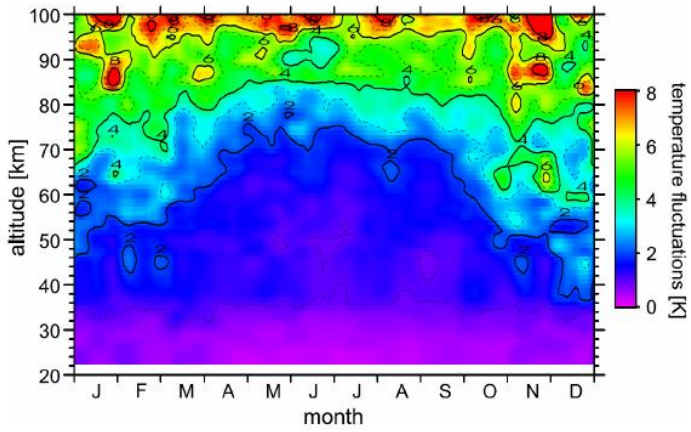


Figure 5.4. Seasonal variation of nightly mean temperature fluctuations derived from the two Leibniz-Institute of Atmospheric Physics lidar; 236 nights of measurements are smoothed with a ±15 day, ±2 km Hanning filter. (Image is Figure 7 from *Rauthe et al.* [2008]. Used with permission.)

during summer and about 4 K during winter. The mid-mesosphere standard deviations are 2 K during summer and 6 K during winter. In the upper mesosphere they range from 5 K during summer to about 7 K during winter.

Two interesting seasonal features in the USU standard deviations are the increases around mid May and early October at 88 km that reach approximately 20 K, and perhaps a third in December. They are strong, but it is difficult to determine if they are anomalous, seasonal, or due to unique local conditions. A similar peak was found in the *Hauchecorne et al.* climatology at 81 km during May. There were no similar peaks found in the *Rauthe et al* climatology. Near 50 km there is a decrease in the standard deviation in May-June and also August-September. We see similar dips in the French climatology at 52 km and in the *Rauthe et al.* climatology at 45 km. Another common feature occurs at 57 km near August where the climatology shows a standard deviation increase that declines with altitude. A similar feature can be seen in the USU data where it occurs

about a month earlier and starts about 5 km higher. This feature is not found in the *Rauthe et al.* climatology.

6. Seasonal Changes in Solar Response

The seasonal solar-noise coefficient may be used to infer a seasonal atmospheric response to solar-noise. To calculate this variation, temperature data from a 121-day (~4-month) window was selected; the same 121 days were selected from each year for which we have data. For example, for each year, all the data between the day-of-year days 5 and 126 were selected and Model (5.1) was fit to that data. From that fit, the solar-noise coefficient was obtained. This selection window was advanced day by day and Model (5.1) was fit to the selected data. Consequently, the data was not compiled into a composite year calculation for this calculation. From this we got an idea of the seasonal variation of the strength of the solar-noise coefficient. The results are shown in Figure 5.5. The atmospheric solar-noise response is greatest at 85 km October-December where

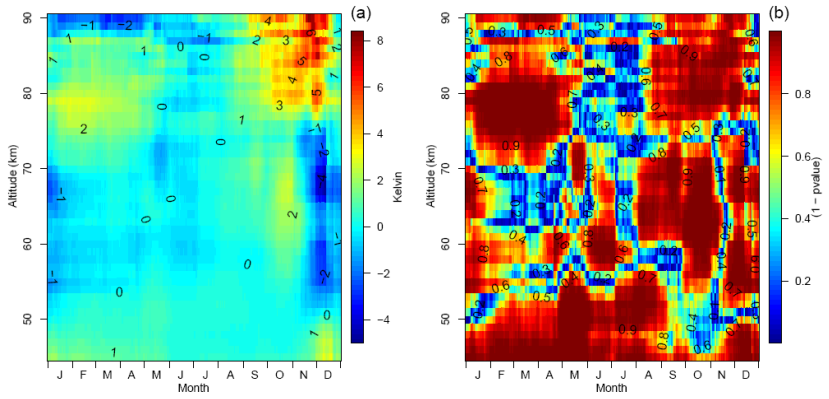


Figure 5.5. Seasonal variation in solar noise coefficient. Plot (a) shows the seasonal variation of the coefficient of the Model (5.1) solarnoise term. Plot (b) shows confidence levels of the coefficient. The confidence levels are between 0 (the lowest confidence level) and 1 (the highest confidence level).

the SN coefficient reaches 8 K. There is another strong response that is about 5 km lower and reaches a maximum of 3 K during January-March. In the middle mesosphere there is a statistically significant SN response from August to January which ranges from 2 K near October to −1 to −2 K from November to January. In the lower mesosphere there is a weakening of the solar-noise term that occurs from June to November and is statistically significant throughout most of the year.

Between 80 and 90 km, the solar-noise response is stronger during winter than it is during summer, which is similar to findings from other researchers who found the winter solar response in the upper mesosphere to be stronger than the summer solar response [*Schwentek*, 1971; *Hood*, 1987; *Keckhut et al.*, 1995; *Pertsev and Perminov*, 2008].

From Figure 5.5b, it is clear that the solar-noise term is either highly significant or insignificant. The rapid transition from high to low significance is due to the rapid transition in coefficient value from positive to negative.

7. Calculating the Residual Autocorrelation Coefficient

The autocorrelation coefficient indicates the degree to which yesterday's temperatures are correlated to today's temperatures. The autocorrelation coefficient varies between 0 (no correlation) and 1 (perfect correlation). This could be interpreted as a stronger memory in the time series or indicate the duration of a process. For example, yesterday's temperature could be strongly influencing today's temperatures or a temperature perturbation could be occurring over a period of a day or more. Given that we are measuring temperature variations and that planetary waves are likely responsible for a significant portion of the variation, we prefer the second interpretation. Because of

the prevalence of planetary waves in winter, the autocorrelation should be bigger in winter. (Finding a method for calculating the autocorrelation coefficient from discontinuous data was challenging. The method for this is described in Appendix E.)

The autocorrelation profile using the entire data set is shown in Figure 5.6. At 45 km the autocorrelation starts at 0.7 and decreases to around 0.6 at 50 km. From 50 to 65 km it oscillates around 0.55; from 65 to 80 km it decreases to 0.15. Above this there is a slight increase before it decreases to zero. On average the autocorrelation decreases at a rate of $0.015\ \mathrm{km}^{-1}$.

It is widely accepted that the seasonal variation in the temperature standard deviation is caused by gravity and/or planetary wave activity. We also know that temperature inversions below the mesopause can last for several days. These processes can inflate temperature standard deviations. Because autocorrelation also inflates residual standard deviations (see Appendix H) a possible connection exists between the autocorrelation and the residual standard deviation. Because the standard deviation follows a seasonal pattern, this opens the question of a seasonal variation in the autocorrelation. To check for this, the residuals were selected as they were for the solar-noise coefficient calculations, according to the method described in the previous section. Residuals from a 121-day (~4-month) window were selected. For example, for each year, all the residuals between the day-of-year days 5 and 126 were selected and the autocorrelation coefficient was calculated for that data. From that fit the solar-noise coefficient was obtained. By advancing the selection window day by day, we created a 365-day climatology of the autocorrelation coefficients, which is shown in Figure 5.7.

It is evident that below 75 km the autocorrelation follows a pattern very similar to

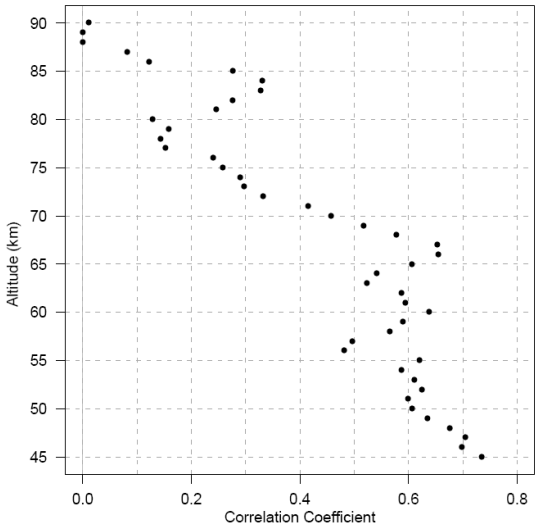


Figure 5.6. The autocorrelation in the residuals of the entire USU data set.

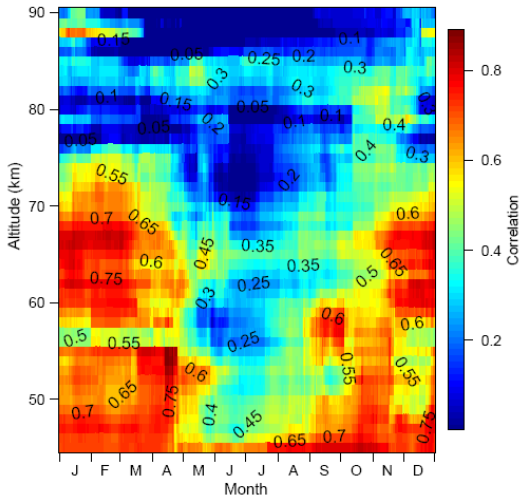


Figure 5.7. Seasonal variation in autocorrelation.

that of the standard deviation in Figure 5.2a: high winter and low summer activity. Like the standard deviations, above 75 km the autocorrelation does not show a seasonal variation. In contrast to the standard deviations, the autocorrelation is small above 75 km and the standard deviation is large. One may infer from the autocorrelation climatology that different physical processes are involved. The elevated wintertime correlations below 75 km originate from processes that occur on time scales of at least a day, and above 75 km, the lower autocorrelations indicate the causative processes are occurring over a shorter time scale.

8. Solar Cycle Variation

In addition to the usual seasonal variation in standard deviation there is also a standard deviation variation that occurs on solar cycle time scales. To check for this the data were divided into four period sections, each 2.5 years long. For each section the standard deviation of the Model (5.1) residuals was calculated. These are shown in Figure 5.8.

While there is a small variation below 70 km, above 70 km there is a marked higher variation from 2.5 to 7.5 years, with the maximum difference occurring near 80 km. The time divisions for the standard deviations are shown in Figure 5.9 along with the Mg II solar flux proxy. In section I (0 to 2.5 y) the standard deviation is around 10 K at 80 km. In section II (2.5 to 5 y) it increases to 14 K. In section III (5 to 7.5 y) it is about 16 K. Finally, in section IV is decreases back to 10 K. *Höppner and Bittner* [2007] found a similar solar-cycle variation in the temperature standard deviation in the hydroxyl (OH*) layer above Wuppertal, Germany (51ºN, 7ºE) at about 87 km. (See also *Bittner et al.*

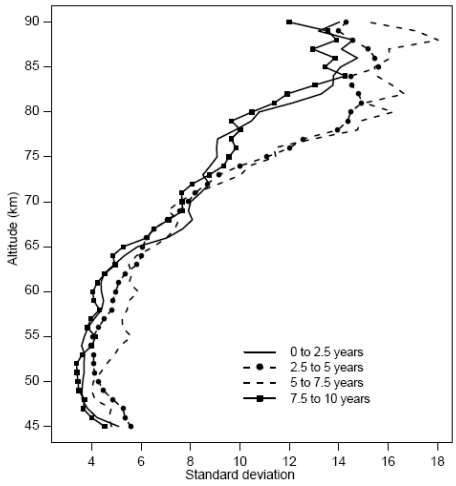


Figure 5.8. The temporal variation in the residual standard deviations over an 11-year period. Each profile is calculated from a 2.5 year block of data.

[2000, 2002] for earlier analyses with similar results.) They found that the temporal evolution of the yearly mean standard deviation of the temperatures reached a maximum about four years before the 2000 solar maximum. (Figure 5.10.) Their standard deviations are considerably smaller than ours. Our max-min difference is about 6 K while theirs is about 2 K. The phases relative to the solar cycle are similar.

This temporal variation in temperature standard deviation was discussed in a paper by *Beig et al.* [2003]. In it they discuss results from *Bittner et al.* [2002], which is an earlier study related to *Höppner and Bittner* [2007]. Beig concludes that it is unlikely the temporal variation in standard deviation is due to measurement error. The standard deviation is a fairly robust statistical parameter and insensitive to instrument bias. Consequently the standard deviation as an atmospheric parameter is sui generis, unique or particular in itself. The later results of *Höppner and Bittner* [2007], as well as our results, support their original conclusions regarding the existence of an 11-year evolution

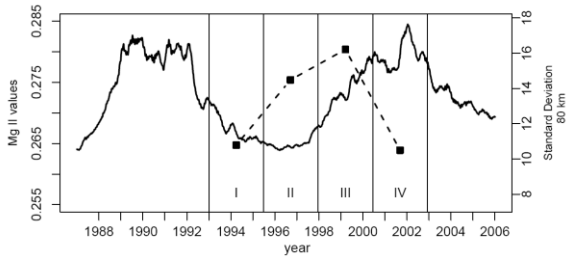


Figure 5.9. The variation of the residual standard deviation from 1994 to 2002. The vertical lines indicate the four data sections. The solid line represents the 81-day smoothed Mg II values. The dashed line indicates the standard deviations at 80 km.

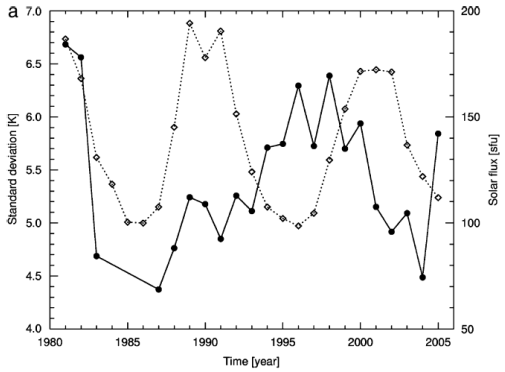


Figure 5.10. The variation of the temperature standard deviation from 1980 to 2005. The solid line shows the temporal evolution of the temperature standard deviation from 1981 to 2005. The dotted line indicates the solar F10.7 solar flux. (Taken from Figure 4a from *Höppner and Bittner* [2007]. Used with permission.)

of the standard deviation.

As far as the differences in the standard deviation between our data and the Wuppertal temperatures is concerned, *Bittner et al.* [2002] found similar differences in the standard deviations between their data and the French data reported in *Leblanc et al.* [1998]. The maximum standard deviation was 12 K at 80 km in the French data, which is closer to our results from the same altitude and latitude. *Bittner et al.* suggested the smaller standard deviation in their data was due to an inherent smoothing effect owing to

the broadness of the OH layer.

One of the theories tested by *Bittner et al.* [2002] is the so called downward control principle, or wave-driven circulation. This theory says that during winter, increased wave breaking at high altitudes increases temperatures at lower altitudes. So there should be a correlation between changes in wave activity and lower altitude temperatures [*Holton et al.*, 1995; *Haynes et al.*, 1991]. To test for this, they calculated the temperature standard deviations at approximately one-year intervals. They did the same for the monthly mean temperatures from November to March and selected the lowest values for a given year in order to reduce the influence of sudden stratospheric warmings. From this a correlation between the lower altitude temperatures and the higher altitude standard deviations was calculated. The lower altitude temperatures are the SSU temperatures from 1 hPa (~47 km).

They found a statistically significant (~95%) correlation of 0.51 for a time lag of 1.5 years, which they tentatively interpret as evidence for the downward control principle. A similar calculation was performed on the USU temperatures by dividing the temperatures from 80 km into sections of 1.5 years and taking the temperature standard deviation for each section. Because of data gaps, we ended up with only seven data points; smaller divisions seemed impractical. The temperatures from 45 km for the months of November through March were selected and from those months temperatures greater than the monthly mean were omitted in order to reduce the influence of sudden stratospheric warmings. The monthly temperatures were averaged and compared to the standard deviations from 80 km. A negative correlation of −0.56 was found. However, as we had only seven data points, this was not statistically significant and because we had so few

points, we were unable to lag our standard deviations. Taking a closer look at the *Bittner et al.* results, it is possible their correlation was strongly influenced by outliers. When the two outermost points were omitted, the correlation dropped to 0.25 and had a significance of 78%. Their correlation calculation was based on 14 data points, which is barely enough for statistical analysis.

9. Conclusions

There is a strong seasonal variability in the residual standard deviations below 75 km. They show the characteristic low summer and high winter variance. Our seasonal variances compare favorably with those from *Hauchecorne et al.* [1991] and *Rauthe et al.* [2008]. We found a low standard deviation of 2-4 K in the lower and mid mesosphere during summer and a high standard deviation during winter, 4-12 K from the same regions. There were two low points at 50 km during May-June and August-September at about 2 K. Similar low points were found in the French data at the same altitude and time of year. Above 75 km the residual standard deviations are fairly constant and relatively high.

The solar-noise coefficient was stronger during winter in the upper mesosphere. In the middle mesosphere it was positive and statistically significant during October-November and negative and statistically significant during December-January. Throughout the remainder of the year is was not statistically significant. In the lower mesosphere it was statistically significant nearly throughout the year, though it was slightly stronger during winter. A strong winter solar response has been found by several researchers and our results from the upper and lower mesosphere show a similar

behavior.

We found a seasonal variation in the autocorrelation coefficient for the residuals below 75 km that correlates strongly with the seasonal variation in the residual standard deviations. A high autocorrelation coefficient indicates a strong correlation between earlier temperatures and later ones. The high autocorrelations are possibly indicative of long-period gravity waves or planetary waves. Above 75 km the autocorrelation coefficient is small and fairly constant.

We investigated the 11-year variation in the residual standard deviation and found an 11-year variation in the middle mesosphere (73-85 km) that leads the 11-year solar cycle by about three years. A similar lag and variation in standard deviation was found by *Höppner and Bittner* [2007]. *Bittner et al.* [2002] also found evidence for wave-driven circulation (downward control) by correlating standard deviations from 87 km with temperatures from 1 hPa (~47km). We did not find any significant correlation, which was expected because we had only seven data points.

Time series plots for the model residuals from every altitude bin are given in Appendix F.

References

Beig, G., et al. (2003), Review of mesospheric temperature trends, *Rev. Geophys.*, *41*(4), doi:10.1029/2002RG000121. [online] Available from: http://www.agu.org/pubs/crossref/2003/2002RG000121.shtml.

Bittner, M., D. Offermann, and H. H. Graef (2000), Mesopause temperature variability above a midlatitude station in Europe, *J. Geophys. Res.*, *105*(D2), 2045-2058, doi:10.1029/1999JD900307.

Bittner, M., D. Offermann, H. Graef, M. Donner, and K. Hamilton (2002), An 18-year time series of OH rotational temperatures and middle atmosphere decadal

variations, *J. Atmos. Sol. Terr. Phys.*, *64*(8-11), 1147-1166, doi:10.1016/S1364-6826(02)00065-2.

Hauchecorne, A., M. Chanin, and P. Keckhut (1991), Climatology and trends of the middle atmospheric temperature (33–87 km) as seen by Rayleigh lidar over the south of France, *J. Geophys. Res.*, *96*(D8), 15297-15309, doi:10.1029/91JD01213.

Haynes, P., M. McIntyre, T. Shepherd, C. Marks, and K. Shine (1991), On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces, *J. Atmos. Sci.*, *48*(4). [online] Available from: http://journals.ametsoc.org/doi/abs/10.1175/1520-0469%281991%29048%3C0651%3AOTCOED%3E2.0.CO%3B2 (Accessed 19 May 2010).

Hines, C. (1971), Generalizations of the Richardson criterion for the onset of atmospheric turbulence, *Q. J. Roy. Meteor. Soc.*, *97*(414), 429-439, doi:10.1002/qj.49709741405.

Holton, J. R., P. H. Haynes, M. E. McIntyre, A. R. Douglass, R. B. Rood, and L. Pfister (1995), Stratosphere-troposphere exchange, *Rev. Geophys.*, *33*(4), 403-439, doi:10.1029/95RG02097.

Hood, L. L. (1987), Solar ultraviolet radiation induced variations in the stratosphere and mesosphere, *J. Geophys. Res.*, *92*(D1), 876-888, doi:10.1029/JD092iD01p00876.

Höppner, K., and M. Bittner (2007), Evidence for solar signals in the mesopause temperature variability?, *J. Atmos. Sol. Terr. Phys.*, *69*(4-5), 431-448, doi:10.1016/j.jastp.2006.10.007.

Kafle, D. M. (2009), *Rayleigh-lidar Observations of Mesospheric Gravity Wave Activity above Logan*, Ph.D., Utah State University, Logan, Utah. Available from: http://digitalcommons.usu.edu/etd/466/ (Accessed 1 June 2010).

Keckhut, P., A. Hauchecorne, and M. L. Chanin (1995), Midlatitude long-term variability of the middle atmosphere: Trends and cyclic and episodic changes, *J. Geophys. Res.*, *100*(D9), 18887–18897.

Leblanc, T., I. S. McDermid, P. Keckhut, A. Hauchecorne, C. Y. She, and D. A. Krueger (1998), Temperature climatology of the middle atmosphere from long-term lidar measurements at middle and low latitudes, *J. Geophys. Res.*, *103*(D14), 17191–17204.

Pertsev, N., and V. Perminov (2008), Response of the mesopause airglow to solar activity inferred from measurements at Zvenigorod, Russia, *Ann. Geophys.*, *26*(5), 1049-1056, doi:10.5194/angeo-26-1049-2008.

Rauthe, M., M. Gerding, and F. Lübken (2008), Seasonal changes in gravity wave activity measured by lidars at mid-latitudes, *Atmos. Chem. Phys.*, *8*(22), 6775-6787.

Schwentek, H. (1971), The sunspot cycle 1958/70 in ionospheric absorption and stratospheric temperature, *J. Atmos. Terr. Phys.*, *33*(12), 1839-1852, doi:10.1016/0021-9169(71)90163-2.

Wilson, R., M. L. Chanin, and A. Hauchecorne (1991), Gravity wave in the middle atmosphere observed by Rayleigh lidar 2: Climatology, *J. Geophys. Res.*, *96*(D3), 5169-5183, doi:10.1029/90JD02610.

CHAPTER 6

SUMMARY AND FUTURE WORK

1. Summary of Research

We have examined collinearity, model specification, solar-noise, atmospheric solar response, linear cooling trends, the possibility that the Pinatubo volcanic eruption affected our initial temperature measurements, and the standard deviation and autocorrelation of the residual noise.

Collinearity can affect an ordinary least squares (OLS) model in several ways. We closely examined two: coefficient correlation and inflated standard errors (SEs). Coefficient correlation arises when two or more regressors have some degree of linear dependence. For a positive correlation, when one coefficient is higher than its true value the other will also be higher than its true value; when they are negatively correlated they have an inverse relationship. There is no a priori way to determine if a coefficient is larger or smaller than its true value. But determining the correlation is a simple calculation and the correlation numbers should be reported. Bootstrapping can be used to better visualize the correlation. In our OLS models there is a strong negative correlation (−0.83) between the linear trend and the *sin ωt* proxy. This means that if the linear trend is higher than its true value, the *sin ωt* coefficient is lower than its true value; or if the linear trend is lower than its true value the *sin ωt* coefficient is higher than its true value. At 90 km we have detected a large linear trend that is greater than −1.5 K/year and also a negative *sin ωt* coefficient. This means if our linear trend is lower (more negative) than its true value then the *sin ωt* coefficient will be high, that is, shifted in the positive

direction, which means the magnitude of the *sin ωt* coefficient is likely to be smaller than its true value. But as already mentioned, there is no a priori way to determine if a given coefficient is lower or higher than its true value.

Another issue related to collinearity is inflated standard errors. The inclusion of the solar-like terms increases the linear trend SEs and limits the confidence that can be placed on the values of the linear coefficient. We found the error limits on the linear coefficient were 45% smaller when the solar-like *sin ωt* and *cos ωt* terms were omitted from the model. Under most circumstances if the solar-like terms are not included the model is under specified and consequently there exists a risk of biasing the linear trend coefficient. However, this is not the case if the phase of the solar proxy (or the *sin ωt* term) is near the time center of the data set, which is the case for the USU temperatures. Another symptom of collinearity is sensitivity to model specification. We found that our linear trend coefficient is fairly stable under model perturbations. When a fixed solar proxy was included in the model, the amplitudes were not abnormally high or low. When the solar proxy was omitted, the linear trend coefficient was still high. Taken together it does not appear that collinearity is behind the large cooling trends in our temperatures.

There is sound evidence that the Pinatubo eruption increased the temperatures of the mesopause region [*Keckhut et al.*, 1995; *She et al.*, 1998; *Bittner et al.*, 2002]. When we divided our data set in half the mesopause temperatures from the first half show a linear trend of approximately −4 K/year and the linear trend from the second half is approximately –1 K/year. Between 76 and 50 km both the first and second halves show a near zero trend. Below 50 km the second half of the data has a cooling trend of −0.5 K/year and the first half data has a warming trend of +0.5 K/year. If the first year is

omitted from the data set then from 80 to 87 km the linear trends are near −1.3 K/year; for the full data set they are about −1.5 K/year. Between 50 and 80 km the linear trends are nearly identical to those from the full data set. Below 50 km the shortened data set has a −0.1 K/year cooling trend and the full data set has a 0.2 K/year warming trend. The differences of the linear trends between 80 and 90 km supports the thesis that a Pinatubo related heating increased our early temperatures in that altitude region. An examination of the residuals at 85 km also shows what appears to be a temperature perturbation of about 10 to 15 K at the beginning of the data set. But this is only true for a fixed-proxy model and may be an unaccounted for solar-like variation. Stronger cooling rates above 70 km at the beginning of the data set are consistent with the existence of a Pinatubo warming effect in our early temperatures. But because we do not have data prior to 1993, we cannot assess for certain if this is the case.

A second OLS issue is model specification. If a model is under specified, then the correctly specified terms in the model are likely to be biased. If the model is cross specified, then the correctly specified terms will also be biased. If a fixed-phase solar proxy is included in a model and a variable phase solar proxy is more appropriate, then for a data set of 12 years, the maximum linear trend bias would be 0.1 K/year for a 1 K solar response. For a 2 K solar response the bias would be 0.2 K/year. This result is not small considering that scientists modeling mesosphere cooling are anticipating cooling trends on the order of 0.1 K/year.

Most researchers looking for secular trends in middle atmosphere temperatures include a fixed-phase solar proxy in their models. Our analysis of the USU temperatures indicates a significant out-of-phase temperature response to the solar input at several

different altitudes. Again, the use of a fixed-phase solar proxy when there is a significant phase difference in the temperatures can lead to bias in the linear trend coefficient. But even in cases where a model has a fixed-phase solar proxy, a selection criteria may be applied to reduce the risk of bias by selecting trends from altitudes where the atmospheric solar response is at a maximum or minimum value. Because the attenuation of the atmospheric solar response cannot exceed one, altitudes where the fixed-phase proxy coefficient has a maximum or minimum value are more likely to be closer to the true value of the atmospheric solar response. These altitudes may be emphasized as locations where the atmospheric solar response is likely to be in phase or out of phase with the solar input, with the understanding that where the solar response is in phase or out of phase with the solar input the linear term is likely to be less biased.

We have concluded that the solar-noise should be included as a separate term in the temperature model. We found this term to be statistically significant at 90% or better between 45 and 54 km and between 75 and 86 km. At 45 km the statistical significance was greater than 99.9%. A Mallow's C_p test was done on the CPC and ERA temperature time series, as well as for every altitude in the USU temperatures. The results confirmed our earlier conclusion that the solar-noise should be included in the temperature models. We also found that if the solar-noise term is not included, the omission could bias the solar proxy coefficient.

We found a seasonal variation in the solar-noise coefficient in the upper mesosphere. The effect of the solar-noise was stronger during winter than during summer, but it was also divided into two distinct periods from January to March and from September to November, the second period being much stronger than the first by about 3 K and higher

in altitude by about 5 km. In the middle mesosphere the solar-noise response is near zero except for October-November when it reaches about 2 K, and November-December at −2 K. During winter the lower mesosphere response is weaker than it is during summer.

As mentioned in connection with model specification, we examined the atmosphere response to solar cycle variations, allowing for a variable phase. The magnitude of the maximum – minimum value of the atmospheric solar response above Logan, Utah is about 3.5 K at 45 km and about 1 K at 50 km. It then steadily increases to about 4 K at the mesopause. Only *Remsberg* has allowed for a variable phase proxy. The amplitudes of the atmospheric solar response from *Remsberg* [2007, 2009] are comparable to ours. At 49 km it is near 1 K. At 80 km it is near 3 K and within or nearly within our error limits. Their lower mesosphere amplitudes are slightly greater than ours and in the upper mesosphere, our amplitudes are slightly greater than theirs. We found significant phase differences in the atmospheric response to the solar input. However, they differed from those found by *Remsberg* [2007, 2009] in the HALOE temperatures. In the lower mesosphere the HALOE data shows a phase of about 10 years (equivalent to −1 year) whereas ours range from 0 to ±5 years. In the upper mesosphere the HALOE phases are close to zero. Our upper mesosphere phases occur later by about one to two years.

We also found a seasonal response to the solar input in the upper mesosphere. The solar response during winter was stronger than it was during summer by a few Kelvin. In the middle mesosphere the solar response was stronger during summer by about the same amount.

We looked for annual and semiannual oscillations in the temperatures. Our annual oscillation (AO) compares favorably in amplitude and phase with the HALOE AO. It also

compares well with the French OHP and CEL AO, with the exception that at 45 km the French AO is 5 K and increases to 6 K at 55 km. The USU AO amplitude is 10 K at 45 km and steadily decreases to 6 K at 55 km. All the AO phases are similar and exhibit warm summers and cold winters in the lower mesosphere and reversed seasons in the upper mesosphere.

Our semiannual amplitudes exhibit a double peak profile, that is, two maxima near 62 and 80 km. The OHP, CEL, HALOE, and SABER data also show the same double-peak pattern. The HALOE profile was a few kilometers higher than the others. The peaks in the OHP, CEL, and SABER SAO coincide with those in our data. The lower altitude peak in the OHP and CEL amplitudes are greater than ours by about two Kelvin. The higher altitude peak of the CEL has a magnitude close to ours; the OHP peak was smaller than ours by about 1 K.

The phases of the SAOs are also similar. At 45 km the first SAO phase is about a month before summer solstice, in mid May. The phase steadily transitions to earlier in the year with increasing altitude until at 83 km it occurs around February. Above 83 km the USU, HALOE, and SABER SAO phase profiles exhibit a shift toward later in the year. This shift is greatest in the USU data, occurring in May at 90 km. At 45 km the USU phase is earlier than the OHP and HALOE phases by about a month. The CEL phase is a month latter. The second SAO phase occurs six months later. Because of the way we did the analysis the second phase is otherwise identical to the first.

A temperature climatology emphasizing the SAO was found by removing the linear trend, AO, solar response, and solar-noise from the temperatures. These results are more complicated than those found for the SAO using least squares. Below 75 km the

climatology shows that the second phase of the SAO is stronger than the first by about 4 K; this is also true for the comparable HALOE climatology. Above 75 km the first phase of the USU climatology is stronger than the second by about 2 K, while in the HALOE climatology the second phase is stronger than the first by about 4 K. There is also a weakening of the SAO amplitude during the first phase at 65 km during March-April and again in the second phase at 70 km during August-September. This feature is also found in the HALOE data. Finally, there is a 1 K temperature anomaly that starts at 45 km during September and penetrates up to 60 km in October and is found in both the USU and HALOE climatologies. The apparent differences between the first phase and second phase SAO amplitudes and the September-October anomaly likely arise from the influence of higher order terms.

In the residual standard deviations, we found a strong seasonal pattern below 75 km where standard deviations (SDs) are lowest during summer and highest during winter. In the lower mesosphere, from May to September, the SD of the residuals are 1 to 2 K. From November-March they are 6 to 8 K. In the middle mesosphere the SDs, are 8 to 12 K from November to February, and 4 to 8 K from March to October. Above 75 km the residuals show a fairly constant SD around 12 K, though there were two peaks at April-May and September, and perhaps a third in December, that reached as high as 20 K.

Our SD climatology is similar to others in the literature. *Hauchecorne et al.* [1991] published a SD climatology that matches ours in both seasonal pattern and in the magnitudes of the SDs. Their climatology also shows SD minima around 52 km that occur May-June and September-October. A similar feature is found in our SDs as well. At 48 km there are minima around May-June and August-September; there are also two

additional minima at 55 km, which occur a month earlier. Our data also shows a slight SD increase of 2-10 K that occurs near 75 km during April and penetrates down to 55 km. In the *Hauchecorne et al.* climatology this feature is not as strong.

At 80 km we found signs of elevated SDs that correlated with the 11-year solar variation that occurred nearly three years before solar maximum. A similar variation at 87 km was also found by *Höppner and Bittner* [2007] whose SDs reached a maximum four years before the solar maximum at 87 km. *Bittner et al.* [2002] found a positive, statistically significant correlation between the temperature SDs at 87 and SSU stratopause temperatures. We did not find a similar correlation in our data, but this could be because of the shortness of our data set.

We created a climatology of the autocorrelation coefficients for the Model (5.1) residuals and found that below 75 km the autocorrelation is considerably stronger during winter than it is during summer. From September to April the correlation averages around 0.6. From May to August it is about 0.35. What this tells us is that during winter, below 75 km, the temperature SDs are inflated because of a seasonal autocorrelation in the residuals. This suggests that the large autocorrelation coefficient in winter arises from planetary waves and mesospheric inversion layers that last more than a day. Above 75 km the autocorrelation remains small throughout the year, near 0.3 with no seasonal variation. In contrast, this region is where the SDs reach their maximum value, but similarly they have no seasonal variations. The smaller autocorrelation suggests that above 75 km the cause of the elevated SDs is a physical process that is likely occurring on timescales of less than 24 hours.

2. Final Comments

Some of the statistical techniques employed in this dissertation have been used before, such as the SD climatology and least squares analysis, to detect cooling trends, the atmospheric solar response, and the influence of the Pinatubo eruption. However, several new tools have been added: The solar-noise term and its climatology along with the autocorrelation coefficient and its climatology. We have also included a mid-latitude SAO climatology (of which there are few). Equations have been derived that can be used to estimate the bias and coefficient attenuation in models that contain a fixed-phase solar proxy, and we used these to developed arguments that address the dangers of linear trend bias in the presence of an atmospheric response that is not in phase to the solar input. We show that altitudes where the fixed-proxy coefficient is at a maximum or minimum value are ideal for identifying middle atmosphere cooling, as well as the magnitude of the atmospheric solar response. We also more fully considered the effects of collinearity between the linear trend and fixed-proxy coefficients.

3. Future Work

Work needs to be done to calculate the coefficient correlations from the OLS models used in the literature. Researchers do not report their coefficient correlations and do not take these correlations into account in their analysis. However, researchers do report their models and that is all that is needed to calculate the coefficient correlations. Further, if error limits are given it might be possible to estimate the height and width of the elliptical correlation pattern.

Also needed is an understanding of how collinearity affects the amplitudes and phases

of the variable-phase solar proxy. For a fixed-phase proxy such as MgII this is simple. But the amplitude of a variable-phase regressor is $[(C_1)^2 + (C_2)^2]^{1/2}$, which is non-linear. The phase angle is given by $\tan^{-1}(C_1/C_2)$ and is monotonic from $-\pi/2$ to $\pi/2$, but further investigation is also needed. We also did not touch on how collinearity affects the solar proxy coefficients.

The error analysis of the autocorrelation coefficient needs to be more fully developed, and a formal proof showing that a correct guess of the autocorrelation coefficient minimizes the residual SD is needed. A better physical interpretation and explanation of the residual temperature autocorrelations needs to be worked out.

Because autocorrelation inflates the residual SDs (see Appendix H) it also inflates coefficient SEs. If the autocorrelation coefficient can be accurately estimated, it might be possible to reduce the coefficient SEs. Another thing that needs inquiry is whether including the solar-noise term in the model reduces autocorrelation in the residuals.

Also needed is a theory connecting gravity-wave activity, the SD variation, and the seasonal variation in the autocorrelation. The effects of the solar-noise also need a physical interpretation.

One interesting result from our solar-noise analysis that we were unable to incorporate into this dissertation because of time constraints is a lag between the solar-noise term and the temperature residuals. When the solar-noise is lagged by 30 days, the correlation between the solar-noise and the temperature residuals increases from 0.17 to 0.20. This was also found in the CPC data, but the correlations were much smaller. We need to check for a lag with other solar proxies and determine if the lagged solar-noise improves the model results. This lagged correlation might also be of value in predictive

models.

More data will improve our analysis. Most of the parameters we worked with are affected by the length of the data set. The collinearity problem can be nearly eliminated if we could acquire about 2.5 more years of data. We could also better estimate the linear trend coefficient and the amplitude and phase of the atmospheric solar response and check for continued variation in the temperature standard deviation with the solar cycle.

The addition of the HALOE data to this dissertation came late and we were unable to fully incorporate in it into the analysis. All the analysis applied to the USU temperatures can also be applied to the HALOE temperatures. It is also possible to repeat our analysis on temperature data available from the National Weather Service Climate Prediction Center (CPC).

References

Bittner, M., D. Offermann, H. Graef, M. Donner, and K. Hamilton (2002), An 18-year time series of OH rotational temperatures and middle atmosphere decadal variations, *J. Atmos. Sol. Terr. Phys.*, *64*(8-11), 1147-1166, doi:10.1016/S1364-6826(02)00065-2.

Hauchecorne, A., M. Chanin, and P. Keckhut (1991), Climatology and trends of the middle atmospheric temperature (33–87 km) as seen by Rayleigh lidar over the south of France, *J. Geophys. Res.*, *96*(D8), 15297-15309, doi:10.1029/91JD01213.

Höppner, K., and M. Bittner (2007), Evidence for solar signals in the mesopause temperature variability?, *J. Atmos. Sol. Terr. Phys.*, *69*(4-5), 431-448, doi:10.1016/j.jastp.2006.10.007.

Keckhut, P., A. Hauchecorne, and M. L. Chanin (1995), Midlatitude long-term variability of the middle atmosphere: Trends and cyclic and episodic changes, *J. Geophys. Res.*, *100*(D9), 18887–18897.

Remsberg, E. E. (2007), A reanalysis for the seasonal and longer-period cycles and the trends in middle-atmosphere temperature from the Halogen Occultation

Experiment, *J. Geophys. Res.*, *112*(D9), doi:10.1029/2006JD007489. [online] Available from: http://www.agu.org/pubs/crossref/2007/2006JD007489.shtml.

Remsberg, E. E. (2009), Trends and solar cycle effects in temperature versus altitude from the Halogen Occultation Experiment for the mesosphere and upper stratosphere, *J. Geophys. Res.*, *114*(D12), doi:10.1029/2009JD011897. [online] Available from: http://www.agu.org/pubs/crossref/2009/2009JD011897.shtml.

She, C., S. W. Thiel, and D. A. Krueger (1998), Observed episodic warming at 86 and 100 km between 1990 and 1997: Effects of Mount Pinatubo eruption, *Geophys. Res. Lett.*, *25*(4), 497, doi:10.1029/98GL00178.

APPENDICES

APPENDIX A

PERMISSION LETTERS

Dear Troy

RE: Citation:Ramsden 72 (477-486): 156
Journal:Proceedings of the Royal Society of London

The Royal Society is pleased to grant permission for use of the above material, subject to the following conditions:

1. Full acknowledgement is given to the original source, with full details of the author(s), title, figure/page numbers, year of publication and the publisher.

2. The permission of the author(s) or the author's estate is obtained where practical.

3. The material is to be used only as described in your email and this permission is granted for one-time use only.

Yours sincerely

Jennifer Kren
Editorial Administrator

Dear Dr. Wynn:

Thank you for requesting permission to reproduce material from American Association of Physics Teachers publications.

Permission is granted – subject to the conditions outlined below – for the following:

Figure 8 from Gilbert N. Plass, "Infrared Radiation in the Atmosphere," American Journal of Physics -- May 1956 -- Volume 24, Issue 5, pp. 303-321

To be used in the following manner:

Reproduced in your Ph.D dissertation titled, STATISTICAL ANALYSIS OF THE USU LIDAR DATA SET WITH REFERENCE TO MESOSPHERIC SOLAR RESPONSE AND COOLING RATE CALCULATIONS, WITH ANALYSIS OF STATISTICAL ISSUES AFFECTING THE REGRESSION COEFFICIENTS,

for submission to Utah State University.

1. The American Association of Physics Teachers grants you the right to reproduce the material indicated above on a one-time, non-exclusive basis, solely for the purpose described. Permission must be requested separately for any future or additional use.

2. This permission pertains only to print use and its electronic equivalent, including CD-ROM or DVD.

3. The following copyright notice must appear with the material (please fill in the information indicated by capital letters): "Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], American Association of Physics Teachers."

Full citation format is as follows: Author names, journal title, Vol. #, Page #, (Year of publication).

For an article, the copyright notice must be printed on the first page of the article or book chapter. For figures, photographs, covers, or tables, the notice may appear with the material, in a footnote, or in the reference list.

4. This permission does not apply to any materials credited to sources other than the copyright holder.

5. If you have not already done so, please attempt to obtain permission from at least one of the authors. The author's address can be obtained from the article.

Please let us know if you have any questions.

Sincerely,
Susann Brailey

Dear Troy,

The HALOE data are in a public archive, so no acknowledgment is required. However, I suggest that you say the form of the data (Level2 files?) and from where (Website or archive) you accessed the dataset that you used.

Best wishes in finishing your degree work,

Ellis Remsberg

Dear Troy Wynn,

It is a pleasure for me to give the permission to reproduce the Figure 6 from the publication "Seasonal Changes in gravity wave activity measured by lidars at mid-latitudes, Atmos. Chem. Phys., 8, 6775-6787, 2008". I would like to ask you to cite the Figure correctly and it would be nice to get the finished Ph.D. dissertation. Please send the dissertation to monika.rauthe@dwd.de because since January 2010 I'm working at the Deutscher Wetterdienst.

Sincerely,

Dr. Monika Rauthe

ELSEVIER LICENSE TERMS AND CONDITIONS

May 18, 2010

This is a License Agreement between Troy Wynn ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Troy Wynn
Customer address	620 E 1000 N Apt. 4 Logan, UT 84321
License Number	2432041388331
License date	May 18, 2010
Licensed content publisher	Elsevier
Licensed content publication	Journal of Atmospheric and Solar-Terrestrial Physics
Licensed content title	An 18-year time series of OH rotational temperatures and middle atmosphere decadal variations
Licensed content author	M. Bittner, D. Offermann, H. -H. Graef, M. Donner, K. Hamilton
Licensed content date	7 May 2002
Volume number	64
Issue number	8-11
Pages	20
Type of Use	Thesis / Dissertation
Portion	Figures/tables/illustrations
Number of Figures/tables/illustrations	10

Format	Both print and electronic
You are an author of the Elsevier article	No
Are you translating?	No
Order Reference Number	
Expected publication date	Aug 2010
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
Value added tax 0.0%	0.00 USD
Total	0.00 USD

ELSEVIER LICENSE TERMS AND CONDITIONS

May 18, 2010

This is a License Agreement between Troy Wynn ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Troy Wynn
Customer address	620 E 1000 N Apt. 4 Logan, UT 84321
License Number	2432040412734
License date	May 18, 2010
Licensed content publisher	Elsevier
Licensed content publication	Journal of Atmospheric and Solar-Terrestrial Physics
Licensed content title	Evidence for solar signals in the mesopause temperature variability?
Licensed content author	Kathrin Höppner, Michael Bittner
Licensed content date	April 2007
Volume number	69
Issue number	4-5
Pages	18
Type of Use	Thesis / Dissertation
Portion	Figures/tables/illustrations
Number of Figures/tables/illustrations	4
Format	Both print and electronic

You are an author of the Elsevier article	No
Are you translating?	No
Order Reference Number	
Expected publication date	Aug 2010
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
Value added tax 0.0%	0.00 USD
Total	0.00 USD

SPRINGER LICENSE TERMS AND CONDITIONS

Jun 04, 2010

This is a License Agreement between Troy Wynn ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2442160225093
License date	Jun 04, 2010
Licensed content publisher	Springer
Licensed content publication	Space Science Reviews
Licensed content title	Signature of the 11-Year Cycle in the Upper Atmosphere
Licensed content author	M.-L Chanin
Licensed content date	Aug 1, 2006
Volume number	125
Issue number	1
Type of Use	Thesis/Dissertation
Portion	Figures
Author of this Springer article	No
Order reference number	
Title of your thesis / dissertation	STATISTICAL ANALYSIS OF THE USU LIDAR DATA SET WITH REFERENCE TO MESOSPHERIC SOLAR RESPONSE AND COOLING RATE CALCULATIONS, WITH ANALYSIS OF STATISTICAL ISSUES AFFECTING THE REGRESSION COEFFICIENTS
Expected completion date	Aug 2010
Estimated size(pages)	150
Total	0.00 USD

APPENDIX B

DERIVATIONS FOR COEFFICIENT BIAS AND ATTENUATION

The proofs for under specified models and cross specified models were worked out by the author. The proof for coefficient bias in under specified models was taken from Draper and Smith *Applied Regression Analysis*, 3rd edition, pp. 235-236.

1. Under Specified Models

Suppose the proposed model is

$$E(\boldsymbol{Y}) = \boldsymbol{X}_1\boldsymbol{b}_1. \tag{B1}$$

The least squares coefficient estimates are

$$E(\boldsymbol{b}_1) = (\boldsymbol{X}_1{}^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1{}^T E(\boldsymbol{Y}_{true}). \tag{B2}$$

Now suppose the true model is this,

$$E(\boldsymbol{Y}_{true}) = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2. \tag{B3}$$

Putting these together gives

$$E(\boldsymbol{b}_1) = (\boldsymbol{X}_1{}^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1{}^T(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2), \tag{B4}$$

$$E(\boldsymbol{b}_1) = (\boldsymbol{X}_1{}^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1{}^T\mathbf{X}_1\boldsymbol{\beta}_1 + (\mathbf{X}_1{}^T\mathbf{X}_1)^{-1}\mathbf{X}_1{}^T\mathbf{X}_2\boldsymbol{\beta}_2, \tag{B5}$$

$$E(\boldsymbol{b}_1) = \boldsymbol{\beta}_1 + (\boldsymbol{X}_1{}^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1{}^T\boldsymbol{X}_2\boldsymbol{\beta}_2. \tag{B6}$$

If the proposed model is under specified, that is, if some but not all of the true model

variables are in the model, and the model contains no spurious explanatory variables, then the coefficients of the proposed model will be bias.

2. Over Specified Models

If the proposed model is

$$E(\boldsymbol{Y}) = \boldsymbol{X}_1\boldsymbol{b}_1 + \boldsymbol{X}_2\boldsymbol{b}_2, \quad \text{(B7)}$$

and the true model is

$$E(\boldsymbol{Y}_{true}) = \boldsymbol{X}_1\boldsymbol{\beta}_1, \quad \text{(B8)}$$

where $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are vectors of coefficients. $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are matrices of explanatory variables. The correctly specified variables will have unbiased coefficients and the incorrectly specified explanatory variables will have coefficients of value zero. The proof is as follows. The expected values for the coefficients $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are

. (B9)

This equation may be rewritten as

, (B10)

where $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ are submatrices. $\boldsymbol{A}$ and $\boldsymbol{D}$ are symmetric. $\boldsymbol{A} = \boldsymbol{X}_1{}^{T}\boldsymbol{X}_1$, $\boldsymbol{B} = \boldsymbol{X}_1{}^{T}\boldsymbol{X}_2$, $\boldsymbol{C} = \boldsymbol{X}_2{}^{T}\boldsymbol{X}_1$, $\boldsymbol{D} = \boldsymbol{X}_2{}^{T}\boldsymbol{X}_2$.The matrix inverse can be calculated using blockwise inversion. That is,

, (B11)

Where $\boldsymbol{A}$ and $\boldsymbol{D}$ must be invertible and therefore square. This formula can be easily checked by multiplying it by the original matrix. Thus

$$\boldsymbol{b}_1 = [(\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1})\boldsymbol{A} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}]\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1 \quad \text{(B12)}$$

$$\boldsymbol{b}_2 = [\ -(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}]\boldsymbol{A} + [(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}]\boldsymbol{C} = \boldsymbol{0} \quad \text{(B13)}$$

For an over specified model all the relevant explanatory variables are included in the proposed model along with some spurious (not-true-model) explanatory variables. The coefficient values for the spurious explanatory variables are zero and the correctly specified variables are unbiased.

3. Cross Specified Models

Suppose the proposed model is this,

$$\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{X}_1\boldsymbol{b}_1 + \boldsymbol{X}_2\boldsymbol{b}_2, \quad \text{(B14)}$$

and the true model is this

$$\mathrm{E}(\boldsymbol{Y}_{\text{true}}) = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_3\boldsymbol{\beta}_3, \quad \text{(B15)}$$

where $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_3$ are coefficients vectors. $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$ are matrices of explanatory variables. The matrix $\boldsymbol{X}_1$ contains explanatory variables common to both the true and proposed models. The coefficients $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are

. (B16)

This may be rewritten as

, (B17)

where $\boldsymbol{Q} = \boldsymbol{X}_1^T\boldsymbol{X}_3$, and $\boldsymbol{R} = \boldsymbol{X}_2^T\boldsymbol{X}_3$, $\boldsymbol{A} = \boldsymbol{X}_1^T\boldsymbol{X}_1$, $\boldsymbol{B} = \boldsymbol{X}_1^T\boldsymbol{X}_2$, $\boldsymbol{C} = \boldsymbol{X}_2^T\boldsymbol{X}_1$, $\boldsymbol{D} = \boldsymbol{X}_2^T\boldsymbol{X}_2$. The matrix may be inverted using the blockwise inversion method mentioned above. Again, $\boldsymbol{A}$ and $\boldsymbol{D}$ are symmetric. From this we get

$$\boldsymbol{b}_1 = \{[\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}]\boldsymbol{A} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\}\boldsymbol{\beta}_1 +$$

$$\{[\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}]\boldsymbol{Q} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{R}\}\boldsymbol{\beta}_3 =$$

$$= \boldsymbol{\beta}_1 + \{[\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}]\boldsymbol{Q} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{R}\}\boldsymbol{\beta}_3 = \boldsymbol{\beta}_1 + \boldsymbol{bias}. \quad \text{(B18)}$$

$$\boldsymbol{b}_2 = \{ -(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{A} + (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C} \}\boldsymbol{\beta}_1 +$$

$$\{ -(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{Q} + (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{R} \}\boldsymbol{\beta}_3 =$$

$$\boldsymbol{0} + \{ -(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{Q} + (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{R} \}\boldsymbol{\beta}_3 = \boldsymbol{(Attenuation\ Factor)}\,\boldsymbol{\beta}_3. \quad \text{(B19)}$$

The correctly specified explanatory variables have bias coefficients and the incorrectly specified coefficients are the attenuated, omitted true model coefficients.

APPENDIX C

CALCULATING THE STANDARD ERRORS IN NONLINEAR REGRESSION MODELS

The formula for the standard errors of coefficients $\boldsymbol{b}$ is found from the estimated variance-covariance matrix of $\boldsymbol{b}$

$$var(\boldsymbol{b}) = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\cdot\sigma^2, \tag{C1}$$

where $\boldsymbol{A}$ forms the bases of the regression; σ^2 is the mean variance. If the σ^2 is unknown then s_2 may be used. The square roots of the diagonal elements are the SEs of the elements of $\boldsymbol{b}$. This calculation is simple for linear models like

$$y = I\cdot\boldsymbol{\iota} + \beta\cdot\boldsymbol{t} + A_1\cdot sin(2\pi\cdot\boldsymbol{t}) + A_2\cdot cos(2\pi\cdot\boldsymbol{t}) + B_1\cdot sin(4\pi\cdot\boldsymbol{t}) + B_2 cos(4\pi\cdot\boldsymbol{t}) + C_1 sin(\omega\cdot\boldsymbol{t}) + C_2\cdot cos(\omega\cdot\boldsymbol{t}) + \boldsymbol{\varepsilon}\,.$$

The data space is $\boldsymbol{A}$=($\boldsymbol{\iota}$, $\boldsymbol{t}$, $sin(2\pi\cdot\boldsymbol{t})$, $cos(2\pi\cdot\boldsymbol{t})$, $sin(4\pi\cdot\boldsymbol{t})$, $cos(4\pi\cdot\boldsymbol{t})$, $sin(\omega\cdot\boldsymbol{t})$, $cos(\omega\cdot\boldsymbol{t})$). However, the following model covers the same data space.

$$y = I\cdot\boldsymbol{\iota} + \beta\cdot\boldsymbol{t} + S\cdot sin(2\pi\cdot\boldsymbol{t} + \gamma) + B\cdot sin(4\pi\cdot\boldsymbol{t} + \varphi) + C\cdot sin(\omega\cdot\boldsymbol{t} + \theta) + \boldsymbol{\varepsilon} \tag{C2}$$

The SEs for (C2) are calculated by taking the partial derivatives with respect to the coefficients of interest. An iterative procedure may be used to solve for I, β, S, γ, B, φ, C, and θ. Once found, their SEs may be calculated by applying Equation (C1), where $\boldsymbol{b}$ = (I, β, S, γ, B, φ, C, θ) and $\boldsymbol{A}$ is comprised of the partial derivatives of (C2) with respect to the coefficients of interest. That is $\boldsymbol{A}$ = [$\partial_I y$, $\partial_\beta y$, $\partial_S y$, $\partial_\gamma y$, $\partial_B y$, $\partial_\varphi y$, $\partial_C y$, $\partial_\theta y$], or $\boldsymbol{A}$ = [$\boldsymbol{\iota}$, $\boldsymbol{t}$,

$sin(2\pi \cdot \boldsymbol{t} + \gamma)$, $S \cdot cos(2\pi \cdot \boldsymbol{t} + \gamma)$, $sin(4\pi \cdot \boldsymbol{t} + \varphi)$, $B \cdot cos(4\pi \cdot \boldsymbol{t} + \varphi)$, $sin(\omega \cdot \boldsymbol{t} + \theta)$, $C \cdot cos(\omega \cdot \boldsymbol{t} + \theta)$].

This is used in Equation (C1). The square root of the diagonal elements of the RHS of (C1) are the SEs of the coefficients of $\boldsymbol{b}$.

APPENDIX D

ISSUES RELATING TO COMMON NOISE

Abstract. One issue of concern is whether a model regressor has a noise component that is correlated to the noise in the dependant variable. That is, if the noise in a signal x is correlated to the noise in y then the possibility of diagnosing x as a good regressor for y is not insignificant, even when x and y are otherwise very dissimilar.

1. Introduction

One of the issues affecting the confidence levels in regression coefficients is noise between independent and dependent variables. Our simulations show that if there is a correlation between the noise in x and y, there is a non-negligible risk that x will be considered a good regressor for y even though the underlying signals may be unrelated. This is important to address because in many cases where an ordinary least squares (OLS) regression is used, it is the coefficient value and not the noise that is of particular interest.

Here we assume that $\boldsymbol{x} = \boldsymbol{s}_2 + \boldsymbol{n}_2$ and $\boldsymbol{y} = \boldsymbol{s}_1 + \boldsymbol{n}_1$, where $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are the underlying signals and $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are the noise components. We derive equations for the correlation between x and y when $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are orthogonal and uncorrelated and a correlation exists only between the noise. We derive an equation for the statistical significance of the coefficient b in $\boldsymbol{y} = \boldsymbol{I} + b \cdot \boldsymbol{x} + \boldsymbol{\varepsilon}$, where $\boldsymbol{I}$ is the intercept and $\boldsymbol{\varepsilon}$ the residual. We use this result to get the equation for the statistical significance of b when $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are uncorrelated and orthogonal.

One of the models employed in this dissertation is,

$$\boldsymbol{T = I + LT + AO + SAO + SOL + N},$$

where ***I*** is the intercept, ***LT*** is the linear term, ***AO*** comprises the annual oscillation, ***SAO*** the semiannual oscillation, ***SOL*** is a solar proxy, and are ***N*** the residuals. The coefficients are implied. This model assumes that ***T*** is a function of ***I***, ***LT***, ***AO***, ***SAO***, and ***SOL***. The explanatory variables ***I***, ***LT***, ***SO***, and ***SAO*** are either constant or functions (such as $sin\ \omega t$ and $cos\ 4\pi t$) and consequently do not have noise components. The remaining variable ***SOL*** is typically a solar proxy, which does have a noise component. It is the noise in ***SOL*** that can be correlated to the noise in ***T***. An example is shown in Figure D1. Though the two underlying functions are dissimilar the correlation between them is 0.10 because of correlation between their noise components; the confidence level for ***x*** regressed on ***y*** is 99.1%.

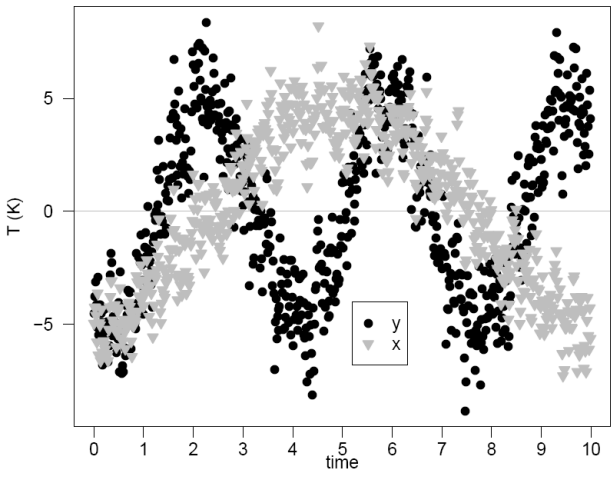


Figure D1. A plot of two signals having common noise. Fitting **x** to **y** produced a statistical confidence level of greater than 99.1%, even though the two signals have unrelated shapes.

2. Some Notation

But before going further, there are some notation issues that must be addressed. One quantity that is frequently used in statistical analysis is the Total Sum of Squares (TSS). For a single variable this is

$$TSS_x = \sum_i (X_i - \underline{X})^2,$$

where $\underline{X}$ is the mean of the X_is. For two variables, TSS_{xy} is

$$TSS_{xy} = \sum_i (X_i - \underline{X})(Y_i - \underline{Y}).$$

3. Derivation of Correlation

The two signals of interest are $\boldsymbol{y} = \boldsymbol{s}_1 + \boldsymbol{n}_1$ and $\boldsymbol{x} = \boldsymbol{s}_2 + \boldsymbol{n}_2$, where $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are the underlying signals, and $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ the noise components. For simplicity $\boldsymbol{y}$, $\boldsymbol{x}$, $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, $\boldsymbol{n}_1$, and $\boldsymbol{n}_2$ are mean centered. We assume that $\boldsymbol{s}_1$ is uncorrelated with $\boldsymbol{n}_1$ and that $\boldsymbol{s}_2$ is uncorrelated with $\boldsymbol{n}_2$. The correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$r = TSS_{xy}/\sqrt{(TSS_x \cdot TSS_y)}, \qquad \text{(D1)}$$

where

$$TSS_x = \sum s_{2,i}^2 + \sum n_{2,i}^2 = s_2^2 + n_2^2,$$

$$TSS_y = \sum s_{1,i}^2 + \sum n_{1,i}^2 = s_1^2 + n_1^2,$$

$$TSS_{xy} = \sum y_i x_i = \sum (s_{1,i} + n_{1,i}) \cdot (s_{2,i} + n_{2,i}) = TSS_{s1,s2} + TSS_{n1,n2}.$$

For this case we assume $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are uncorrelated and orthogonal. In terms of a linear regression this means that x has no explanatory value for $\boldsymbol{y}$. With these assumptions we

then have $TSS_{xy} = \sum n_{1,i} \cdot n_{2,i}$. We take $n_{12} \equiv \sum n_{1,i} \cdot n_{2,i}$. Putting these terms into (D1) we get

$$\frac{}{} \quad \frac{}{} \quad \frac{}{}, \qquad \text{(D2)}$$

where ρ is the correlation between n_1 and n_2. (D2) can be rewritten as

$$\frac{}{}. \qquad \text{(D3)}$$

From (D3) we see that the correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$ is directly proportional to the correlation between their noise components. The ratio terms in the denominator are the variance of the signal to the variance of the noise.

4. Statistical Test on b

The equation for the linear coefficient b from $\boldsymbol{y} = \boldsymbol{I} + b \cdot \boldsymbol{x}$ is

$$b = (y_i x_i) / x_i^2$$

A t-test can be applied to determine its level of significance. It was easier to work with the squared t-statistic, which is

$$t^2 = b^2 / SE^2,$$

where $SE^2 = se^2 / x_i^2$ and $se^2 = (y_i^2 - (x_i y_i)^2 / x_i^2) / (n - 1)$. If $t_{theoretical} \leq t_{coef}$ then b is considered statistically significant to the predetermined level of confidence. Substituting se^2 and SE^2 into t^2 we get

Some simplifying steps

until we get

. (D4)

For 95% confidence with n>100 t_{th} = 1.96, which can be rounded to 2. Recognizing that the RHS term is the squared correlation we get

. (D5)

If this inequality holds, then $\boldsymbol{x}$ will be considered a good regressor for $\boldsymbol{y}$ with at least 95% confidence. As $\boldsymbol{n}$ increases this will become more likely. A correlation of 8% will satisfy the minimum requirement for a data set of 593 data points. A correlation of 6% will satisfy this requirement for a data set of 1000 points.

We can now use Equation (D4) to look at cases where $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are uncorrelated and orthogonal, where the correlation is between $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$. Applying this condition to Equation (D4) we get

$$\overline{\quad} \qquad \overline{\quad} \qquad . \tag{D6}$$

For b to considered statistically significant at 95% or greater, we apply the same condition for the t-statistic as in the previous case (n>100, t = 2). We now get

$$\overline{\quad} \quad \frac{\qquad\qquad}{\overline{\quad} \quad \overline{\quad}} \ , \tag{D7}$$

where the σs are variances of $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, $\boldsymbol{n}_1$, and $\boldsymbol{n}_2$; ρ is the correlation between $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$, and n is the number of data points. As with Equation (D5), Equation (D7) shows us that more data increases the statistical significance of b, and consequently the likelihood that x will be considered a good regressor for y when, in fact, this is not true.

This result has particular significance in correlation studies. Several studies posit a relationship between solar input and an atmospheric response by looking for a correlation between a solar proxy and atmospheric temperatures. For example *Beig and Fadnavis* [2009]: “In order to detect whether real solar signal is present in the time series, the

correlation between the time series (before subjecting to the regression model) and F10.7 cm solar flux is computed at every altitude" (parenthesis original). Another example of this is from *Mohanakumar* [1985] who found correlations between sunspot number and temperatures ranging from −0.13 to 0.95. Yet another example is from *Höppner and Bittner* [2007] who note that *Batista et al.* [1994] found a positive correlation between the F10.7 cm solar flux and OH* temperatures over Paulista, Brazil for the time period from 1977 to 1986. However, their data from 1990 to 2005 shows an anticorrelaton. When combining their OH* data from Wuppertal, Germany with that of Batista they found no conclusive evidence for a correlation between the F10.7 cm solar flux and OH* temperatures. (For two other correlation studies see *White et al.* [1997] and *Cùrrie* [1991] and references therein.)

5. Conclusions

The presence of autocorrelated noise in regressor $\boldsymbol{x}$ and a regressand $\boldsymbol{y}$ can cause $\boldsymbol{x}$ to appear to have explanatory value for $\boldsymbol{y}$ even when their underlying signals are unrelated. This risk increases as the number of data points increases. In practical terms one cannot say for sure if a statistically significant correlation between a solar proxy and a temperature time series is due to a common sine-like variation in the temperatures or a correlation between the temperature residuals and the proxy residuals.

We also know that the atmospheric solar response can be significantly out of phase to the solar input, but this might not apply to the solar-noise. If they are not separated then there exists a risk of a false positive: $\boldsymbol{x}$ could be considered a good regressor for $\boldsymbol{y}$ when it is not.

References

Batista, P., H. Takahashi, and B. Clemesha (1994), Solar cycle and the QBO effect on the mesospheric temperature and nightglow emissions at a low latitude station, *Adv. Space Res.*, *14*(9), 221-224, doi:10.1016/0273-1177(94)90139-2.

Beig, G., and S. Fadnavis (2009), Solar response in the temperature over the equatorial middle atmosphere, *J. Atmos. Sol. Terr. Phys.*, *71*(13), 1450-1455, doi:10.1016/j.jastp.2008.07.007.

Cùrrie, R. G. (1991), Luni-solar 18.6-year signal in tree-rings from Argentina and Chile, *Pure Appl. Geophys*., *137*(3), 281-300, doi:10.1007/BF00876993.

Höppner, K., and M. Bittner (2007), Evidence for solar signals in the mesopause temperature variability?, *J. Atmos. Sol. Terr. Phys.*, *69*(4-5), 431-448, doi:10.1016/j.jastp.2006.10.007.

Mohanakumar, K. (1985), An investigation on the influence of solar cycle on mesospheric temperature, *Planet. Space Sci.*, *33*(7), 795-805, doi:10.1016/0032-0633(85)90033-9.

White, W. B., J. Lean, D. R. Cayan, and M. D. Dettinger (1997), Response of global upper ocean temperature to changing solar irradiance, *J. Geophys. Res.*, *102*(C2), PAGES 3255–3266.

APPENDIX E

A METHOD TO ESTIMATE THE AUTOCORRELATION COEFFICIENT IN DATA WITH LARGE DATA GAPS

Abstract. This presents a way to estimate the autocorrelation coefficient for autocorrelated data containing large data gaps.

1. Introduction

Autocorrelation occurs when a given data point is correlated to one or more of the previous data points. There are ways to calculate the autocorrelation coefficient if there are no data gaps. But as is often the case with measurements that depend on the weather, significant data gaps exist in some data sets. The USU data set has 593 nightly profiles out of 3,623 days covering approximately 10 years. This makes fitting for the correlation coefficient more challenging. The USU data set has 16% of possible daily observations. While this seems low, it is not unusual. Over a similar time span (October 1978 through December 1989) the French lidar at Haute-Provence collected a total of 872 nights of observations out of 4,110 possible, amounting to 21.3% of possible observations [*Hauchecorne et al.*, 1991]. In cases where the data gaps are small and infrequent interpolation is acceptable, otherwise modeling the correlation structure is more difficult.

In the USU data set, out of 593 data points 420 (71%) are consecutive, 161 have two-day spacing, 86 have three-day spacing, and 57 have four-day spacing—there is some overlap owing to the fact that some data points that are one day apart are also two days from another data point, etc.

2. Autocorrelation

This paper focuses on first order autocorrelation AR(1). This can be described by the following equations.

$$N_1 = \varphi N_0 + \varepsilon_1$$

$$N_2 = \varphi N_1 + \varepsilon_2$$

$$N_3 = \varphi N_2 + \varepsilon_3$$

$$N_4 = \varphi N_3 + \varepsilon_4$$

...

Any given measurement N depends on the previous N multiplied by the autocorrelation coefficient φ, plus some noise ε. These equations can be expanded by substituting N_i into the N_{i+1} etc., which yields

$$N_1 = \varphi N_0 + \varepsilon_1$$

$$N_2 = \varphi^2 N_0 + \varphi\varepsilon_1 + \varepsilon_2$$

$$N_3 = \varphi^3 N_0 + \varphi^2\varepsilon_1 + \varphi\varepsilon_2 + \varepsilon_3$$

$$N_4 = \varphi^4 N_0 + \varphi^3\varepsilon_1 + \varphi^2\varepsilon_2 + \varphi\varepsilon_3 + \varepsilon_4$$

$$N_5 = \varphi^5 N_0 + \varphi^4\varepsilon_1 + \varphi^3\varepsilon_2 + \varphi^2\varepsilon_3 + \varphi\varepsilon_4 + \varepsilon_5$$

$$N_6 = \varphi^6 N_0 + \varphi^5\varepsilon_1 + \varphi^4\varepsilon_2 + \varphi^3\varepsilon_3 + \varphi^2\varepsilon_4 + \varphi\varepsilon_5 + \varepsilon_6$$

...

This can be written as

$$N_k = \varepsilon_k + \varphi\,\varepsilon_{k-1} + \varphi^2\,\varepsilon_{k-2} + \varphi^3\,\varepsilon_{k-3} + \ldots + \varphi^{n-1}\,\varepsilon_{k-n-1} + \varphi^n\,\varepsilon_{k-n} \qquad \text{(E1)}$$

where the εs are assumed to be random with identical variance and zero mean.

3. A Simple Case with no Data Gaps

If the correlation coefficient φ is known it is possible to solve for the εs and unzip the time series. Because we do not have data prior to N_0, N_0 must be taken to be equal to ε_0. Solving for the εs we get

$$\varepsilon_0 = N_0 \tag{E2}$$

$$\varepsilon_1 = N_1 - \varphi N_0$$

$$\varepsilon_2 = N_2 - \varphi N_1$$

...

This can be expanded as

$$\varepsilon_0 = N_0$$

$$\varepsilon_1 = N_1 - (\varphi\varepsilon_0)$$

$$\varepsilon_2 = N_2 - (\varphi^2\varepsilon_0 + \varphi\varepsilon_1)$$

$$\varepsilon_3 = N_3 - (\varphi^3\varepsilon_0 + \varphi^2\varepsilon_1 + \varphi\varepsilon_2)$$

$$\varepsilon_4 = N_4 - (\varphi^4\varepsilon_0 + \varphi^3\varepsilon_1 + \varphi^2\varepsilon_2 + \varphi\varepsilon_3)$$

$$\varepsilon_5 = N_5 - (\varphi^5\varepsilon_0 + \varphi^4\varepsilon_1 + \varphi^3\varepsilon_2 + \varphi^2\varepsilon_3 + \varphi\varepsilon_4)$$

$$\varepsilon_6 = N_6 - (\varphi^6\varepsilon_0 + \varphi^5\varepsilon_1 + \varphi^4\varepsilon_2 + \varphi^3\varepsilon_3 + \varphi^2\varepsilon_4 + \varphi\varepsilon_5)$$

...

In the presence of data gaps, this sequence looks a little different. If there are missing data points, then we sum the terms in the parenthesis leaving out the missing εs. For

example if the first six days in the data set are located at days one, two, three, five, nine, and sixteen then we would have (counting from zero)

$$\begin{aligned}
\varepsilon_0 &= N_0 \\
\varepsilon_1 &= N_1 - \varphi\varepsilon_0 \\
\varepsilon_2 &= N_2 - (\varphi^2\varepsilon_0 + \varphi\varepsilon_1) \\
\varepsilon_4 &= N_4 - (\varphi^4\varepsilon_0 + \varphi^3\varepsilon_1 + \varphi^2\varepsilon_2) \\
\varepsilon_8 &= N_8 - (\varphi^8\varepsilon_0 + \varphi^7\varepsilon_1 + \varphi^6\varepsilon_2 + \varphi^4\varepsilon_4) \\
\varepsilon_{15} &= N_{15} - (\varphi^{15}\varepsilon_0 + \varphi^{14}\varepsilon_1 + \varphi^{13}\varepsilon_2 + \varphi^{11}\varepsilon_4 + \varphi^7\varepsilon_8) \qquad \text{(E3)} \\
&\ldots
\end{aligned}$$

The following procedure is used to estimate the autocorrelation coefficient. First a value for φ is guessed; the sequence (E3) is then used to unzip the εs, then their variance is taken. This process is repeated for a different value of φ. The φ that minimizes $\sigma^2(\varepsilon)$ is selected as representative of the true correlation coefficient. This claim is a conjecture supported by Monte Carlo simulations.

A plot of $\sigma^2(\varepsilon)$ versus φ is shown in Figure E1 for data without gaps. It clearly shows that the minimum σ^2 is located near the true value of the autocorrelation coefficient. In practice we employed the Newton-Raphson method to locate the minimum value: $\varphi_{i+1} = \varphi_i - \sigma^2(\varphi)'/\sigma^2(\varphi)''$, where $\sigma^2(\varphi)'$ and $\sigma^2(\varphi)''$ are first and second derivatives, calculated numerically.

4. For Data Gaps

The following derivation demonstrates more precisely how a correct guess of the

autocorrelation coefficient can minimize the residuals of the εs.

The following equation represents an first order autoregressive, AR(1), signal.

, (E4)

where N_n is the measured signal, ε_n is the random component, and φ the correlation coefficient. This can be rewritten as

. (E5)

To represent a data set containing missing data points multiply the terms in the summation by an indicator function I.

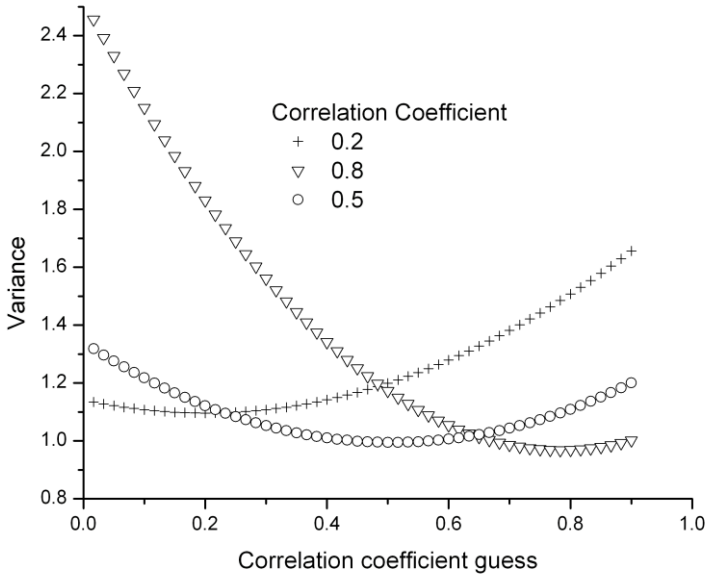


Figure E1. How the variance of the εs change with guessed correlation coefficient. The minimum can be seen to be near the true value of the correlation coefficient. The data has no gaps.

. (E6)

The indicator I_i has a value of 1 if the ith data point is present and 0 if it is absent. I^* is the converse indicator. If I_i* has a value of 1 then the ith data point is missing, and 0 if it is present. Note that $I_i + I_i^* = 1$ for all i. What (E6) represents is something like

$$\varepsilon_7 = N_7 - (\varphi^7 N_0 + \varphi^6 \varepsilon_1 + \varphi^5 \varepsilon_2 + \varphi^4 \varepsilon_3 + \varphi^3 \varepsilon_4 + \varphi^2 \varepsilon_5 + \varphi\, \varepsilon_7),$$

where the grayed out terms represent examples of missing data points excluded from the summation by the indicator function. Equation (E6) can be taken to represent our attempt to unzip the εs when some of the data points are missing. However, a measured data point contains information about the past and hence, an actual measurement does not experience gaps. Equation (E4) can be taken to represent an actual data measurement. Replacing N_n in (E6) with (E4) gives

, (E7)

where ξ is the true correlation coefficient and a_i is the noise. Multiplying the terms in the first summation by $I + I^*$ (this is just multiplying by 1) gives

. (E8)

Then collecting terms

. (E9)

And finally taking the variance of (E9)

. (E10)

If $\varphi = \xi$ and $a_i \approx \varepsilon_i$, then the variance is minimized by minimizing the third term on the RHS. While this is not a proof, it helps to illustrate the validity of the concept.

One issue that needs to be addressed is the distribution of the coefficient estimate. Every data set has a unique pattern of data gaps. So how does one estimate the distribution of the autocorrelation estimate? This must be done with a Monte Carlo simulation. When a correlation estimate is obtained, a set of autocorrelated noise is generated from that correlation value and data gaps matching the original data are applied. The autocorrelation coefficient is estimated from the simulated data set. When the above procedure is performed repeatedly and a distribution for that autocorrelation value is obtained.

Simulations were conducted to test this procedure. A series of autocorrelated Ns with a known φ were generated and the above procedure was applied. A distribution for φ was generated. This was done for several values of φ. This was repeated for simulated data with no data gaps, for USU data gaps, and for USU data gaps but only including consecutive points. The distributions for the φs for all these cases are given at the end of this section of the Appendix in Figures E3, E4, and E5. A plot of the standard deviations for the φ-distributions for the three simulations is given in Figure E2. We see that the method that produced the smallest standard deviation for φ comes for the simulation with no data gaps. The next best estimate is for the simulation with USU data gaps. And the worst estimate is obtained from simulations with USU data gaps selecting only the

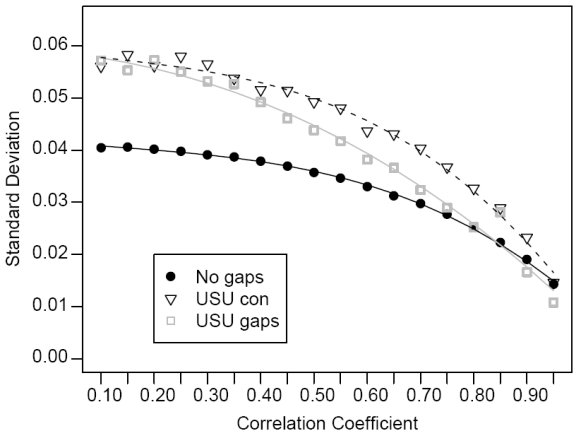


Figure E2. The standard deviations for the no gaps case, the USU data gaps, and USU consecutive (USU con) data points only.

adjacent points.

References

Hauchecorne, A., M. Chanin, and P. Keckhut (1991), Climatology and trends of the middle atmospheric temperature (33–87 km) as seen by Rayleigh lidar over the south of France, *J. Geophys. Res.*, *96*(D8), 15297-15309, doi:10.1029/91JD01213.

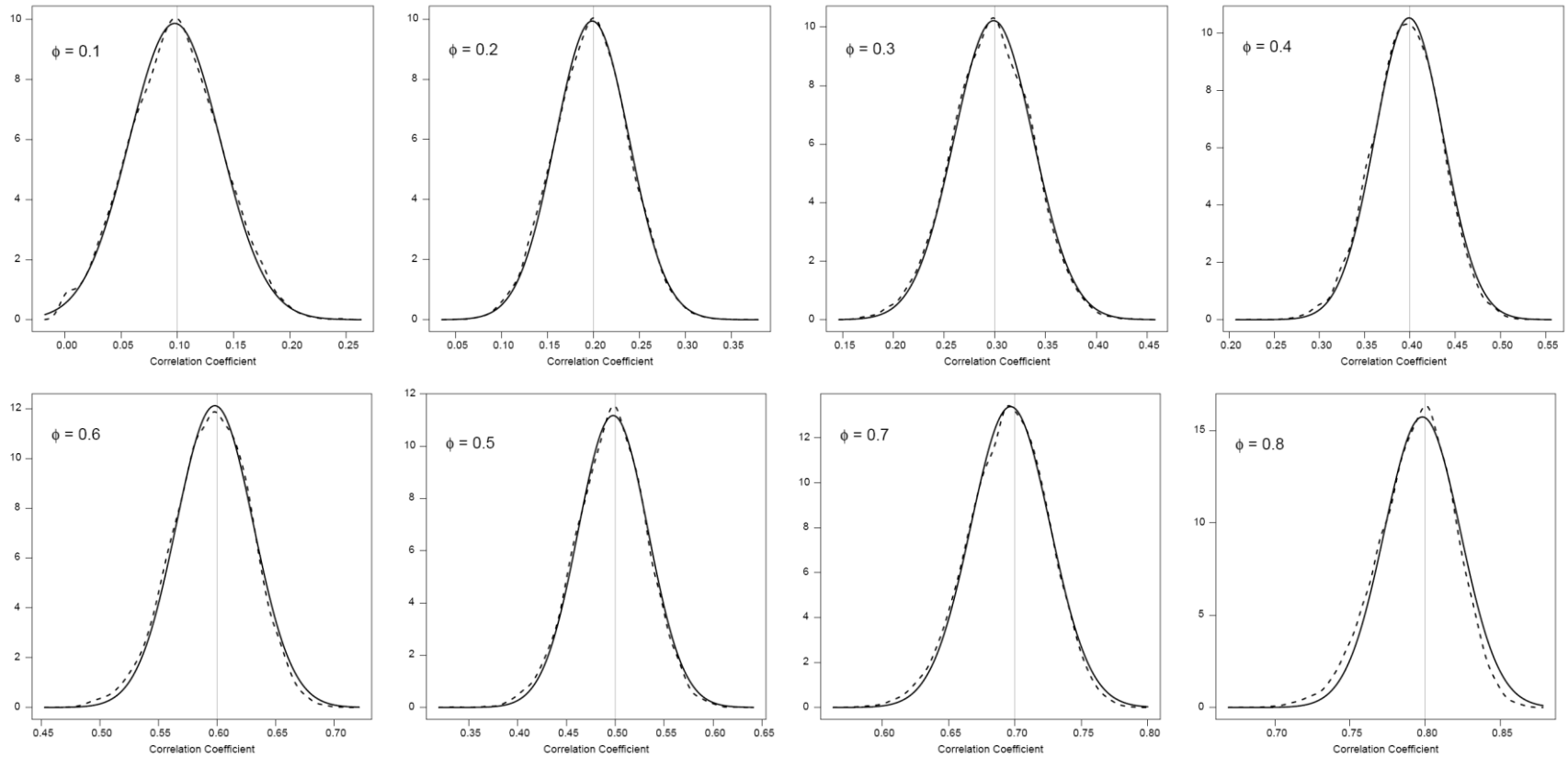


Figure E3. Distributions for various correlation coefficients from Monte Carlo simulations, no data gaps. The distributions for various correlation coefficients from Monte Carlo simulations are shown in the dashed lines. A Gaussian fit are shown in the solid lines. This figure is for data containing no data gaps.

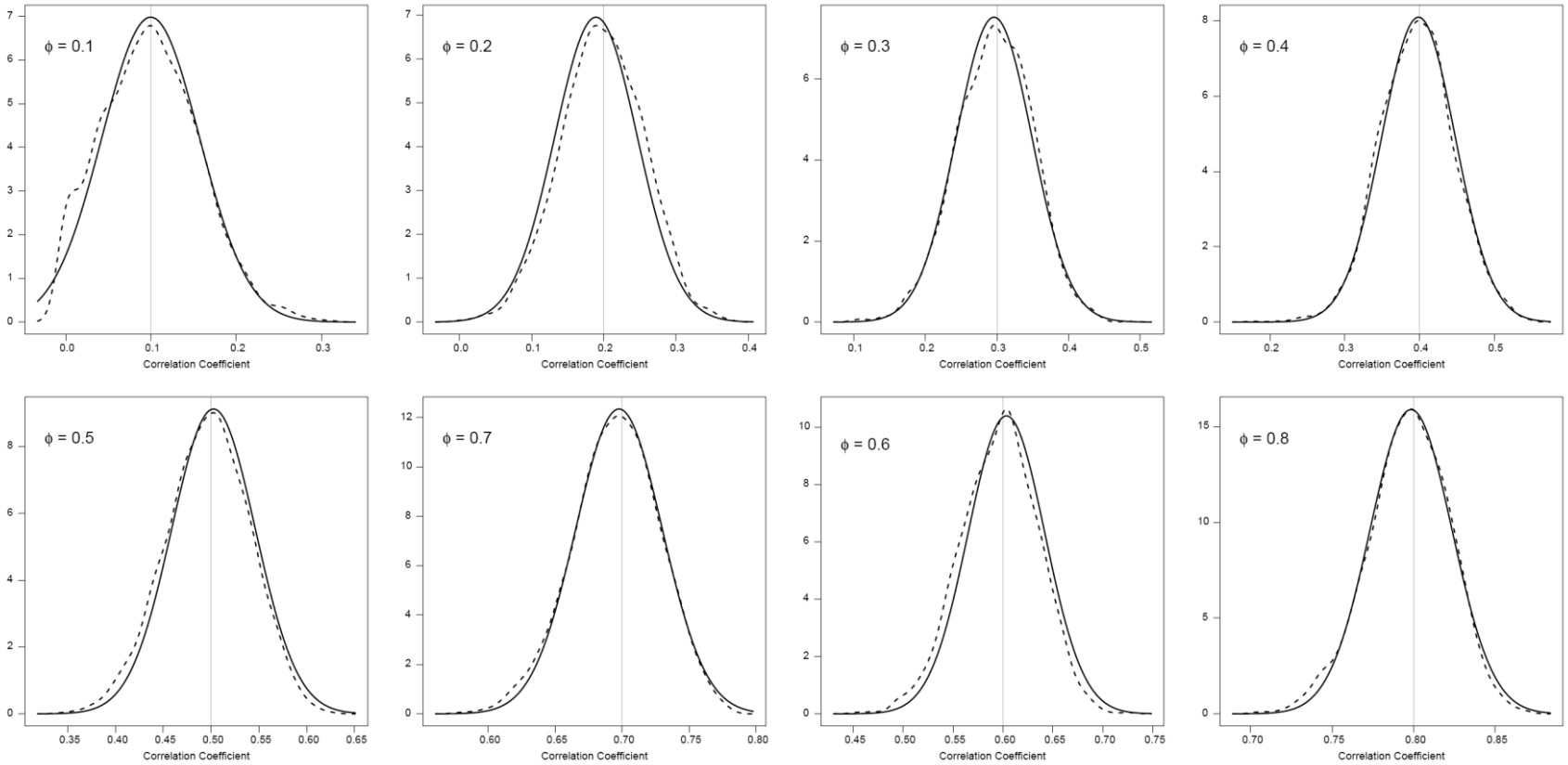


Figure E4. Same as Figure E2 but for data replicating the data gaps in the USU data set.

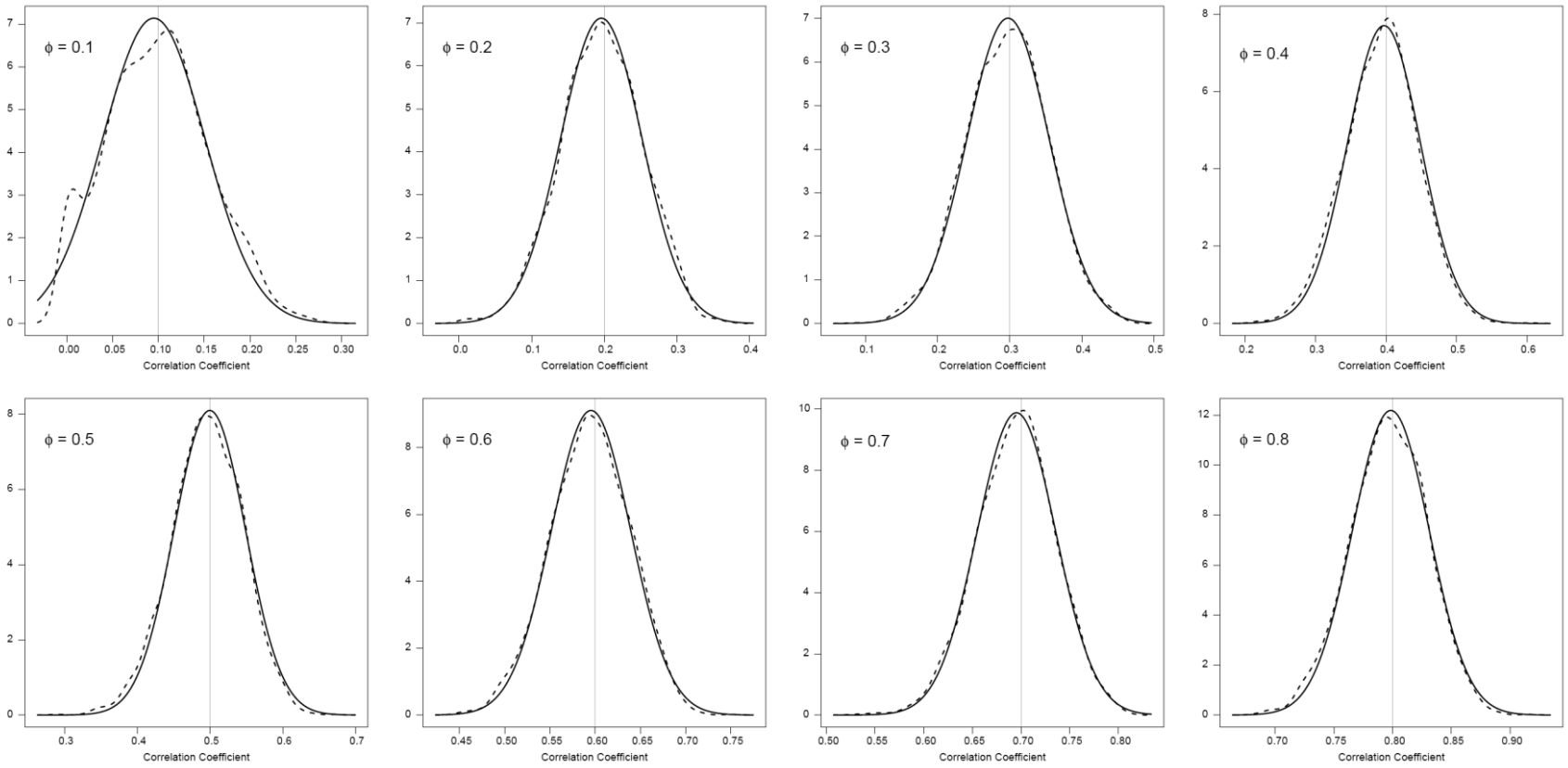


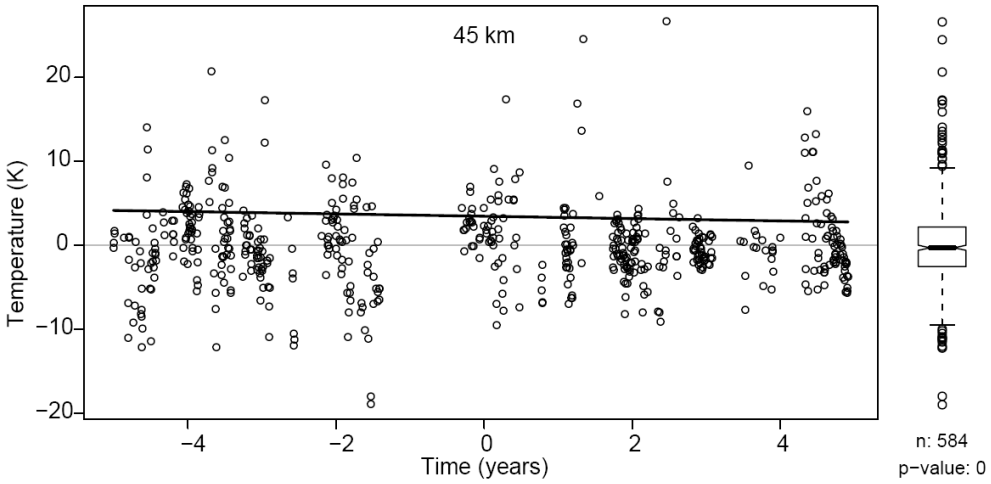
Figure E5. Same as Figure E2 but for USU data gaps where only consecutive data points are selected.

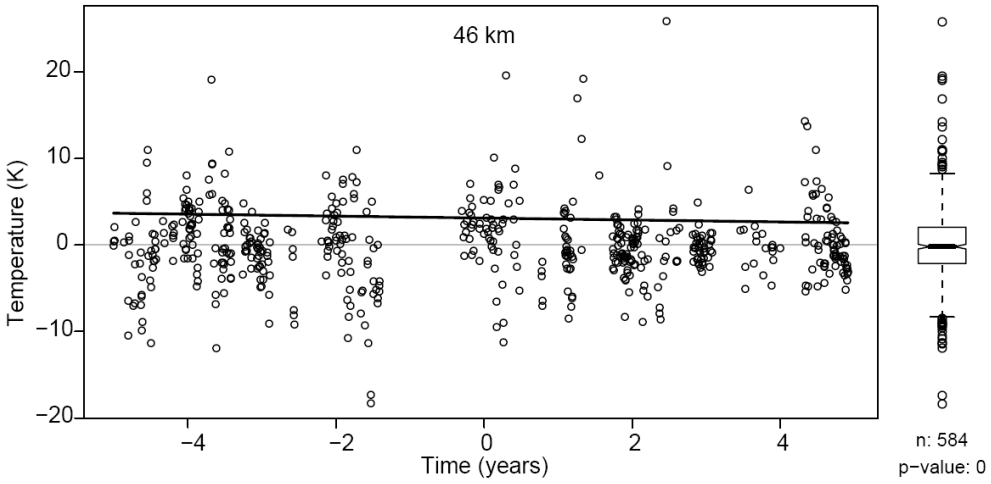
APPENDIX F

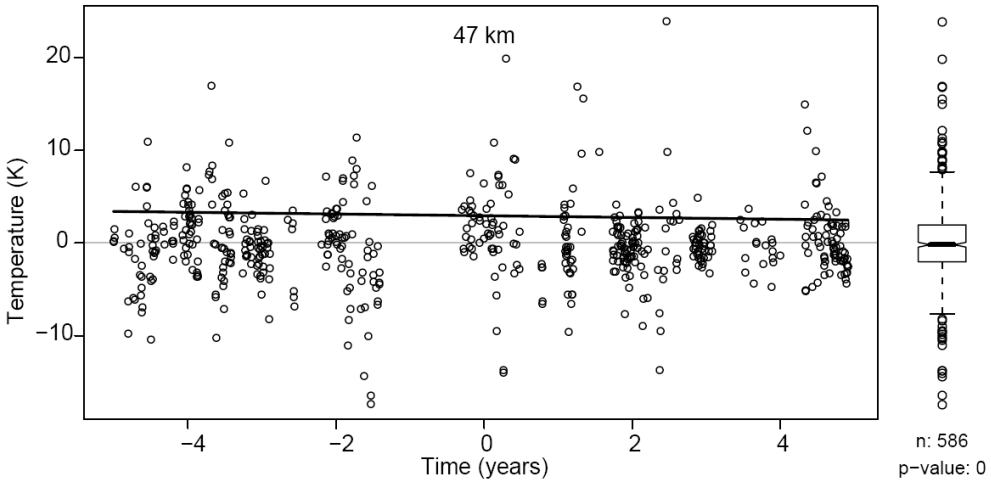
RESIDUALS PLOTS: 45-90 km

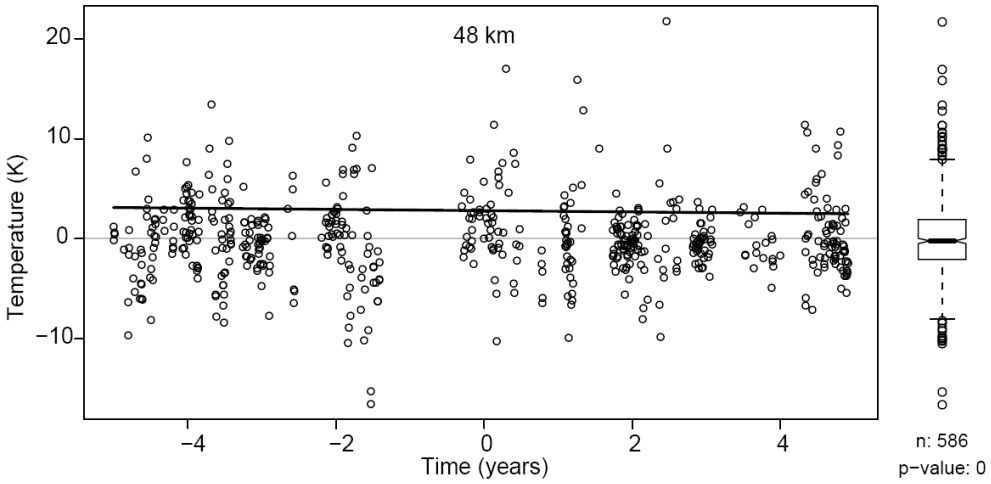
One easily accessible measure of heteroscedasticity is the Breusch–Pagan test. In this test the squares of the residuals are fit to a linear model $\varepsilon^2 \sim \beta_o + \beta_1 t$. If an F-test shows that the model variables are jointly significant, then there is evidence for the presence of heteroscedasticity. The test was performed on the residuals from the USU data using the bptest function in the lmtest library in the R programming language.

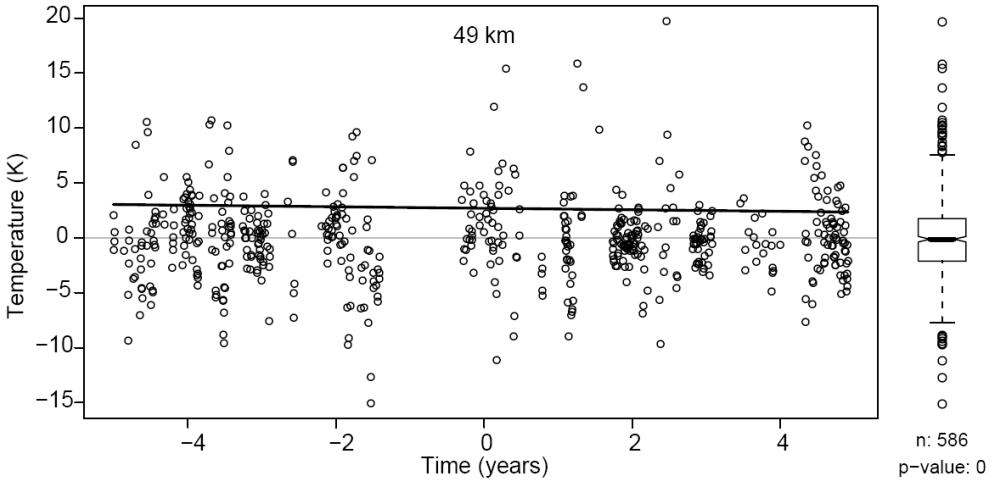
Given below are residual plots with a linear fit to the $|\varepsilon|$ shown above the abscissa. This was added to make the heteroscedasticity more visually accessible. A box plot of the residuals is shown to the right along with the p-value from Breusch–Pagan (BP) test along with the number of data points. A lower p-value indicates a higher probability that the data is heteroscedastic. P-values less than 0.1 and 0.05 correspond to 90% and 95% confidence, respectively. From the BP tests it appears that at most altitudes the residuals show some heteroscedasticity. The numbers shown in the plots are rounded to three significant figures so at many altitudes it rounds to zero. At 6 altitude levels the BP p-value was greater than 0.1, indicating that at most altitudes heteroscedasticity was a problem. Is not entirely surprising due to the fact that we found an 11-year variation in residual standard deviation.

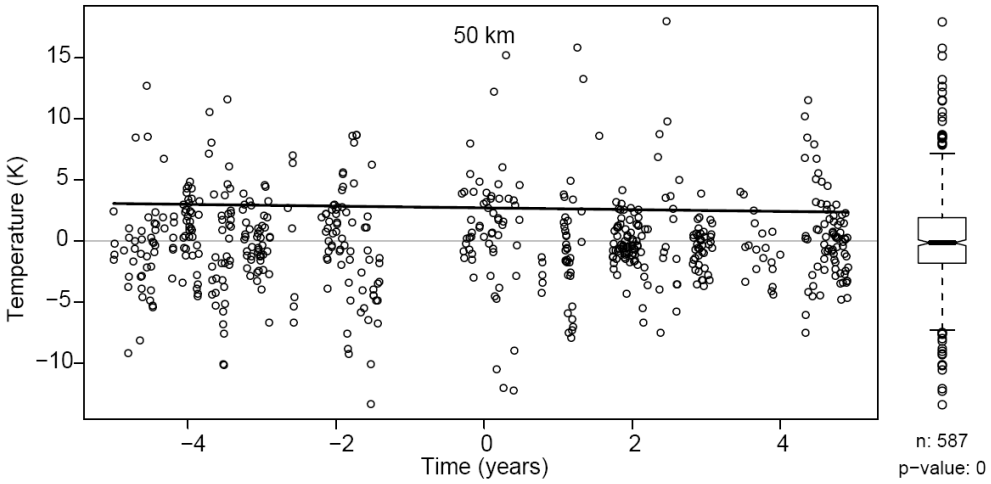

45 km
Temperature (K)
Time (years)
n: 584
p−value: 0

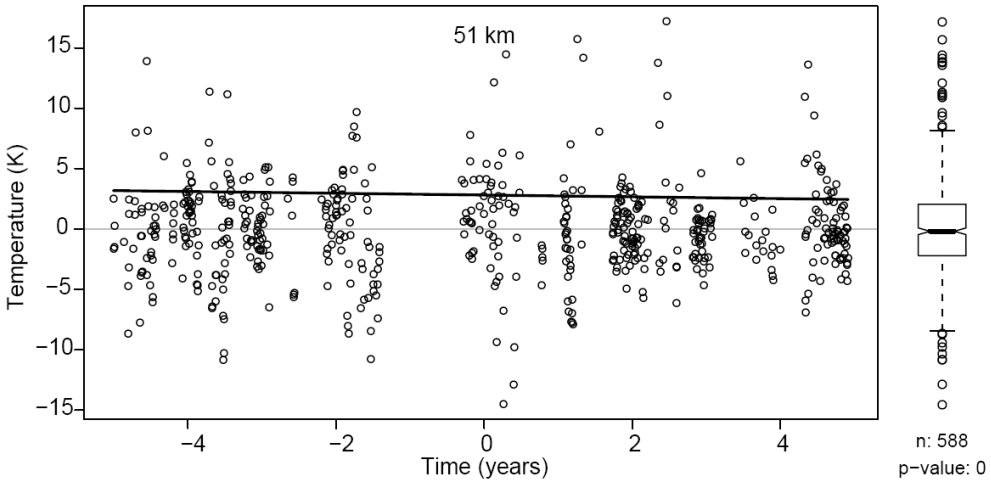

46 km
Temperature (K)
Time (years)
n: 584
p−value: 0

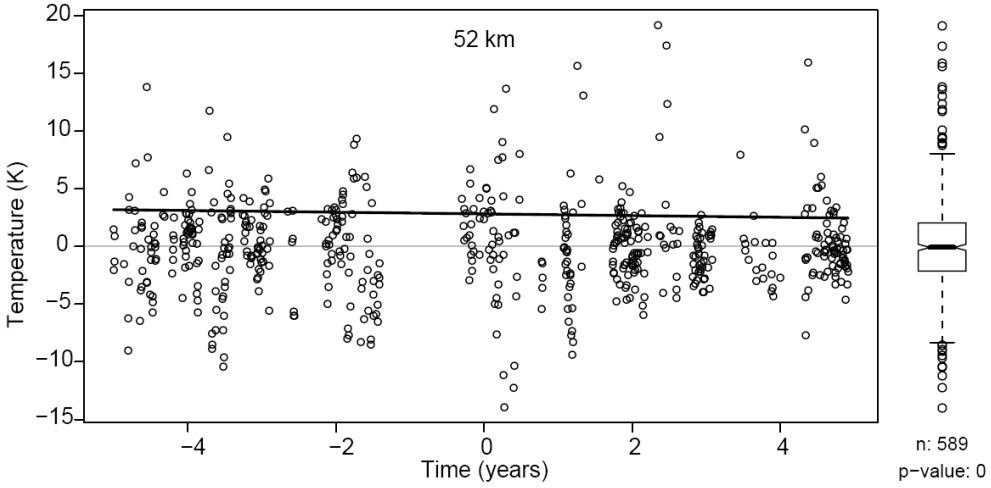

47 km
Temperature (K)
Time (years)
n: 586
p−value: 0

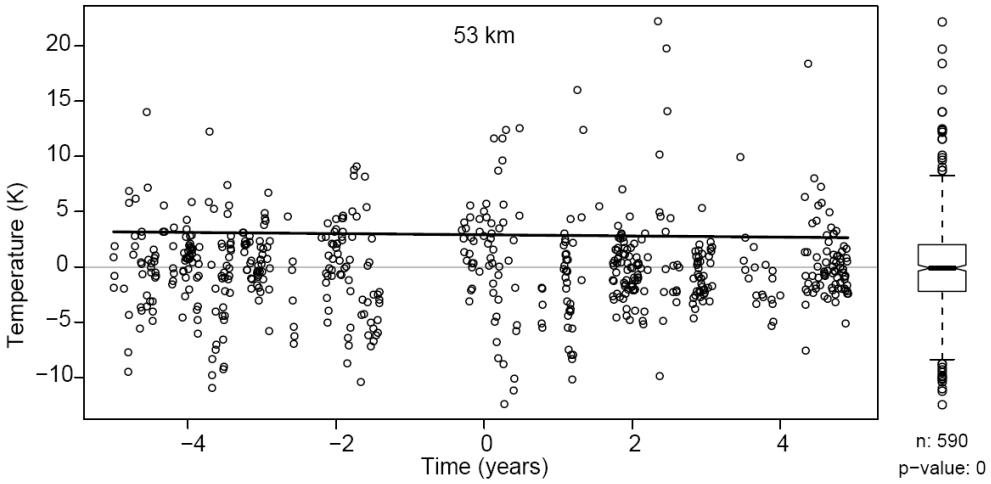

48 km
Temperature (K)
Time (years)
n: 586
p−value: 0

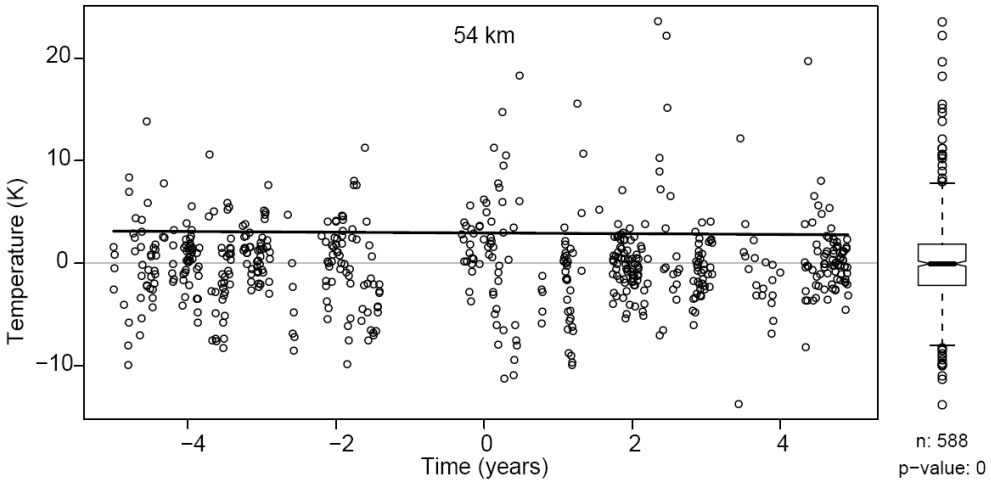

49 km
Temperature (K)
Time (years)
n: 586
p−value: 0

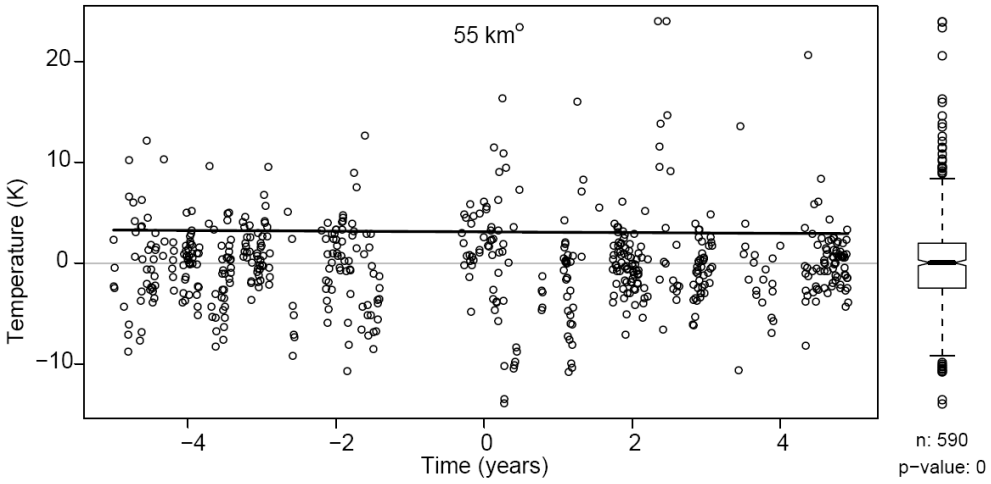

50 km
Temperature (K)
Time (years)
n: 587
p−value: 0

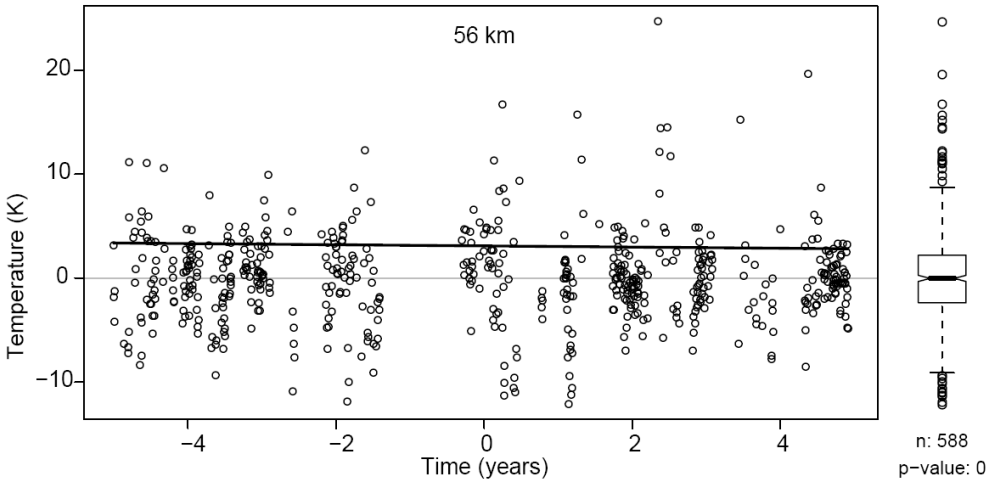

51 km
Temperature (K)
Time (years)
n: 588
p−value: 0

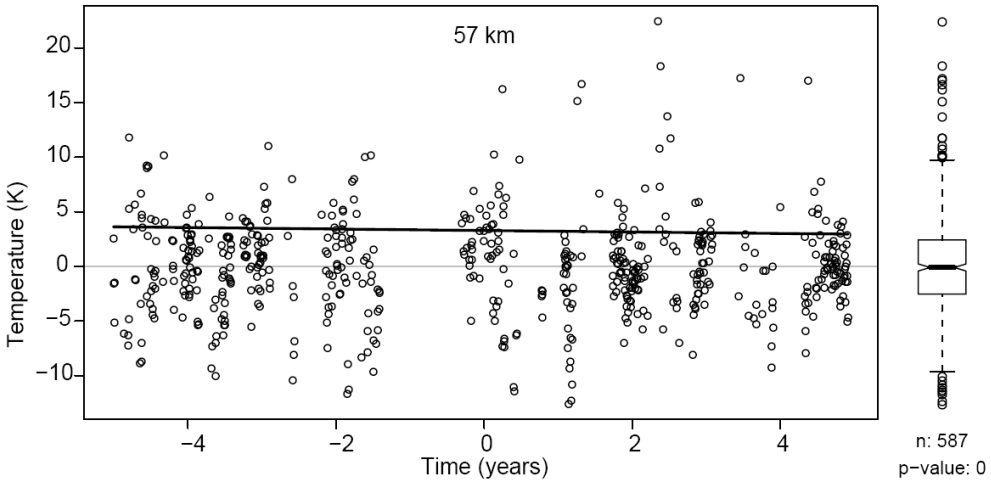

52 km
Temperature (K)
Time (years)
n: 589
p−value: 0

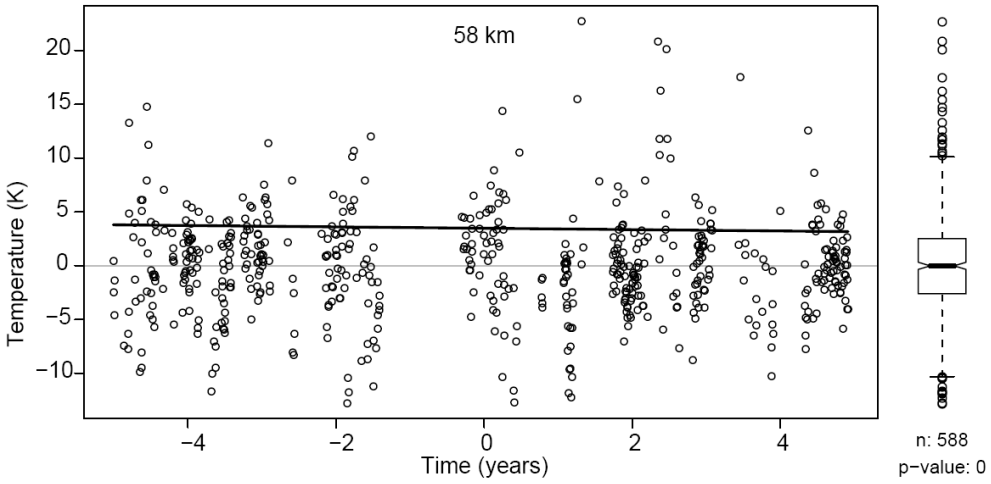

53 km
Temperature (K)
Time (years)
n: 590
p−value: 0

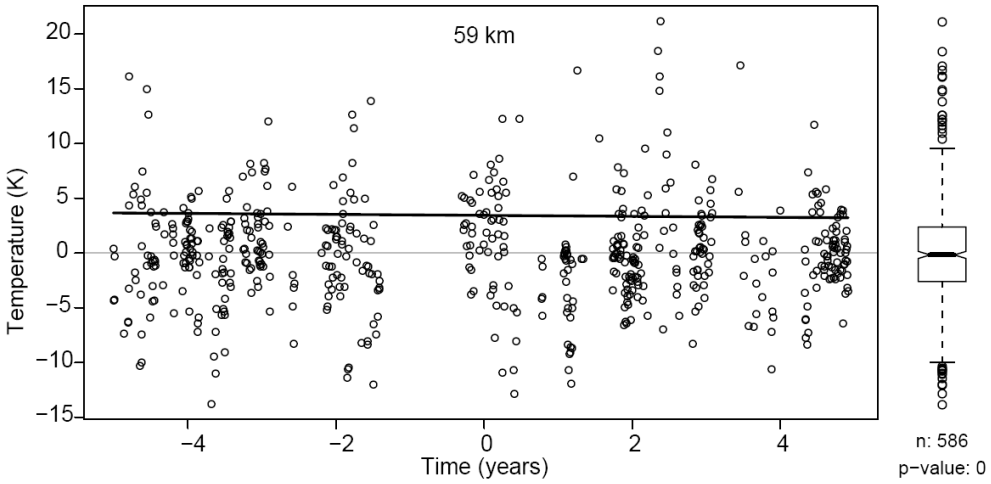

54 km
Temperature (K)
Time (years)
n: 588
p−value: 0

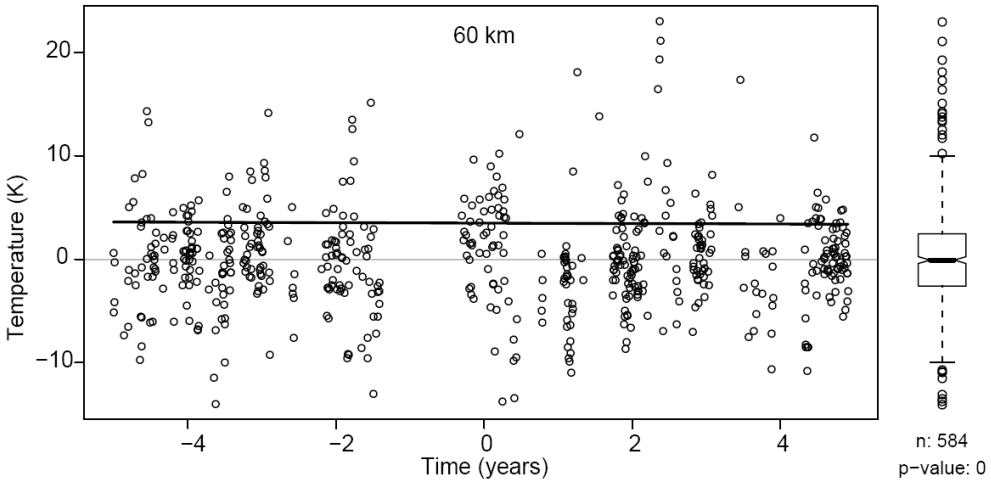

55 km
Temperature (K)
Time (years)
n: 590
p−value: 0

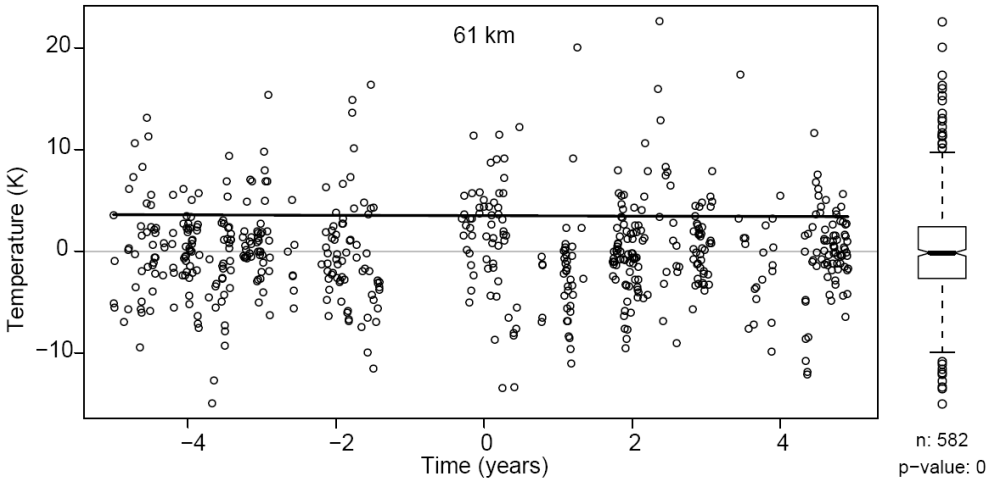

56 km
Temperature (K)
Time (years)
n: 588
p−value: 0

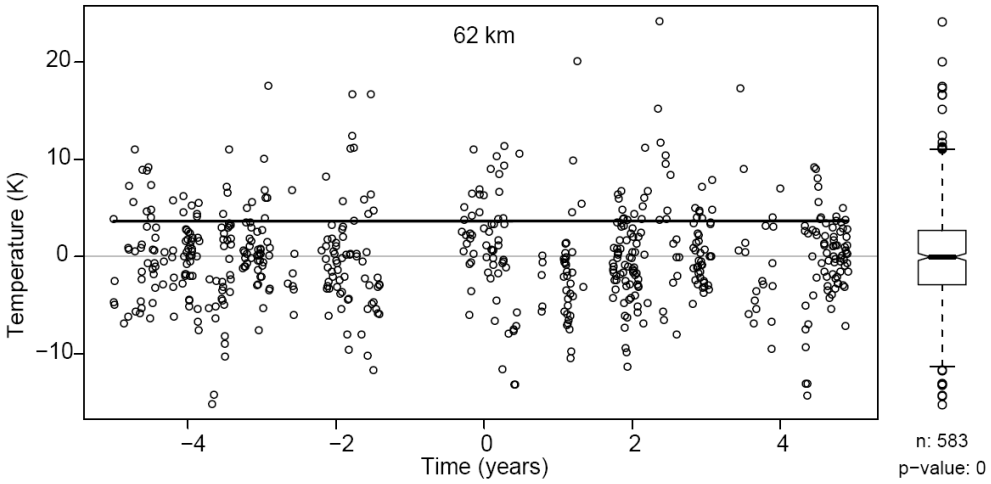

57 km
Temperature (K)
Time (years)
n: 587
p−value: 0

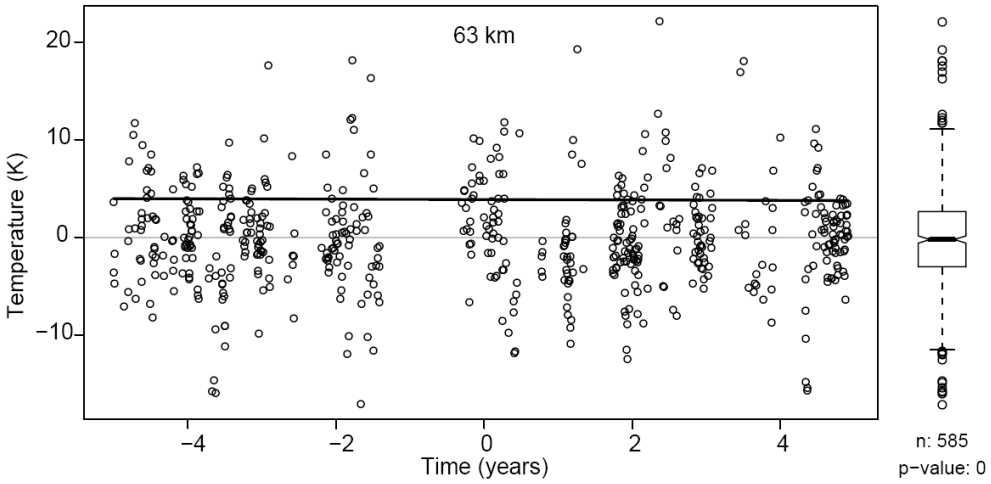

58 km
Temperature (K)
Time (years)
n: 588
p−value: 0

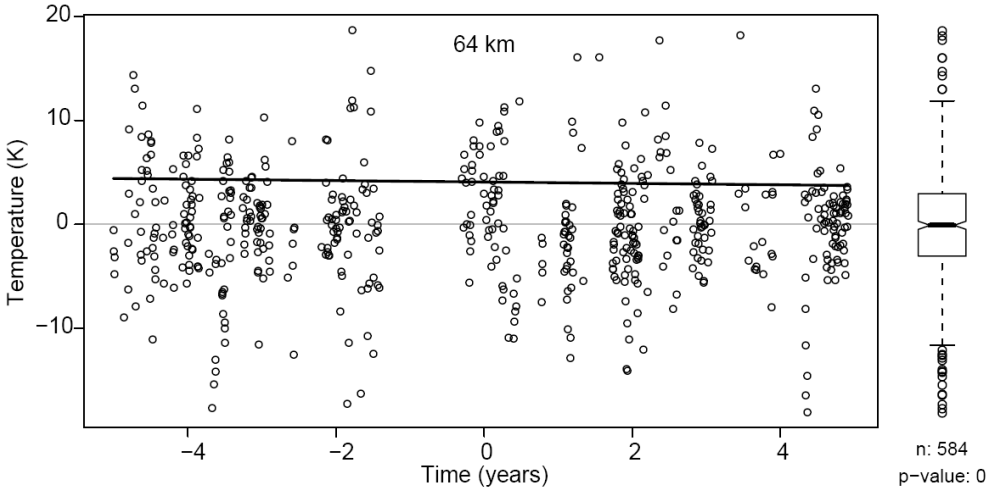

59 km
Temperature (K)
Time (years)
n: 586
p−value: 0

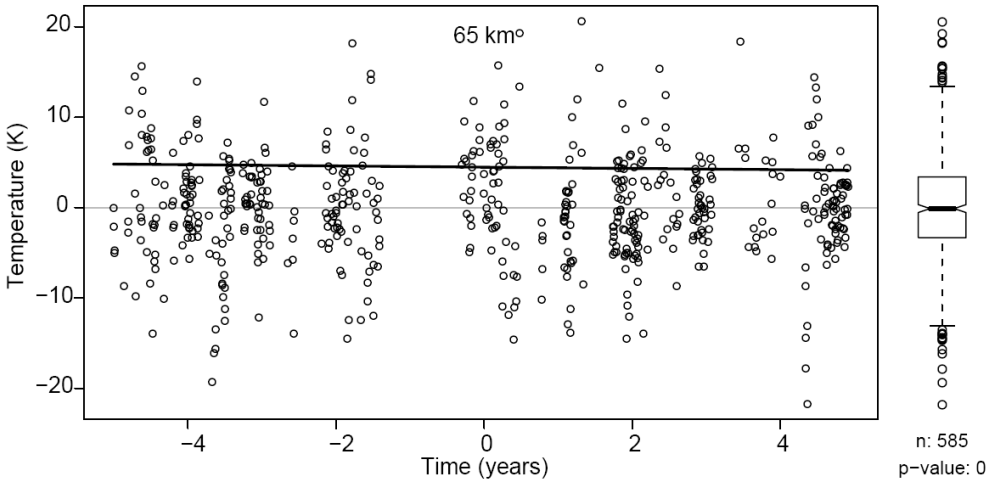

60 km
Temperature (K)
20
10
0
−10
−4
−2
0
2
4
Time (years)
n: 584
p−value: 0

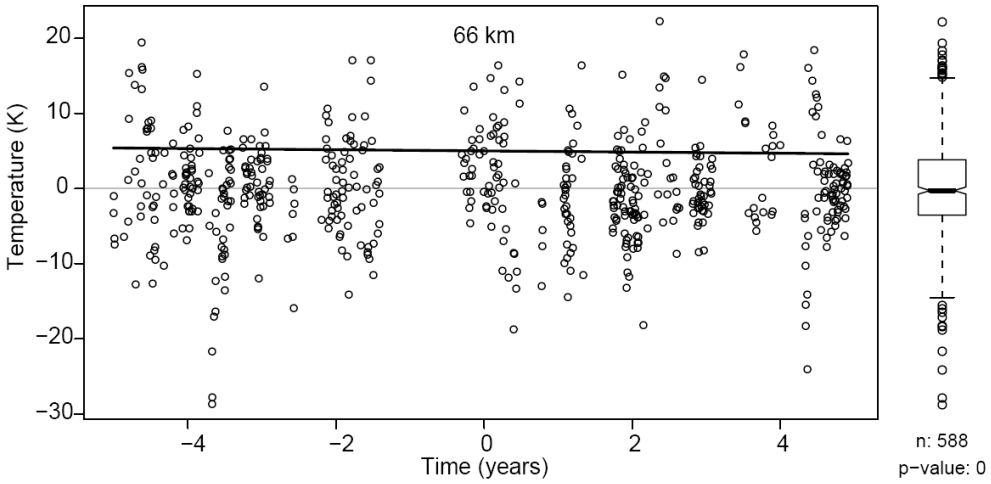

61 km
Temperature (K)
20
10
0
−10
−4
−2
0
2
4
Time (years)
n: 582
p−value: 0

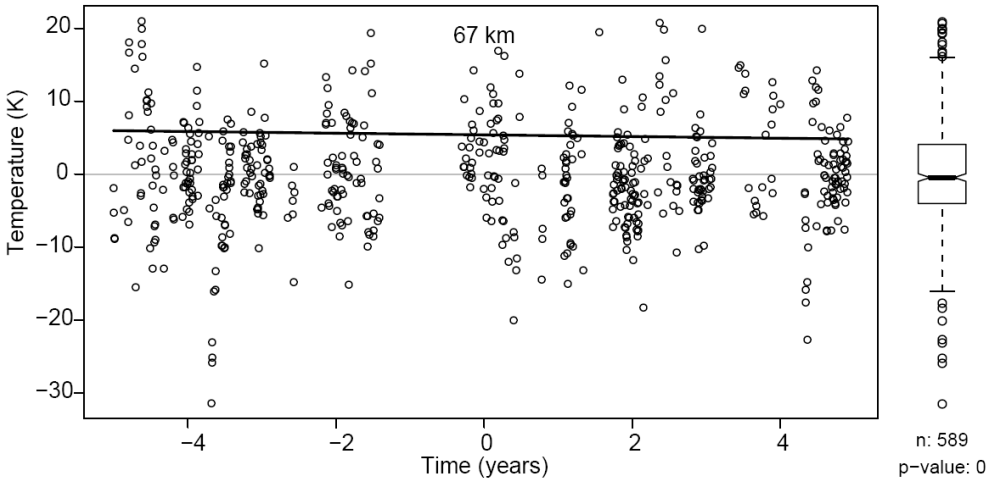

62 km
Temperature (K)
20
10
0
−10
−4
−2
0
2
4
Time (years)
n: 583
p−value: 0

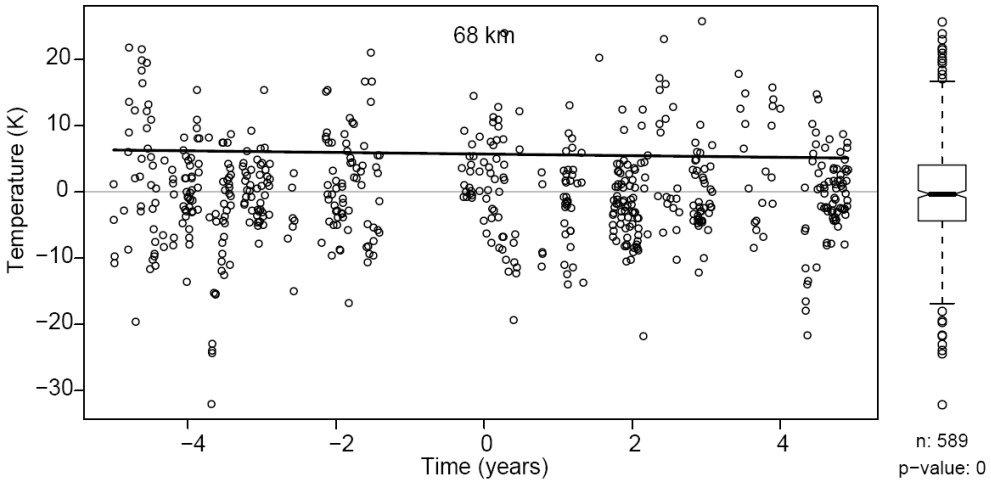

63 km
Temperature (K)
20
10
0
−10
−4
−2
0
2
4
Time (years)
n: 585
p−value: 0

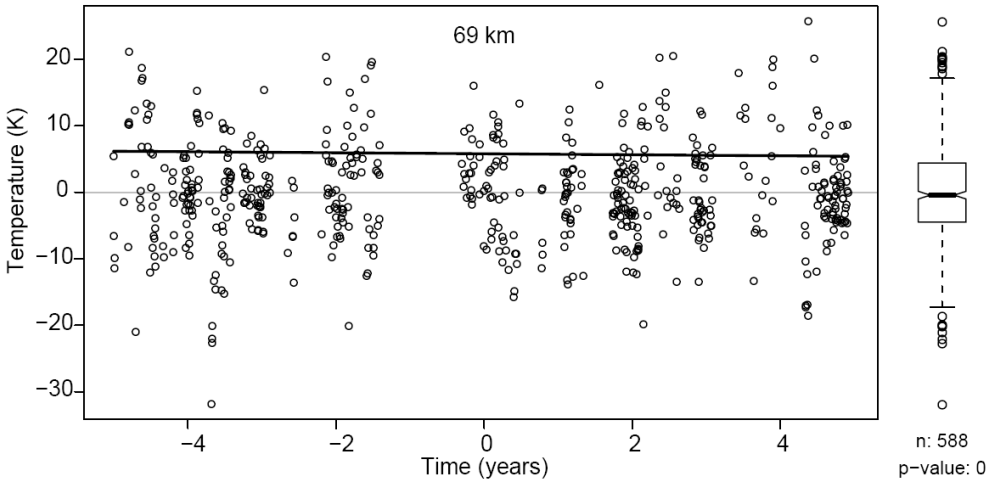

64 km
Temperature (K)
20
10
0
−10
−4
−2
0
2
4
Time (years)
n: 584
p−value: 0

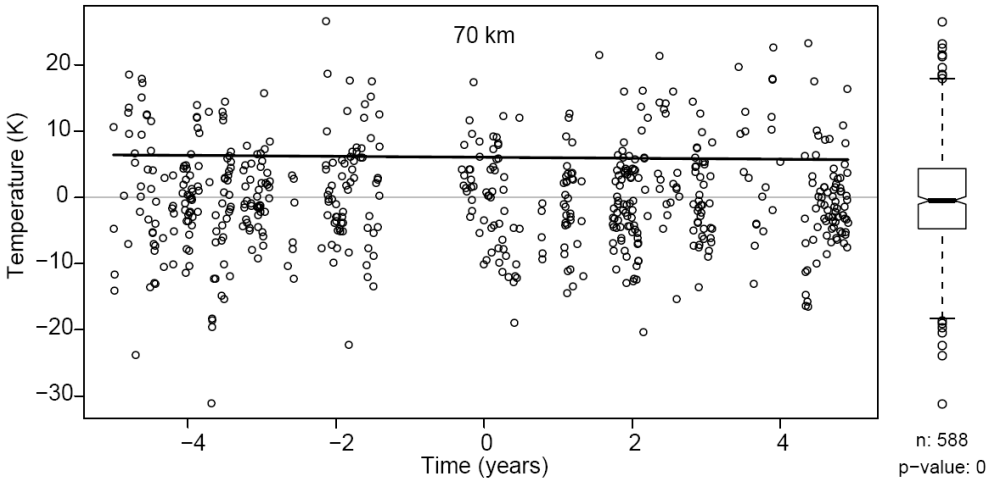

65 km
Temperature (K)
20
10
0
−10
−20
−4
−2
0
2
4
Time (years)
n: 585
p−value: 0

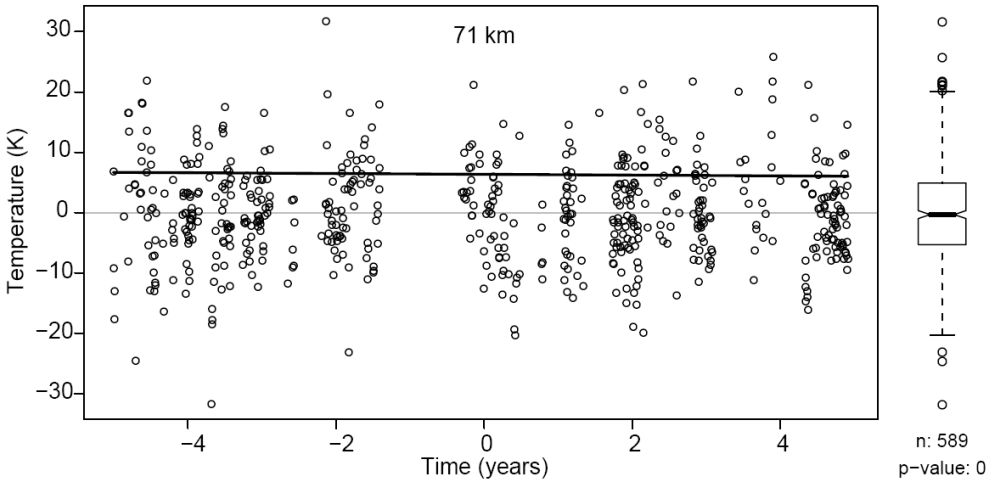

66 km
Temperature (K)
Time (years)
n: 588
p−value: 0

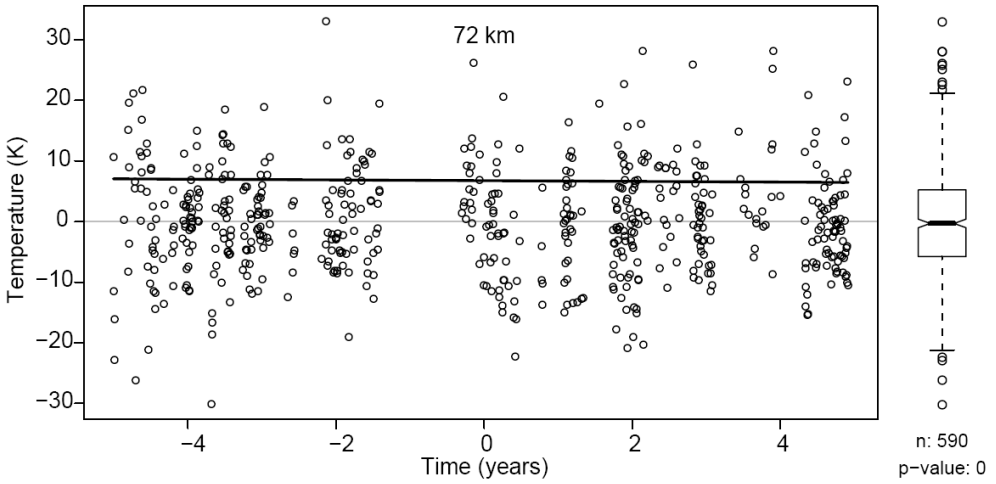

67 km
Temperature (K)
Time (years)
n: 589
p−value: 0

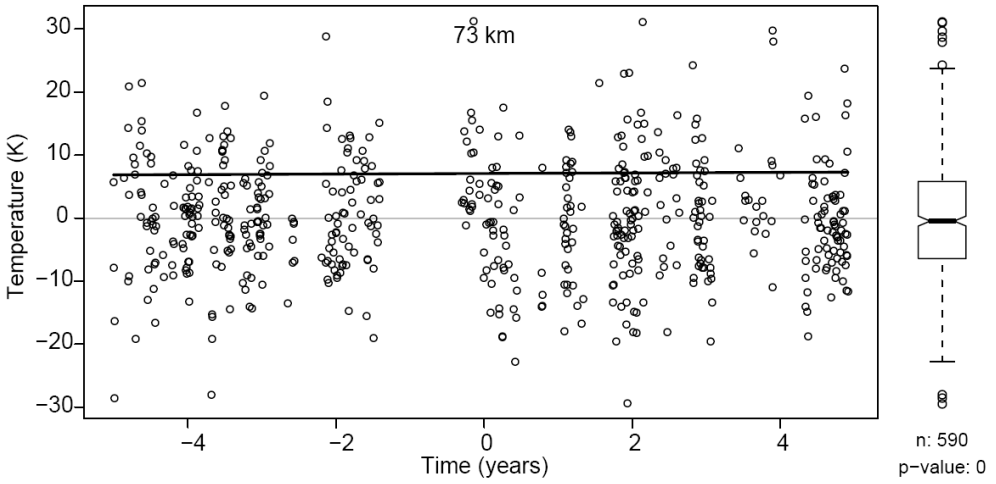

68 km
Temperature (K)
Time (years)
n: 589
p−value: 0

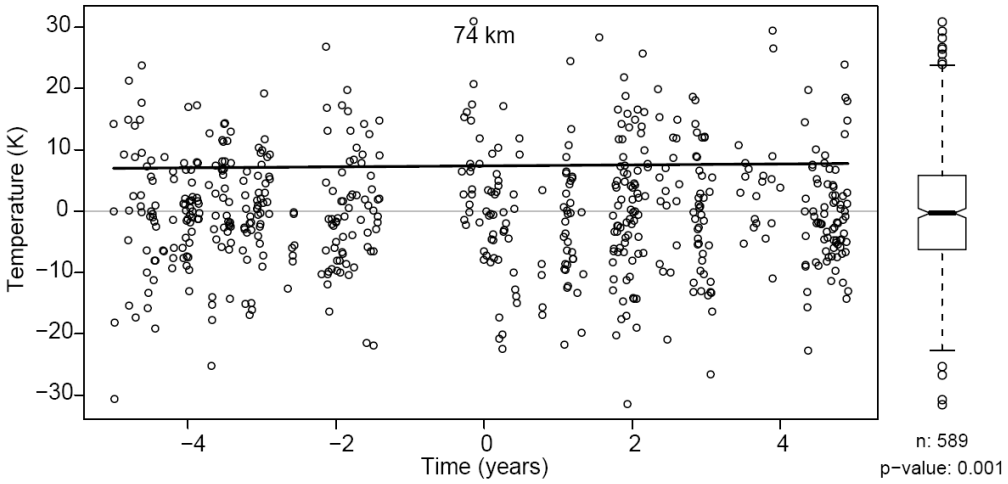

69 km
Temperature (K)
20
10
0
−10
−20
−30
−4
−2
0
2
4
Time (years)
n: 588
p−value: 0

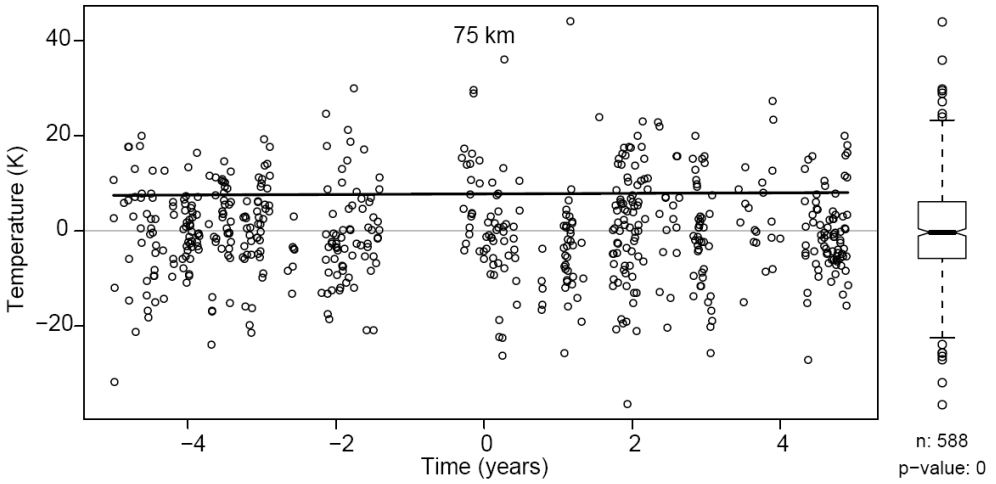

70 km
Temperature (K)
20
10
0
−10
−20
−30
−4
−2
0
2
4
Time (years)
n: 588
p−value: 0

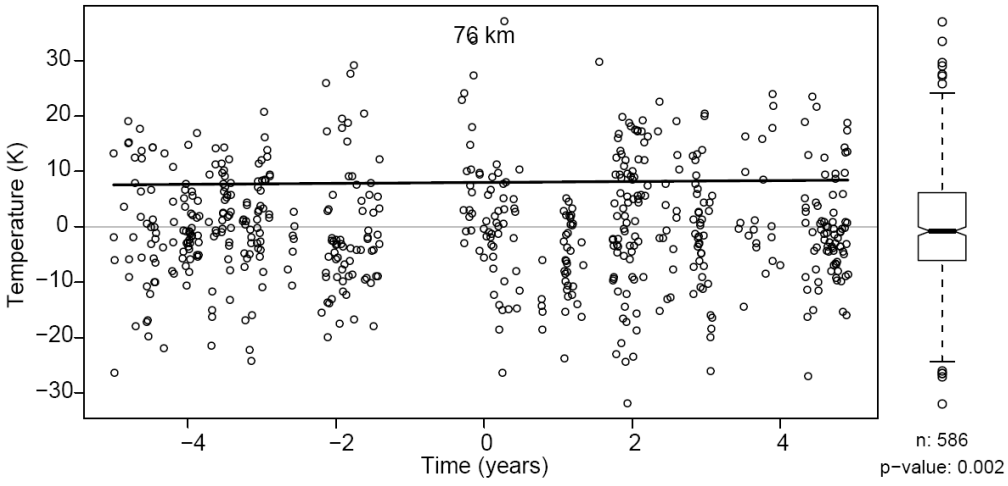

71 km
Temperature (K)
30
20
10
0
−10
−20
−30
−4
−2
0
2
4
Time (years)
n: 589
p−value: 0

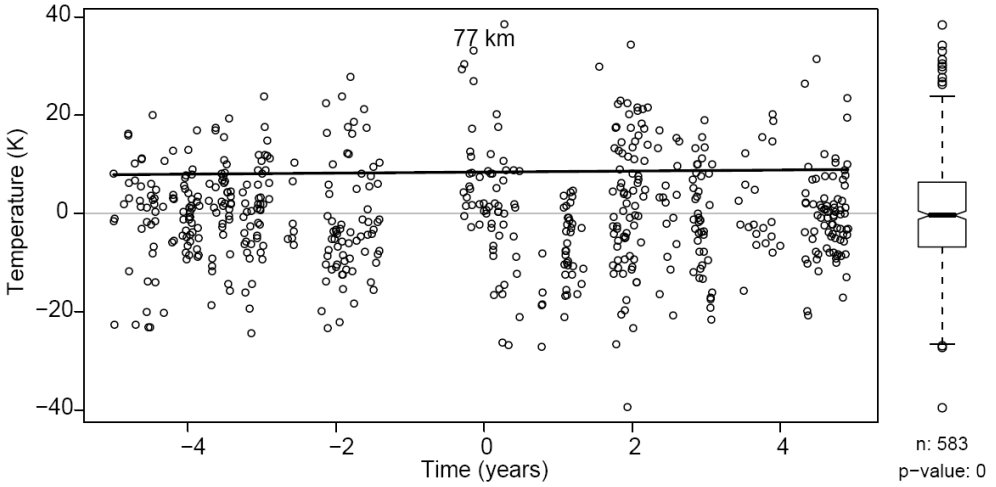

72 km
Temperature (K)
Time (years)
n: 590
p−value: 0

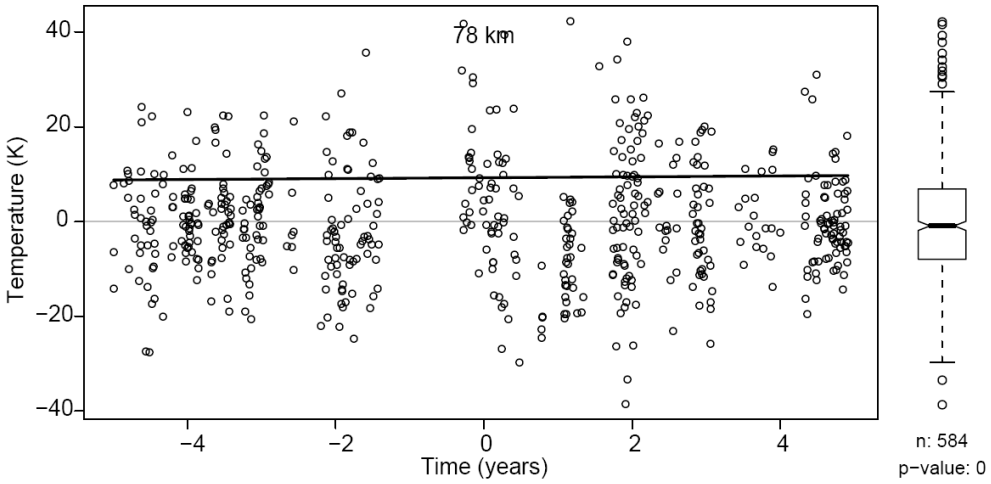

73 km
Temperature (K)
Time (years)
n: 590
p−value: 0

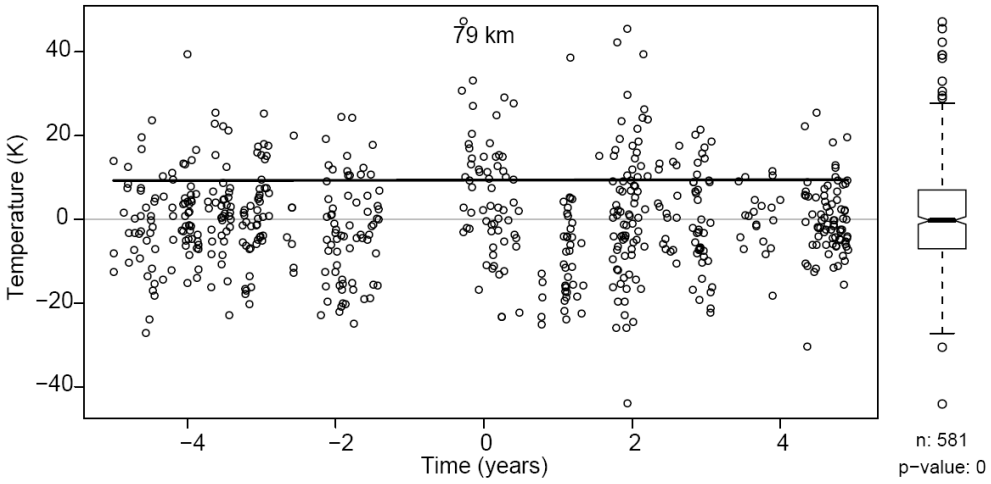

74 km
Temperature (K)
Time (years)
n: 589
p−value: 0.001

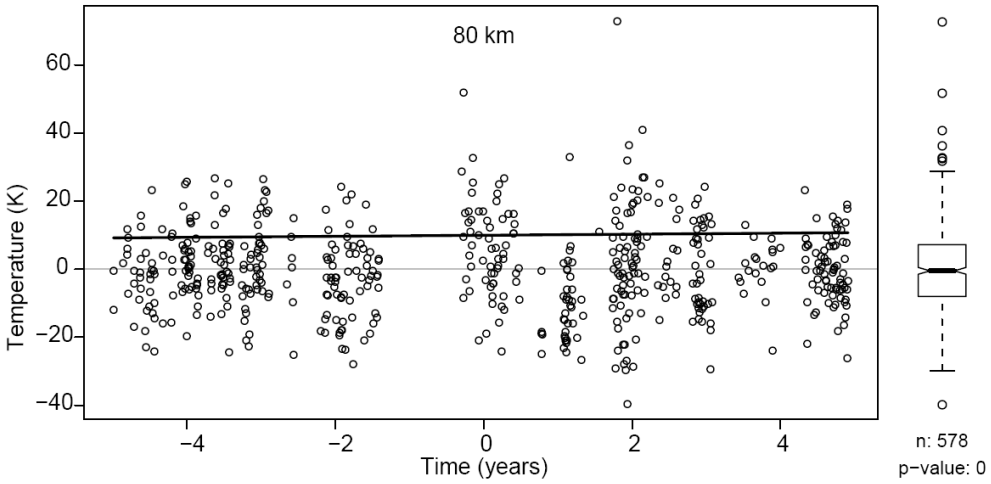

75 km
Temperature (K)
Time (years)
n: 588
p−value: 0

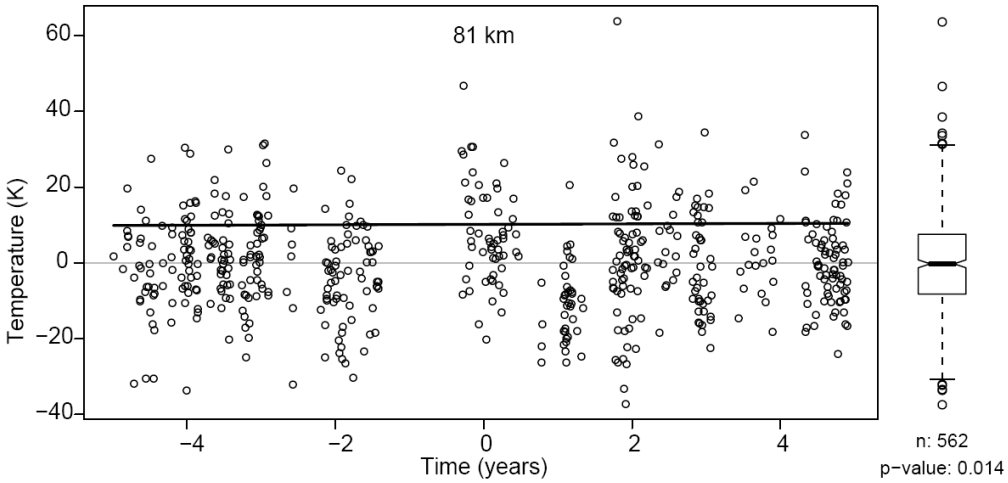

76 km
Temperature (K)
Time (years)
n: 586
p−value: 0.002

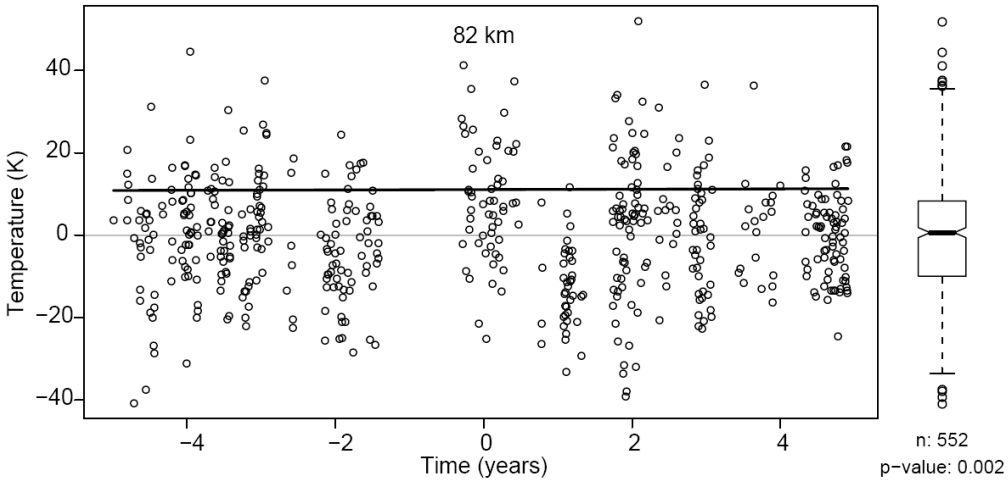

77 km
Temperature (K)
Time (years)
n: 583
p−value: 0

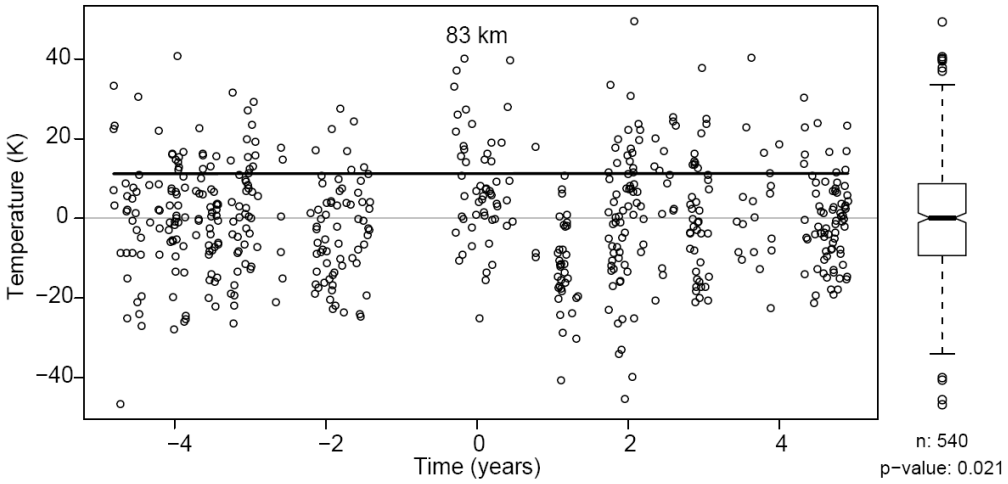

78 km
Temperature (K)
Time (years)
n: 584
p−value: 0

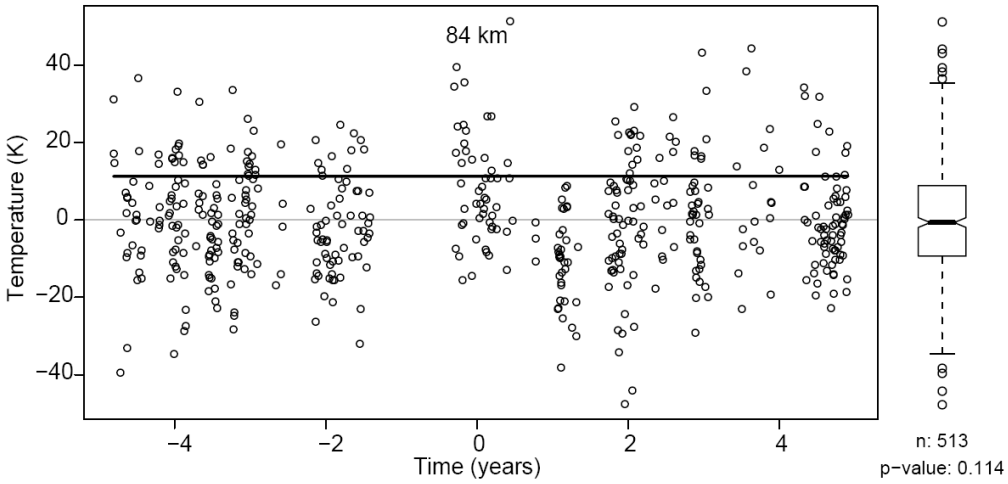

79 km
Temperature (K)
Time (years)
n: 581
p−value: 0

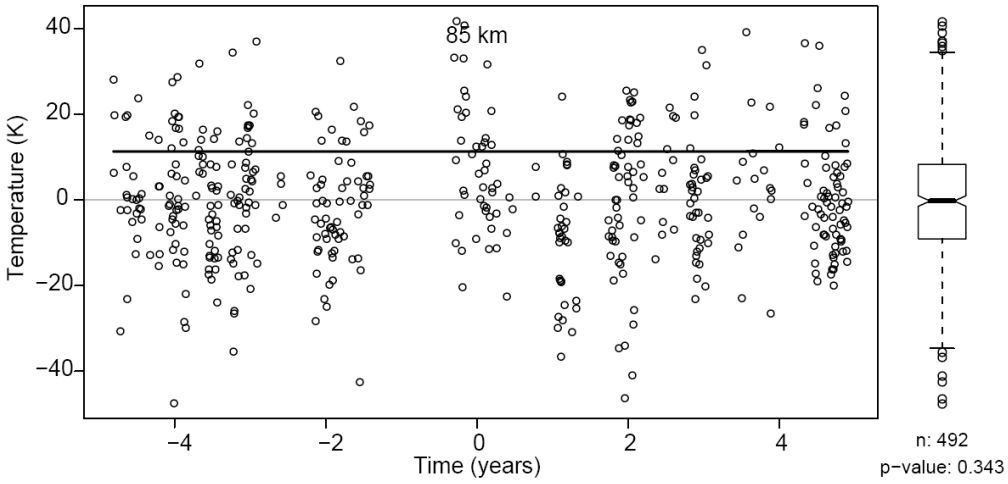

80 km
Temperature (K)
Time (years)
n: 578
p−value: 0

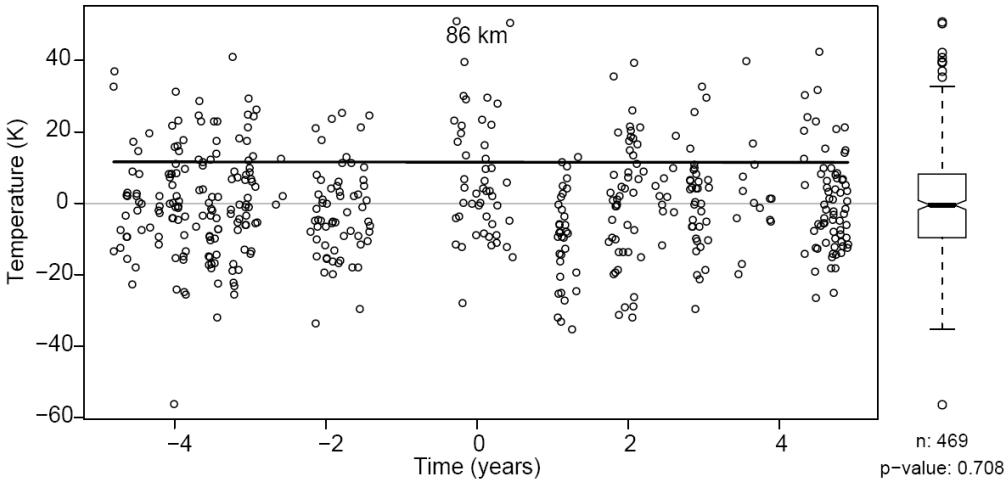

81 km
Temperature (K)
Time (years)
n: 562
p−value: 0.014

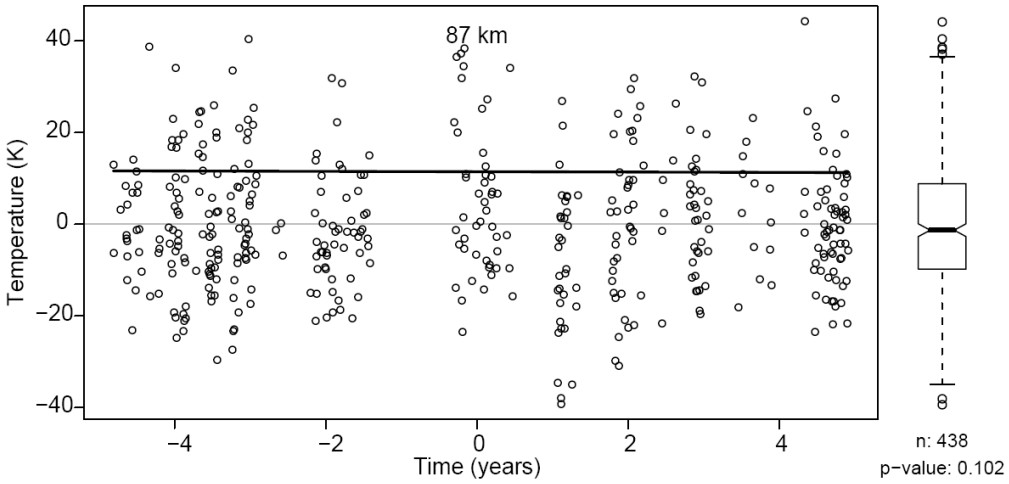

82 km
Temperature (K)
Time (years)
n: 552
p−value: 0.002

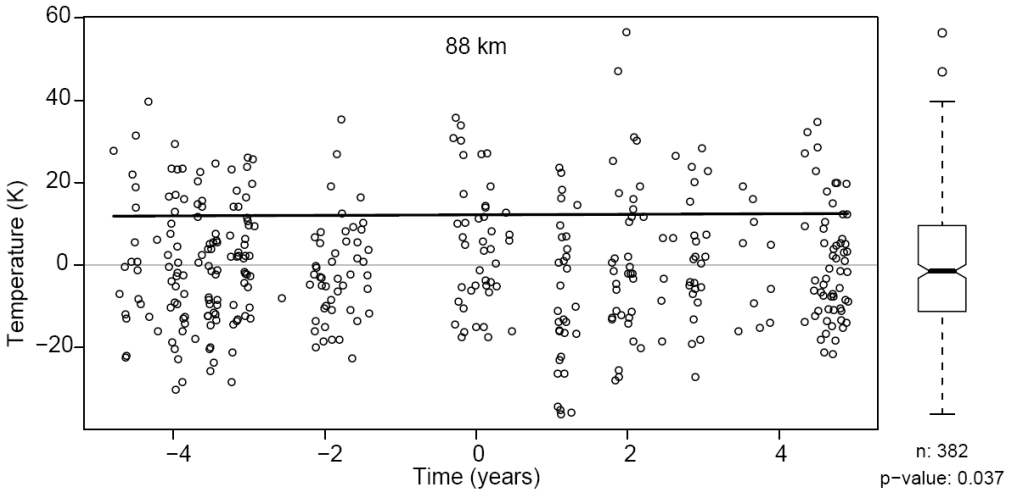

83 km
Temperature (K)
Time (years)
n: 540
p−value: 0.021

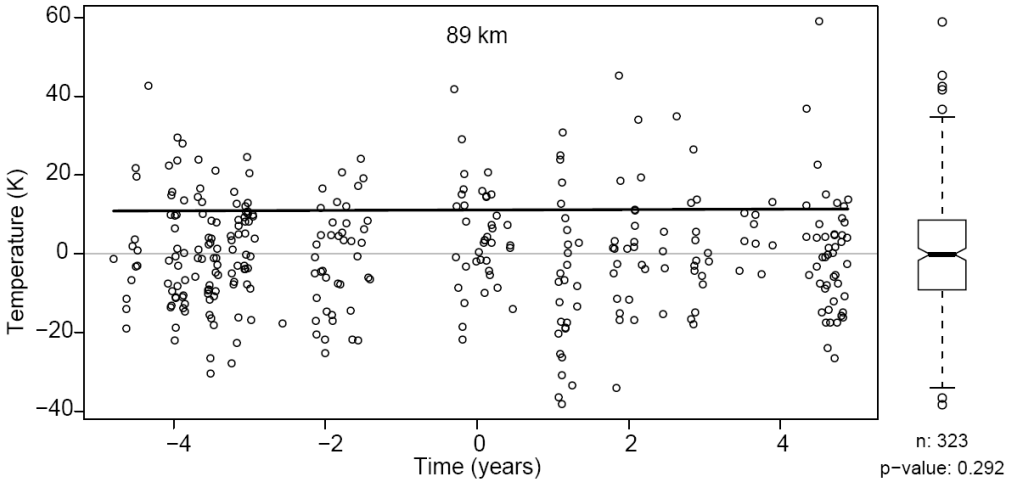

84 km
Temperature (K)
40
20
0
−20
−40
−4
−2
0
2
4
Time (years)
n: 513
p−value: 0.114

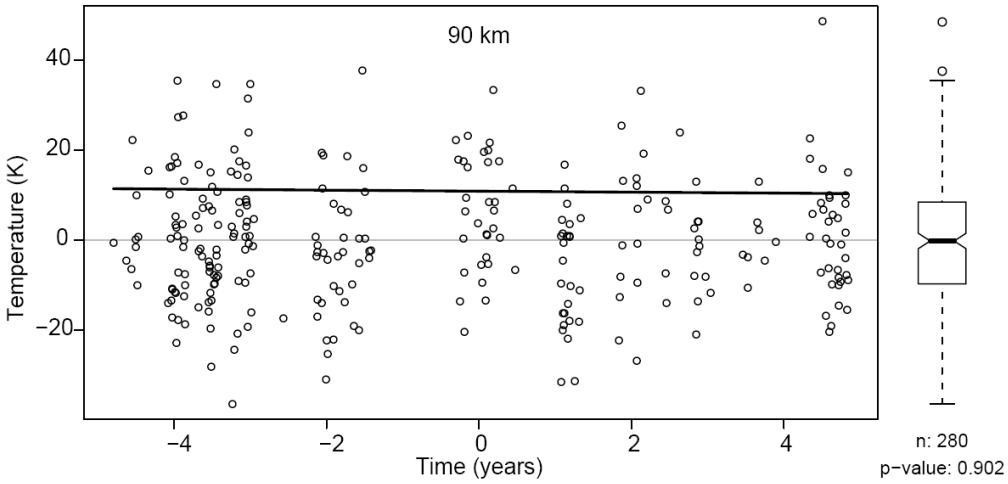

85 km
Temperature (K)
40
20
0
−20
−40
−4
−2
0
2
4
Time (years)
n: 492
p−value: 0.343


86 km
Temperature (K)
40
20
0
−20
−40
−60
−4
−2
0
2
4
Time (years)
n: 469
p−value: 0.708


87 km
Temperature (K)
40
20
0
−20
−40
−4
−2
0
2
4
Time (years)
n: 438
p−value: 0.102


88 km
Temperature (K)
60
40
20
0
−20
−4
−2
0
2
4
Time (years)
n: 382
p−value: 0.037


89 km
Temperature (K)
60
40
20
0
−20
−40
−4
−2
0
2
4
Time (years)
n: 323
p−value: 0.292


90 km
Temperature (K)
40
20
0
−20
−4
−2
0
2
4
Time (years)
n: 280
p−value: 0.902

APPENDIX G

CALCULATION OF THE INNER PRODUCT TERMS

FOR THE BIAS AND ATTENUATION

Abstract. In this section of the appendices the inner product terms from Chapter 2 are derived.

1. The Inner Products

We have the following inner product terms to calculate from the following two equations.

$$\alpha' = \alpha''\{ (t^T t)(SP^T SR) - (SP^T t)(SR^T t)\} / \gamma$$

$$\beta_1' = \beta_1'' + \alpha''\{ (SP^T SP)(SR^T t) - (SP^T t)(SP^T SR)\} / \gamma,$$

where $\gamma = (t^T t)(SP^T SP) - (SP^T t)^2$. Let

$$SP = sin(\omega t + \theta),\ SR = sin(\omega t + \varphi).$$

The summation may be approximated as an integration. Therefore, the inner products $(t^T t)$, $(SP^T SP)$ and $(SP^T t)$ may be approximated with integrals. That is $\sum f(x_i)^2 \Delta t \approx \int f(x)^2 dt$, or $\sum f(x_i)^2 \approx \int f(x)^2 dt/\Delta t$. So, for example, $t^T t \approx$ $= 1/3t_0^3$. Let $s1 = sin\ \omega t$, $c1 = cos\ \omega t$, $s2 = sin\ 2\omega t$, $c2 = cos\ 2\omega t$. The vector t, SP, and SR are mean centered so the average must be subtracted and we will need

_____ ,

_____ .

The inner product terms are evaluated below.

$(SP^{T}SP) = \sum SP_{i}SP_{i} =$ ~

$(SP^{T}SP) = \{ t_0 - s_2\cos(2\theta)/2\omega - 2s_1^{2}\sin^{2}\theta/t_0\omega^{2} \} / \Delta t.$

$(t^{T}t) = \sum t_i t_i =$

$(t^{T}t) = 2t_0^{3}/(3\cdot\Delta t).$

$(SR^{T}t) = \sum(sin(\omega t_i + \varphi) - S_{\varphi})t_i =$

$(SR^{T}t) = \{2\cdot\cos(\varphi)\{ s_1 - \omega t_0 c_1\} / \omega^{2}\}/\Delta t.$

$(SP^{T}t) = \sum(sin(\omega t_i + \theta) - S_{\theta})t_i =$

$(SP^{T}t) = (2\cdot\cos(\theta)\{ s_1 - \omega t_0 c_1\} / \omega^{2})/\Delta t.$

$(SP^{T}SR) =$

$\sum(sin(\omega t_i + \theta) - S_{\theta}) (sin(\omega t_i + \varphi) - S_{\varphi}) =$

$(SP^{T}SR) = (t_0\cos(\varphi - \theta) - s_2\cos(\varphi + \theta)/2\omega - 2s_1^{2}\sin(\varphi)\sin(\theta)/t_0\omega^{2}) /\Delta t.$

The expected values of the coefficients α' are β_1'

$E(A') = \alpha'' \{ [2t_0^{3}/3][t_0\cos(\varphi - \theta) - s_2\cos(\varphi + \theta)/2\omega - 2s_1^{2}\sin(\varphi)\sin(\theta)/t_0\omega^{2}] -$

$[2\cdot\cos(\theta)(s_1 - \omega t_0 c_1) / \omega^{2}][2\cdot\cos(\varphi)(s_1 - \omega t_0 c_1) / \omega^{2}]\} / \gamma.$

$E(A') = \beta_1'' + \alpha''\{ [t_0 - s_2\cos(2\theta)/2\omega - 2s_1^{2}\sin^{2}\theta/t_0\omega^{2}][2\cdot\cos(\varphi)(s_1 - \omega t_0 c_1) / \omega^{2}] -$

$[2\cdot\cos(\theta)(s_1 - \omega t_0 c_1) /\omega^{2}][t_0\cos(\varphi - \theta) - s_2\cos(\varphi + \theta)/2\omega - 2s_1^{2}\sin(\varphi)\sin(\theta)/t_0\omega^{2}] \} / \gamma,$

where $\gamma = (t^T t)(SP^T SP) - (SP^T t)^2 = (2t_0^3/3)(t_0 - s_2 cos(2\theta)/2\omega - 2s_1^2 sin^2\theta/t_0\omega^2) - (2 \cdot cos(\theta)\{ s_1 - \omega t_0 c_1\} / \omega^2)^2$.

2. Special Cases

If the solar phase angle is $\theta = 0$, then

$$\beta' = \beta''$$

$$\alpha' = \alpha'' \cdot cos(\varphi) \{ 2/3 \cdot t_0^3 (t_0 - s_2/2\omega) - 4(s_1 - \omega t_0 c_1)^2 /\omega^4) \} / \gamma .$$

If the solar phase angle is $\theta = \pi/2$, then

$$\beta' = \beta'' + 2\alpha'' \cdot cos(\varphi) \{ (t_0 + s_2/2\omega - 2s_1^2/t_0\omega^2)(s_1 - \omega t_0 c_1) \} / \gamma\omega^2$$

$$\alpha' = 1.5 \cdot \alpha'' \cdot sin(\varphi) \cdot t_0^3 (t_0 + s_2/2\omega - 2s_1^2/t_0\omega^2) / \gamma.$$

APPENDIX H

INFLATION OF STANDARD ERRORS

1. Inflation of Variance

Autocorrelation inflates the variance of a time series of data. Shown in this appendix is a derivation of how this occurs.

A first order autocorrelation process may be written as

$$N_k = \varepsilon_k + \varphi\, \varepsilon_{k-1} + \varphi^2\, \varepsilon_{k-2} + \varphi^3\, \varepsilon_{k-3} + \ldots + \varphi^{n-1}\, \varepsilon_{k-n-1} + \varphi^n\, \varepsilon_{k-n}\,, \qquad \text{(H1)}$$

where φ is the autocorrelation coefficient and ε is random noise and k is the time index. Taking the variance of N_k yields

$$\mathrm{VAR}(N_k) = \mathrm{VAR}(\varepsilon_k + \varphi\, \varepsilon_{k-1} + \varphi^2\, \varepsilon_{k-2} + \varphi^3\, \varepsilon_{k-3} + \ldots + \varphi^{n-1}\, \varepsilon_{k-n-1} + \varphi^n\, \varepsilon_{k-n}). \qquad \text{(H2)}$$

Because the εs are independent and identically distributed, and because φ is constant, this becomes.

$$\mathrm{VAR}(N_k) = \mathrm{VAR}(\varepsilon_k) + \varphi^2 VAR(\varepsilon_{k-1}) + \varphi^4 \mathrm{VAR}(\varepsilon_{k-2}) + \varphi^6\, \mathrm{VAR}(\varepsilon_{k-3}) + \ldots + \varphi^{2(n-1)} \mathrm{VAR}(\varepsilon_{k-n-1}) + \varphi^{2n} \mathrm{VAR}(\varepsilon_{k-n}). \qquad \text{(H3)}$$

Because all the εs have the same variance, the variance of the εs can be factored out. We then get

$$\sigma^2_N = \sigma^2_\varepsilon(1 + \varphi^2 + \varphi^4 + \ldots + \varphi^{2(n-1)} + \varphi^{2n}). \qquad \text{(H4)}$$

Applying a geometric series to the φs we get

$$\sigma^2_N = \sigma^2_\varepsilon/(1 - \varphi^2). \quad \text{(H5)}$$

The standard deviation of the Ns is increased by the factor $(1 + \varphi^2)^{-1}$.

APPENDIX I

COMPTUER CODE

```
#########################################################################
######### THIS CODE GENERATES FIGURE 5 FROM CHAPTER 2 ###########
#########################################################################

m(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)

source("dir.txt")
ppppp <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Data\\fit solar\\F107_A.txt"


T <- matrix(NA, 380, 2)

T[1,] <-  c( 1968.95094,206.18091)
T[2,] <-  c( 1969,217.96483)
T[3,] <-  c( 1969.04906,214.97487)
T[4,] <-  c( 1969.09811,199.84924)
T[5,] <-  c( 1969.14717,238.36684)
T[6,] <-  c( 1969.14717,226.23116)
T[7,] <-  c( 1969.24529,234.84924)
T[8,] <-  c( 1969.24529,223.59297)
T[9,] <-  c( 1969.29434,234.14572)
T[10,] <-  c( 1969.3434,222.71356)
T[11,] <-  c( 1969.39246,210.92964)
T[12,] <-  c( 1969.44151,214.27136)
T[13,] <-  c( 1969.49056,224.82413)
T[14,] <-  c( 1969.58868,225.87939)
T[15,] <-  c( 1969.58868,214.09547)
T[16,] <-  c( 1969.68679,232.56282)
T[17,] <-  c( 1969.73585,223.94472)
T[18,] <-  c( 1969.73585,223.94472)
T[19,] <-  c( 1969.73585,221.13065)
T[20,] <-  c( 1969.7849,227.63818)
T[21,] <-  c( 1969.7849,235.72864)
T[22,] <-  c( 1969.88302,225)
T[23,] <-  c( 1969.88302,215.85428)
T[24,] <-  c( 1969.98113,233.09045)
T[25,] <-  c( 1969.98113,238.01508)
T[26,] <-  c( 1970.03019,219.19598)
T[27,] <-  c( 1970.07925,228.34171)
T[28,] <-  c( 1970.07925,242.58794)
T[29,] <-  c( 1970.17736,233.96985)
T[30,] <-  c( 1970.17736,222.88945)
T[31,] <-  c( 1970.27547,251.03015)
T[32,] <-  c( 1970.27547,260.70352)
T[33,] <-  c( 1970.32453,218.49246)
T[34,] <-  c( 1970.32453,207.58794)
T[35,] <-  c( 1970.47169,226.58292)
T[36,] <-  c( 1970.52075,214.62312)
T[37,] <-  c( 1970.61887,222.88945)
T[38,] <-  c( 1970.61887,211.45729)
T[39,] <-  c( 1970.66792,235.02513)
T[40,] <-  c( 1970.71698,223.59297)
T[41,] <-  c( 1970.76604,249.62312)
T[42,] <-  c( 1970.76604,238.7186)
```

```
T[43,] <-  c( 1970.91321,241.18091)
T[44,] <-  c( 1970.91321,232.91457)
T[45,] <-  c( 1970.91321,215.50252)
T[46,] <-  c( 1970.96227,254.37186)
T[47,] <-  c( 1970.96227,243.64322)
T[48,] <-  c( 1971.01132,237.48744)
T[49,] <-  c( 1971.01132,232.91457)
T[50,] <-  c( 1971.06038,220.60301)
T[51,] <-  c( 1971.15849,213.39195)
T[52,] <-  c( 1971.15849,211.98492)
T[53,] <-  c( 1971.20755,227.98994)
T[54,] <-  c( 1971.20755,208.64322)
T[55,] <-  c( 1971.20755,218.49246)
T[56,] <-  c( 1971.30566,207.58794)
T[57,] <-  c( 1971.40377,205.65326)
T[58,] <-  c( 1971.40377,216.38191)
T[59,] <-  c( 1971.45283,207.9397)
T[60,] <-  c( 1971.50188,211.98492)
T[61,] <-  c( 1971.50188,219.0201)
T[62,] <-  c( 1971.6,222.88945)
T[63,] <-  c( 1971.64906,221.65829)
T[64,] <-  c( 1971.64906,232.73869)
T[65,] <-  c( 1971.74717,230.9799)
T[66,] <-  c( 1971.79623,219.0201)
T[67,] <-  c( 1971.84528,217.96483)
T[68,] <-  c( 1971.84528,230.10051)
T[69,] <-  c( 1971.9434,237.13568)
T[70,] <-  c( 1971.9434,225.70352)
T[71,] <-  c( 1971.99245,215.67839)
T[72,] <-  c( 1971.99245,227.46231)
T[73,] <-  c( 1972.09057,218.1407)
T[74,] <-  c( 1972.09057,229.39699)
T[75,] <-  c( 1972.13963,228.16583)
T[76,] <-  c( 1972.23774,217.43718)
T[77,] <-  c( 1972.23774,215.15076)
T[78,] <-  c( 1972.23774,225.17587)
T[79,] <-  c( 1972.33585,231.68341)
T[80,] <-  c( 1972.33585,220.07538)
T[81,] <-  c( 1972.43396,224.12061)
T[82,] <-  c( 1972.43396,221.30653)
T[83,] <-  c( 1972.48302,212.1608)
T[84,] <-  c( 1972.58113,232.56282)
T[85,] <-  c( 1972.58113,220.7789)
T[86,] <-  c( 1972.67924,226.58292)
T[87,] <-  c( 1972.7283,214.799)
T[88,] <-  c( 1972.77736,210.5779)
T[89,] <-  c( 1972.77736,222.01006)
T[90,] <-  c( 1972.82642,223.41708)
T[91,] <-  c( 1972.97359,210.75377)
T[92,] <-  c( 1972.97359,223.94472)
T[93,] <-  c( 1973.02264,211.80905)
T[94,] <-  c( 1973.0717,223.06532)
T[95,] <-  c( 1973.0717,213.9196)
T[96,] <-  c( 1973.21886,225)
T[97,] <-  c( 1973.26792,215.15076)
T[98,] <-  c( 1973.31698,217.43718)
T[99,] <-  c( 1973.31698,207.23618)
T[100,] <-  c( 1973.36604,204.24623)
T[101,] <-  c( 1973.46415,204.77386)
T[102,] <-  c( 1973.51321,206.70854)
T[103,] <-  c( 1973.51321,219.0201)
T[104,] <-  c( 1973.56226,215.85428)
T[105,] <-  c( 1973.61132,224.64824)
```

```
T[106,] <-  c( 1973.66038,212.51256)
T[107,] <-  c( 1973.75849,235.72864)
T[108,] <-  c( 1973.90566,211.98492)
T[109,] <-  c( 1973.95472,209.87437)
T[110,] <-  c( 1973.95472,223.41708)
T[111,] <-  c( 1974.00378,222.01006)
T[112,] <-  c( 1974.15094,216.55779)
T[113,] <-  c( 1974.15094,227.46231)
T[114,] <-  c( 1974.24905,212.1608)
T[115,] <-  c( 1974.29811,210.40201)
T[116,] <-  c( 1974.29811,221.65829)
T[117,] <-  c( 1974.44528,226.75879)
T[118,] <-  c( 1974.49434,223.06532)
T[119,] <-  c( 1974.59245,214.97487)
T[120,] <-  c( 1974.64151,226.75879)
T[121,] <-  c( 1974.69057,206.70854)
T[122,] <-  c( 1974.69057,218.66833)
T[123,] <-  c( 1974.73962,206.00502)
T[124,] <-  c( 1974.78868,229.2211)
T[125,] <-  c( 1974.83774,217.43718)
T[126,] <-  c( 1974.93585,204.4221)
T[127,] <-  c( 1974.93585,216.55779)
T[128,] <-  c( 1975.03397,212.51256)
T[129,] <-  c( 1975.08302,225.52763)
T[130,] <-  c( 1975.13207,228.16583)
T[131,] <-  c( 1975.18113,210.40201)
T[132,] <-  c( 1975.18113,220.7789)
T[133,] <-  c( 1975.3283,206.00502)
T[134,] <-  c( 1975.37736,222.36182)
T[135,] <-  c( 1975.42641,211.10553)
T[136,] <-  c( 1975.47547,217.08543)
T[137,] <-  c( 1975.52453,230.45226)
T[138,] <-  c( 1975.57359,218.66833)
T[139,] <-  c( 1975.57359,228.34171)
T[140,] <-  c( 1975.72076,204.24623)
T[141,] <-  c( 1975.72076,216.73367)
T[142,] <-  c( 1975.81887,213.74371)
T[143,] <-  c( 1975.81887,228.69347)
T[144,] <-  c( 1975.86793,235.55276)
T[145,] <-  c( 1975.91698,223.06532)
T[146,] <-  c( 1975.91698,225.87939)
T[147,] <-  c( 1976.06415,218.31659)
T[148,] <-  c( 1976.06415,230.27638)
T[149,] <-  c( 1976.16226,212.86432)
T[150,] <-  c( 1976.16226,235.02513)
T[151,] <-  c( 1976.21132,220.7789)
T[152,] <-  c( 1976.26038,207.9397)
T[153,] <-  c( 1976.26038,224.12061)
T[154,] <-  c( 1976.40755,219.19598)
T[155,] <-  c( 1976.4566,208.81909)
T[156,] <-  c( 1976.50566,205.30151)
T[157,] <-  c( 1976.50566,225.52763)
T[158,] <-  c( 1976.60378,213.39195)
T[159,] <-  c( 1976.65283,216.03015)
T[160,] <-  c( 1976.65283,229.04523)
T[161,] <-  c( 1976.70189,218.31659)
T[162,] <-  c( 1976.8,206.18091)
T[163,] <-  c( 1976.8,223.59297)
T[164,] <-  c( 1976.84906,210.40201)
T[165,] <-  c( 1976.84906,226.58292)
T[166,] <-  c( 1976.94717,214.27136)
T[167,] <-  c( 1976.99622,217.43718)
T[168,] <-  c( 1976.99622,228.51759)
```

```
T[169,] <-  c( 1977.04528,222.36182)
T[170,] <-  c( 1977.09434,211.45729)
T[171,] <-  c( 1977.14339,223.59297)
T[172,] <-  c( 1977.14339,227.63818)
T[173,] <-  c( 1977.19245,213.39195)
T[174,] <-  c( 1977.19245,216.73367)
T[175,] <-  c( 1977.38868,213.74371)
T[176,] <-  c( 1977.38868,226.58292)
T[177,] <-  c( 1977.43774,215.15076)
T[178,] <-  c( 1977.43774,228.34171)
T[179,] <-  c( 1977.53585,221.30653)
T[180,] <-  c( 1977.53585,233.09045)
T[181,] <-  c( 1977.63396,212.51256)
T[182,] <-  c( 1977.63396,224.29648)
T[183,] <-  c( 1977.73208,216.90955)
T[184,] <-  c( 1977.73208,228.16583)
T[185,] <-  c( 1977.83018,211.2814)
T[186,] <-  c( 1977.83018,223.06532)
T[187,] <-  c( 1977.83018,235.201)
T[188,] <-  c( 1977.9283,225.35176)
T[189,] <-  c( 1978.07547,224.64824)
T[190,] <-  c( 1978.07547,231.50754)
T[191,] <-  c( 1978.22264,218.49246)
T[192,] <-  c( 1978.32076,222.88945)
T[193,] <-  c( 1978.32076,230.45226)
T[194,] <-  c( 1978.41887,238.19095)
T[195,] <-  c( 1978.56604,234.67337)
T[196,] <-  c( 1978.66415,219.37186)
T[197,] <-  c( 1978.71321,221.30653)
T[198,] <-  c( 1978.76226,230.10051)
T[199,] <-  c( 1978.86037,231.33167)
T[200,] <-  c( 1979.00755,225.35176)
T[201,] <-  c( 1979.15472,223.24121)
T[202,] <-  c( 1979.20377,219.54774)
T[203,] <-  c( 1979.4,225.35176)
T[204,] <-  c( 1979.44906,232.21106)
T[205,] <-  c( 1979.44906,232.21106)
T[206,] <-  c( 1979.54717,244.34674)
T[207,] <-  c( 1979.64529,241.35678)
T[208,] <-  c( 1979.69434,224.64824)
T[209,] <-  c( 1979.74339,221.83417)
T[210,] <-  c( 1979.84151,228.86935)
T[211,] <-  c( 1979.93962,230.80402)
T[212,] <-  c( 1979.98868,227.46231)
T[213,] <-  c( 1980.13585,226.58292)
T[214,] <-  c( 1980.23396,221.65829)
T[215,] <-  c( 1980.38113,216.90955)
T[216,] <-  c( 1980.47925,229.39699)
T[217,] <-  c( 1980.47925,224.82413)
T[218,] <-  c( 1980.62641,223.41708)
T[219,] <-  c( 1980.72453,212.1608)
T[220,] <-  c( 1980.77358,226.40703)
T[221,] <-  c( 1980.96981,227.81407)
T[222,] <-  c( 1981.11698,224.82413)
T[223,] <-  c( 1981.26415,228.69347)
T[224,] <-  c( 1981.31321,217.78894)
T[225,] <-  c( 1981.41132,221.83417)
T[226,] <-  c( 1981.50944,228.86935)
T[227,] <-  c( 1981.55849,233.44221)
T[228,] <-  c( 1981.6566,239.24623)
T[229,] <-  c( 1981.70566,215.32663)
T[230,] <-  c( 1981.80377,229.2211)
T[231,] <-  c( 1981.85283,215.85428)
```

```
T[232,] <-  c( 1982,226.75879)
T[233,] <-  c( 1982.04906,223.59297)
T[234,] <-  c( 1982.14717,228.34171)
T[235,] <-  c( 1982.24529,218.66833)
T[236,] <-  c( 1982.39246,220.07538)
T[237,] <-  c( 1982.39246,221.13065)
T[238,] <-  c( 1982.49056,223.41708)
T[239,] <-  c( 1982.63773,234.49748)
T[240,] <-  c( 1982.68679,225.17587)
T[241,] <-  c( 1982.7849,207.58794)
T[242,] <-  c( 1982.83396,217.61307)
T[243,] <-  c( 1982.93208,216.20602)
T[244,] <-  c( 1983.03019,223.94472)
T[245,] <-  c( 1983.1283,220.42714)
T[246,] <-  c( 1983.17736,219.0201)
T[247,] <-  c( 1983.22642,217.43718)
T[248,] <-  c( 1983.37359,216.55779)
T[249,] <-  c( 1983.47169,207.9397)
T[250,] <-  c( 1983.61887,217.08543)
T[251,] <-  c( 1983.71698,221.83417)
T[252,] <-  c( 1983.86415,220.7789)
T[253,] <-  c( 1983.96227,227.98994)
T[254,] <-  c( 1984.06038,224.64824)
T[255,] <-  c( 1984.15849,220.95477)
T[256,] <-  c( 1984.30566,221.48241)
T[257,] <-  c( 1984.35471,217.08543)
T[258,] <-  c( 1984.35471,213.04021)
T[259,] <-  c( 1984.6,223.94472)
T[260,] <-  c( 1984.6,216.73367)
T[261,] <-  c( 1984.79623,208.99498)
T[262,] <-  c( 1984.84528,210.22614)
T[263,] <-  c( 1984.89434,214.799)
T[264,] <-  c( 1985.04151,219.19598)
T[265,] <-  c( 1985.13963,224.47237)
T[266,] <-  c( 1985.18868,216.03015)
T[267,] <-  c( 1985.28679,211.63316)
T[268,] <-  c( 1985.33585,215.15076)
T[269,] <-  c( 1985.33585,209.87437)
T[270,] <-  c( 1985.53207,210.92964)
T[271,] <-  c( 1985.67924,222.53769)
T[272,] <-  c( 1985.67924,221.13065)
T[273,] <-  c( 1985.82642,212.51256)
T[274,] <-  c( 1985.87547,210.22614)
T[275,] <-  c( 1985.97359,218.66833)
T[276,] <-  c( 1986.02264,220.07538)
T[277,] <-  c( 1986.16982,221.48241)
T[278,] <-  c( 1986.21886,219.54774)
T[279,] <-  c( 1986.31698,214.62312)
T[280,] <-  c( 1986.46415,213.39195)
T[281,] <-  c( 1986.51321,209.52261)
T[282,] <-  c( 1986.56226,215.50252)
T[283,] <-  c( 1986.75849,212.86432)
T[284,] <-  c( 1986.75849,206.88441)
T[285,] <-  c( 1986.85661,207.58794)
T[286,] <-  c( 1986.85661,219.0201)
T[287,] <-  c( 1987.00378,212.1608)
T[288,] <-  c( 1987.10189,215.85428)
T[289,] <-  c( 1987.15094,219.54774)
T[290,] <-  c( 1987.29811,216.20602)
T[291,] <-  c( 1987.34717,212.1608)
T[292,] <-  c( 1987.44528,206.88441)
T[293,] <-  c( 1987.49434,209.87437)
T[294,] <-  c( 1987.59245,217.08543)
```

```
T[295,] <-  c( 1987.73962,219.0201)
T[296,] <-  c( 1987.78868,211.45729)
T[297,] <-  c( 1987.78868,218.1407)
T[298,] <-  c( 1987.93585,209.69849)
T[299,] <-  c( 1988.03397,218.1407)
T[300,] <-  c( 1988.23019,217.08543)
T[301,] <-  c( 1988.23019,215.50252)
T[302,] <-  c( 1988.27924,219.19598)
T[303,] <-  c( 1988.37736,213.21608)
T[304,] <-  c( 1988.47547,209.34674)
T[305,] <-  c( 1988.57359,208.46733)
T[306,] <-  c( 1988.62264,227.11055)
T[307,] <-  c( 1988.72076,225.87939)
T[308,] <-  c( 1988.81887,230.80402)
T[309,] <-  c( 1988.86793,222.71356)
T[310,] <-  c( 1988.96603,214.27136)
T[311,] <-  c( 1989.06415,221.30653)
T[312,] <-  c( 1989.21132,220.42714)
T[313,] <-  c( 1989.30943,216.73367)
T[314,] <-  c( 1989.40755,219.0201)
T[315,] <-  c( 1989.50566,210.40201)
T[316,] <-  c( 1989.55472,212.33669)
T[317,] <-  c( 1989.75095,216.90955)
T[318,] <-  c( 1989.84906,213.21608)
T[319,] <-  c( 1989.94717,214.97487)
T[320,] <-  c( 1990.04528,222.18593)
T[321,] <-  c( 1990.09434,218.31659)
T[322,] <-  c( 1990.29057,219.72362)
T[323,] <-  c( 1990.29057,222.01006)
T[324,] <-  c( 1990.48679,214.44724)
T[325,] <-  c( 1990.48679,210.40201)
T[326,] <-  c( 1990.63396,213.74371)
T[327,] <-  c( 1990.73208,210.22614)
T[328,] <-  c( 1990.83018,218.84422)
T[329,] <-  c( 1990.87924,213.04021)
T[330,] <-  c( 1991.02641,210.22614)
T[331,] <-  c( 1991.07547,205.30151)
T[332,] <-  c( 1991.12453,209.17085)
T[333,] <-  c( 1991.22264,216.03015)
T[334,] <-  c( 1991.2717,226.40703)
T[335,] <-  c( 1991.41887,215.50252)
T[336,] <-  c( 1991.41887,214.27136)
T[337,] <-  c( 1991.56604,211.2814)
T[338,] <-  c( 1991.6151,206.70854)
T[339,] <-  c( 1991.71321,207.23618)
T[340,] <-  c( 1991.81132,214.799)
T[341,] <-  c( 1991.90943,197.21106)
T[342,] <-  c( 1992.00755,204.07036)
T[343,] <-  c( 1992.10566,218.66833)
T[344,] <-  c( 1992.15472,214.799)
T[345,] <-  c( 1992.15472,225.87939)
T[346,] <-  c( 1992.30189,221.48241)
T[347,] <-  c( 1992.4,226.23116)
T[348,] <-  c( 1992.44906,221.83417)
T[349,] <-  c( 1992.54717,212.86432)
T[350,] <-  c( 1992.64529,205.47739)
T[351,] <-  c( 1992.69434,200.37688)
T[352,] <-  c( 1992.79245,211.98492)
T[353,] <-  c( 1992.79245,217.78894)
T[354,] <-  c( 1992.98868,218.84422)
T[355,] <-  c( 1992.98868,213.9196)
T[356,] <-  c( 1993.03773,209.34674)
T[357,] <-  c( 1993.18491,212.86432)
```

```
T[358,] <-  c( 1993.28302,224.64824)
T[359,] <-  c( 1993.38113,220.95477)
T[360,] <-  c( 1993.43019,211.63316)
T[361,] <-  c( 1993.52831,206.53267)
T[362,] <-  c( 1993.72453,212.51256)
T[363,] <-  c( 1993.82264,217.61307)
T[364,] <-  c( 1993.8717,210.40201)
T[365,] <-  c( 1993.92075,216.20602)
T[366,] <-  c( 1994.01887,204.59799)
T[367,] <-  c( 1994.11698,212.51256)
T[368,] <-  c( 1994.2151,220.42714)
T[369,] <-  c( 1994.31321,215.32663)
T[370,] <-  c( 1994.36227,213.39195)
T[371,] <-  c( 1994.50944,217.61307)
T[372,] <-  c( 1994.6566,211.45729)
T[373,] <-  c( 1994.6566,202.31155)
T[374,] <-  c( 1994.75471,217.26131)
T[375,] <-  c( 1994.85283,209.17085)
T[376,] <-  c( 1995.04906,199.49748)
T[377,] <-  c( 1995.09811,213.39195)
T[378,] <-  c( 1995.19623,206.70854)
T[379,] <-  c( 1995.29434,208.64322)
T[380,] <-  c( 1995.39246,197.03517)

t <- T[,1]
y <- T[,2]

i.step <- 188   #  1977.9525
step <- matrix(0,length(t))
step[i.step:length(step)] <- 1
step <- step - mean(step)
sol <- solar.abs(t, unit=TRUE)
fit <- lm( y ~ t + sol + step)
r <- resid(fit)
sd.r <- sd(r)

yy <- predict(fit)
get <- matrix(NA, 2000, 3)


for (i in 1:nrow(get)) {
   noise <- sample(r, length(r), replace=TRUE)
   noise <- rnorm(length(t),0, sd(r))
   y.new <- yy + noise
   fit <- lm( y.new ~ t + sol + step)
   get[i,] <- coef(fit)[2:4]
}

op <- par(no.readonly = TRUE)
par(mfrow = c(2, 2))
par(mar = c(2.5, 2.5, 1, 0.8))
par(mgp = c(1.5, 0.5, 0))
par(oma = c(0.5, 0.5, 0.5, 0))
par(bg="white")

plot(t, y, pch=19, ylab="Average Temperature 55-75 km (K)",
xlab="Year")
abline(v=1977.9525, lwd=2)
points(t, predict(fit), col="red", pch=19)

plot(get[,2], get[,1], pch=19, xlab="Solar Coefficient (K/sfu)",
ylab="Time Coefficient (K/year)")
abline(v=mean(get[,2]), h=mean(get[,1]), col="red")
```

```
plot(get[,2], get[,3], pch=19, xlab="Solar Coefficient (K/sfu)",
ylab="Step Coefficient (K)")
abline(v=mean(get[,2]), h=mean(get[,3]), col="red")
plot(get[,3], get[,1], pch=19, xlab="Step Coefficient (K)", ylab="Time
Coefficient (K/year)")
m.step <- mean(get[,3])    ;    m.time <- mean(get[,1])
abline(v=m.step, h=m.time, col="red")


##########################################################################
#
######### THIS CODE GENERATES FIGURE 6 FROM CHAPTER 2 ##########
##########################################################################

## This will do a strait regression on the alo temperatures.
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(simpleboot)


#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
lm.mg <- lm(MgII ~ s1 + s2, data=df) ; mc <- coef(lm.mg) ; A.mg <-
sqrt(mc[2]^2 + mc[3]^2)
df$MgII <- df$MgII/A.mg
alt <- seq(45, 90, by=1)
n <- length(alt)
df$time <- t - mean(t)

get.lt1 <- matrix(NA, n, 3)


###    45 km
i <- 1 ; # i <- length(alt)
h.col <- 26 + i
df$y <- df[, h.col]
lm45 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + MgII +
sol.n, data=df) ; lm1 <- lm45
lm1.b <- lm.boot(lm1, 2000, FALSE)
lm1.c <- samples(lm1.b, name="coef")

par(mar=c(5,5,1,1))
plot(lm1.c[7,], lm1.c[2,], type="p", pch=20, axes=FALSE, xlab=NA,
ylab=NA)
sr. <- quantile(lm1.c[7,], probs=c(0.05, 0.5, 0.95))
lt. <- quantile(lm1.c[2,], probs=c(0.05, 0.5, 0.95))
abline(v=sr.[2], h=lt.[2], col="gray")
abline(v=sr.[-2], h=lt.[-2], lty=2, col="gray")
points(lm1.c[7,], lm1.c[2,], pch=20)
axis(side=1, tick=TRUE)
axis(side=2)
mtext("MgII coefficient (K/solar cycle)", 1, line=2.5)
mtext("Linear trend (K/year)", 2, line=2.5)
text(0.39,0.7, "(b) 45 km")
box()
```

```
t <- lm1$model$time
m <- lm1$model$MgII
plot(t, m, axes=FALSE, xlab=NA, ylab=NA, type="n")
abline(v=0, h=0)
points(t, m, pch=20)
axis(side=1, tick=TRUE)
axis(side=2)
mtext("Normalized MgII data", 2, line=2.5)
mtext("Time (years)", 1, line=2.5)
text(-4,1.3, "(a)")
box()

### 90 km

i <- length(alt)
h.col <- 26 + i
df$y <- df[, h.col]
lm90 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + MgII +
sol.n, data=df)
lm90.b <- lm.boot(lm90, 2000, FALSE)
lm90.c <- samples(lm90.b, name="coef")

par(mar=c(5,5,1,1))
plot(lm90.c[7,], lm90.c[2,], type="p", pch=20, axes=FALSE, xlab=NA,
ylab=NA)
sr. <- quantile(lm90.c[7,], probs=c(0.05, 0.5, 0.95))
lt. <- quantile(lm90.c[2,], probs=c(0.05, 0.5, 0.95))
abline(v=sr.[2], h=lt.[2], col="gray")
abline(v=sr.[-2], h=lt.[-2], lty=2, col="gray")
points(lm90.c[7,], lm90.c[2,], pch=20)
axis(side=1, tick=TRUE)
axis(side=2)
mtext("MgII coefficient (K/solar cycle)", 1, line=2.5)
mtext("Linear trend (K/year)", 2, line=2.5)
text(6,0.6, "(c) 90 km")
box()

##########################################################################
######### THIS CODE GENERATES FIGURE 1 FROM CHAPTER 3 ##########
##########################################################################

source("dir.txt")


calc.solar.phase <- function(time, mg2) {

      w <-  0.5752152710
      t <- time - (max(time) + min(time))/2
      sol.s <- sin(w*t)
      sol.c <- cos(w*t)
      df <- data.frame(sol.s, sol.c, mg2=mg2)
    lm.s <- lm( 'mg2 ~ sol.s + sol.c', df)
      ph <- atan2(coef(lm.s)[3], coef(lm.s)[2])  ;  names(ph) <- NULL
      A <- sqrt(coef(lm.s)[3]^2 +  coef(lm.s)[2]^2)
    I <- matrix(1, length(t))
      nl <- nls(mg2 ~ a*I + b*sin(ww*t + p), start=list(a=0, b=A, p=ph,
ww=w))
      #print(A)
      #plot(t, mg2)  ;  points(t, A*solar2(ph, t, noise=FALSE, A = 1),
pch=19, col="red")
      #points(t, predict(lm.s), col="blue")
      cc <- coef(nl)
```

```
      w <- cc[4]  ;  names(w)  <- NULL
      ph <- cc[3] ;  names(ph) <- NULL
      int <- cc[1] ; names(int) <- NULL
      A <- cc[2]  ; names(A) <- NULL
      return(list(ph=ph, A=A, w=w, int=int, lm=nl))
}


alo <- purify.alo2()

dn <- alo$doy   ;   get <- (dn > 91) & (dn < 273)
alo.sum <- alo[get,]  ;  alo.wint <- alo[!get,]

alo <- alo

t <- alo$time  ;  t <- t - mean(t)
w <-  0.5752152710
MgII <- alo$MgII  ;  MgII <- MgII/sd(MgII) ; MgII <- MgII - mean(MgII)
sol.out <- calc.solar.phase(time=t, mg2=MgII)  ;  w <- sol.out$w
mg.a <- sol.out$A  ;  MgII <- MgII/mg.a  ;  MgII <- MgII - mean(MgII);
sol.out <- calc.solar.phase(time=t, mg2=MgII)
th <- sol.out$ph ;  w <- sol.out$w  ;  alo$MgII <- MgII
sol.sine <- sin(w*t + th)  ;  alo$sol.n <- resid(sol.out$lm)  ;
alo$sol.n <- alo$sol.n/sd(alo$sol.n)

op <- par(no.readonly = TRUE)
par(mfrow = c(2, 1))
par(mar = c(2.5, 2.5, 1, 1))
par(mgp = c(1.5, 0.5, 0))
par(oma = c(0.5, 0.5, 0.5, 0.5))
par(bg="white")
par(cex.axis=1)

off <- 0.25
plot(alo$time.y, MgII, type="p", pch=19, ylab="", xlab="Time (years)",
cex=0.8)
mtext(text="Scaled Mg II", side=2, line=1.6)
abline(h=0)
lines(alo$time.y, sol.sine, lwd=2)
arrows(1996.3, 0+off, 1997.9, -0.4, code=2, length=0.14, col="gray",
lwd=2)
text(1995.5, 0.25+off, "solar like" )
text(1995.5, 0+off, expression("sin "*omega*"t"))
text(1994, 1, "(a)")

par(mar = c(2.5, 2.5, 0.8, 1))
plot(t, alo$sol.n, type="p", ylab="", xlab="", pch=19,  cex=0.8)
mtext(text="Mean Centered Time (years)", side=1, at=0, line=1.5)
mtext(text="Solar-noise (sd=1)", side=2, at=1, line=1.6)
text(-4.65, 3.5, "(b)")
abline(h=0)
```

```
#########################################################################
######### THIS CODE GENERATES FIGURE 2 FROM CHAPTER 3 ##########
#########################################################################
rm(list=ls(all=TRUE))

getBA <- function(t0, ph, th, A=1) {
      s1 <- sin(w*t0)  ;  c1 <- cos(w*t0)
      s2 <- sin(2*w*t0)  ;  c2 <- cos(2*w*t0)

      sisr <- t0*cos(ph - th) - s2*cos(ph + th)/(2*w) -
2*sin(ph)*sin(th)*s1^2/(t0*w*w)
      srti <- 2*cos(ph)/(w^2)*(s1 - w*t0*c1)
      siti <- 2*cos(th)/(w^2)*(s1 - w*t0*c1)
      si2 <- t0 - s2*cos(2*th)/(2*w) - 2*s1^2*sin(th)^2/(t0*w^2)
      ti2 <- 2/3*t0^3
      gamma <- ti2*si2 - (siti)^2
      lt.b <- A/gamma*(si2*srti - siti*sisr)
      A.b <- A/gamma*(ti2*sisr - siti*srti)
      return(list(b=lt.b, a=A.b))
}

t0. <- 20
t <- seq(1, t0., by=0.1)
w <-  0.57521
phs <- seq(0, 2*pi, by=0.01)
ths <- seq(0, 2*pi, by=0.01)
get.bias <- matrix(NA, length(t))
get.ph <- matrix(NA, length(t))
A <- 1
for (i in 1:length(t)) {
   get.lt <- matrix(NA, length(phs), length(ths))
   t0 <- t[i]
   for (ii in 1:length(phs)) {

        ph <- phs[ii]  ;  th <- ths
            BA <- getBA(t0, ph, th)
            lt.b <- BA$b
            A.b <- BA$a
            get.lt[,ii] <- lt.b
      }

      print(round(i/length(t)*100,1))
      get.bias[i] <- max(abs(get.lt))
}

op <- par(no.readonly = TRUE)
par(mar = c(4, 4, 1, 1))   ## Outside Margins  (bottom, left, top,
right)
par(mgp = c(0.5, 0, 0))  ## The margin line (in mex units) for the axis
title, axis labels and axis line. The default is c(3, 1, 0).
par(bg="white")
par(cex.axis=1)
op1 <- par()


plot(2*t, get.bias, pch=19, xlim=c(0,30), axes=FALSE, ylim=c(0,2),
ylab="", xlab="", type="l", lwd=2)
mtext( "Length of Data Set (years)", side=1, at=15, cex=1, line=1.8)
mtext("Cooling Rate Bias (K/year)", side=2, at=0.9, cex=1, line=2)
box()
axis(side=1, at=(0:15)*2, labels=TRUE, cex=0.5, padj=0.8)
axis(side=2, at=(0:20)/10, padj=-1)
abline(v=(0:15)*2, h=(0:20)/10, lt=2, col="gray")
```

```
lines(2*t, get.bias, pch=19, lwd=2)

close.screen(all=TRUE)
split.screen(c(2,2),erase=FALSE)
screen(2)
par(mar = c(2, 2, 2, 2), cex.axis=0.9)  ## Outside Margins  (bottom,
left, top, right)
plot(2*t, get.bias, pch=19, ylim=c(0,0.07), xlim=c(15,30), axes=FALSE,
ylab="", xlab="", type="l", lwd=2)
box()
axis(side=2, at=(0:11)/100, labels=TRUE, padj=-1.2)
axis(side=1, at=(0:15)*2, padj=0.9)
abline(v=(0:15)*2, h=(0:11)/100, lt=2, col="gray")
lines(2*t, get.bias, lwd=2)

close.screen(all=TRUE)
op <- op1
box()
text( -2.5, -0.13, "(a)")
text( 27, 1.63, "(b)")


#########################################################################
######### THIS CODE GENERATES FIGURE 3 FROM CHAPTER 3 ##########
#########################################################################

alt  <-
c(45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,
68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90)
pr.mg2  <-
c(0.2458,0.4141,0.7233,0.8134,0.9244,0.9566,0.9522,0.9592,0.9611,0.8603
,0.8459,0.6774,0.4956,0.3902,0.1238,0.2261,0.6681,0.7375,0.6728,0.8084,
0.8906,0.9014,0.8278,0.865,0.9932,0.9914,0.9879,0.9595,0.9102,0.558,0.6
222,0.749,0.681,0.3252,0.3584,0.5389,0.1717,0.8338,0.684,0.5795,0.4591,
0.6169,0.8581,0.8147,0.0409,0.616)
pr.sin  <-
c(0.9853,0.8481,0.4775,0.1632,0.2931,0.6238,0.7433,0.8365,0.8852,0.8116
,0.8941,0.7142,0.6137,0.6258,0.4623,0.0251,0.653,0.7508,0.7658,0.8286,0
.906,0.866,0.6421,0.6789,0.9714,0.978,0.982,0.9184,0.8545,0.3957,0.6674
,0.8218,0.8162,0.4118,0.2703,0.2711,0.1326,0.6225,0.6503,0.4688,0.4418,
0.389,0.8235,0.8347,0.245,0.4272)
pr.sin_p  <-
c(0.9998,0.9996,0.9965,0.9904,0.9822,0.9415,0.8779,0.8915,0.9305,0.9032
,0.9157,0.7169,0.667,0.7806,0.5255,0.4461,0.665,0.7541,0.7725,0.8291,0.
9117,0.871,0.7603,0.9069,0.9972,0.9967,0.9979,0.9924,0.9896,0.9843,0.99
23,0.9992,0.999,0.9995,0.9996,0.9626,0.9546,0.8095,0.9796,0.9767,0.9949
,0.7909,0.93,0.8951,0.6545,0.6854)
pr.soln  <-
c(1,0.9999,0.9996,0.9983,0.9975,0.9927,0.972,0.9304,0.8137,0.1988,0.378
3,0.0322,0.4501,0.8132,0.8304,0.7056,0.31,0.0799,0.34,0.0858,0.2546,0.4
897,0.7395,0.6814,0.6578,0.4117,3.38000000000001E-
02,0.2109,0.0252,0.2705,0.8072,0.9338,0.9728,0.9548,0.9799,0.9799,0.964
2,0.9692,0.8654,0.9354,0.8788,0.9013,0.7785,0.4565,0.411,0.5155)



par(mar = c(4, 4, 1, 1), mfrow=c(1,2))
plot(pr.mg2, alt, xlim=c(0.85, 1), lty=3, type="l", lwd=3,
ylim=c(46,89), ylab="", xlab="", axes=FALSE)
text(0.98, 88, "(a)")
lines(pr.sin, alt, lty=2, lwd=3)
lines(pr.sin_p, alt, lty=1, lwd=3)
grid(col="dark gray")
```

```
mtext("Confidence Level of
Solar-like Variations", side=1, cex=1, line=2.6)
mtext("Altitude (km)", side=2, cex=1, line=2)
box()
axis(side=2, at=seq(45, 90, by=5), labels=TRUE, cex=0.5, padj=0.5)
axis(side=1, padj= -0.6)
legend(0.915, 64, c(expression(alpha[1]*": Mg II"),
expression(alpha[2]*": sin "*omega * t), expression(alpha[3]*":
sin("*omega*t*"+"*phi*") ")), lty = c(3,2,1), bg = 'white' , cex=.9,
lwd=2)

par(mar = c(4, 0, 1, 6))
plot(pr.soln, alt, xlim=c(0.85, 1), type="l", lwd=3, ylim=c(46,89),
ylab="", xlab="", axes=FALSE)
text(0.98, 87, "(b)")
grid(col="dark gray")
mtext("Confidence Level of
Solar-noise Term", side=1, cex=1, line=2.6)
box()
axis(side=1, padj= -0.6)

##########################################################################
######### THIS CODE GENERATES FIGURE 4 FROM CHAPTER 3 ##########
##########################################################################

rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(simpleboot)

fix.phase.one <- function(phase, control=2) {
    n <- length(phase)

      for (i in 2:n) {
            diff <- phase[i] -       phase[i-1]
            if (abs(diff)> control) {
                  if (sign(diff)>0) {
                        phase[i] <- phase[i] - 2*pi
                        i <- i - 1
                  } else if (sign(diff)<0) {
                        phase[i] <- phase[i] + 2*pi
                        i <- i - 1
                  }
            }
      }

      diff <- phase[n] - phase[n-1]
      if (abs(diff)> control) {
            if (sign(diff)>0) {
                  phase[n] <- phase[n] - 2*pi
            } else if (sign(diff)<0) {
                  phase[n] <- phase[n] + 2*pi
            }
      }

      if (min(phase)<0 ) { phase <- phase + 2*pi }
      if (max(phase)>2*pi) { phase <- phase - 2*pi }
      return(phase)
}

calc.solar.phase <- function(time, mg2) {

      w <-  0.5752152710
```

```
      t <- time - (max(time) + min(time))/2
      sol.s <- sin(w*t)
      sol.c <- cos(w*t)
      df <- data.frame(sol.s, sol.c, mg2=mg2)
    lm.s <- lm( 'mg2 ~ sol.s + sol.c', df)
      ph <- atan2(coef(lm.s)[3], coef(lm.s)[2])  ;  names(ph) <- NULL
      A <- sqrt(coef(lm.s)[3]^2 +  coef(lm.s)[2]^2)
    I <- matrix(1, length(t))
      nl <- nls(mg2 ~ a*I + b*sin(ww*t + p), start=list(a=0, b=A, p=ph,
ww=w))
      #print(A)
      #plot(t, mg2)  ;  points(t, A*solar2(ph, t, noise=FALSE, A = 1),
pch=19, col="red")
      #points(t, predict(lm.s), col="blue")
      cc <- coef(nl)
      w <- cc[4]  ;  names(w)  <- NULL
      ph <- cc[3] ;  names(ph) <- NULL
      int <- cc[1] ; names(int) <- NULL
      A <- cc[2]  ; names(A) <- NULL
      return(list(ph=ph, A=A, w=w, int=int, lm=nl))
}

SEs <- function(lm.object, ph, A, col.names= c("Int.", "LT", "A", "ph",
"sol.n") ) {
  w <-  0.5752152710
  sig <- resid(lm.object)
  t <- lm.object$model[,2]
  m <- lm.object$model  ;  m[,1] <- 1  ;  m[,3] <- sin(w*t + ph)
  m[,4] <- A*cos(w*t + ph)  ;  m <- as.matrix(m)
  AA <- t(m)%*%m
  AA. <- solve(AA)*var(sig)
  AA.d <- sqrt(diag(AA.))
  names(AA.d) <- NULL
  names(AA.d) <- col.names
  return(AA.d)
}

get.sol.max <- function(lm, tt) {
   c <- coef(lm)
   ang <- atan2(c[3],c[4])
   if (ang<0) { ang <- ang + 2*pi}
   w <-  0.5752152710
   # print(ang/w)
   t <- seq(min(tt), max(tt), by=0.025)
   s2 <- c[3]*sin(w*t) ; c2 <- c[4]*cos(w*t)
   # plot(t, s2 + c2) # ; points(tt, c[3]*sin(w*tt) + c[4]*cos(w*tt))
   get <- max(s2 + c2)==(s2 + c2)
   # print(get) ; print(sum(get))
   s.max <- t[get]
   if (s.max<0) { s.max <- s.max + 2*pi/w }
   return(s.max)
}
#  get.sol.max(fit3, tt)

#####################################
#####################################

A. <- matrix(NA, 46, 3)
ph. <- matrix(NA, 46, 3)
mg. <- matrix(NA, 46, 3)
mga. <- matrix(NA, 46, 3)
soln. <- matrix(NA, 46)
```

```
alo <- purify.alo2()

dn <- alo$doy   ;   get <- (dn > 91) & (dn < 273)
alo.sum <- alo[get,]  ;  alo.wint <- alo[!get,]

alo <- alo
t <- alo$time  ;  t <- t - mean(t)
w <-  0.5752152710
MgII <- alo$MgII  ;  MgII <- MgII/sd(MgII) ; MgII <- MgII - mean(MgII)
sol.out <- calc.solar.phase(time=t, mg2=MgII)  ;  w <- sol.out$w
mg.a <- sol.out$A  ;  MgII <- MgII/mg.a  ;  MgII <- MgII - mean(MgII)
th <- sol.out$ph ;  w <- sol.out$w  ;  alo$MgII <- MgII
sol.sine <- sin(w*t + th)  ;  alo$sol.n <- resid(sol.out$lm)  ;
alo$sol.n <- alo$sol.n/sd(alo$sol.n)

th <- 0
solar.max <- "2/12/2002" #"7/31/2001"
get.sm <- solar.max==alo$date
adj <- date.string(1,1,2002,2,12,2002)
time.offset <- -t[get.sm] + adj  ; tt <- t + time.offset
sol.s <- sin(w*tt)
sol.c <- cos(w*tt)
rnd <- rnorm(length(t), 0, 20)
SE.s <- matrix(NA, 46, 4)
I <- matrix(1, length(t))
data <- data.frame( alo, I = I, t, sol.s=sol.s, sol.c=sol.c,
sol.sine=sol.sine)
#  keep <- alo$time > 0 ; data <- data[keep,]

sol.noise <- sol.sine <- MgII <- matrix(NA, 46)
SP <- matrix(NA, 46)
lt1 <- lt2 <- lt3 <- matrix(NA, 46)
t0 <- matrix(NA, 46)
phase <- phase2 <- amp <- matrix(NA, 46)
alt <- 45:90
pr.mg2 <- pr.s <- pr.sp <- pr.soln <- matrix(NA, 46)

for (j in 45:90) {

      ii <- j - 44
      fmla1 <- ccat( "X", j, "km ~ t + sol.sine")
      #fmla1 <- "rnd ~ t + mg2"
      fit1 <- fit.alo(fmla=fmla1, kill=FALSE, data=data)$lm
      sol.sine[ii] <- coef(fit1)[3]
      lt1[ii] <- coef(fit1)[2]
      soln.[ii] <- coef(fit1)[4]
      pr.s[ii] <- 1-summary(fit1)$coefficients[3,4]

      fmla2 <- ccat( "X", j, "km ~ t + MgII")
      #fmla2 <- "rnd ~ t + MgII"
      fit2 <- fit.alo(fmla=fmla2, kill=FALSE, data=data)$lm
      MgII[ii] <- coef(fit2)[3]  ;  b.mg <- MgII[ii]
      lt2[ii] <- coef(fit2)[2]
      SP[ii] <- summary(fit2)$coefficients[3,2]
      pr.mg2[ii] <- 1-summary(fit2)$coefficients[3,4]

      fmla3 <-  ccat( "X", j, "km ~ t + sol.s + sol.c")
      #fmla3 <-  "rnd ~ t + sol.s + sol.c"
      fit3 <- fit.alo(fmla=fmla3, kill=FALSE, data=data)$lm  ;  cc <-
coef(fit3)
      b <- coef(fit3)[2]  ;  c <- coef(fit3)[1] ; tt2 <- fit3$model[,2]
+ time.offset
      mod <- fit3$model
```

```
      ph2 <- atan2(coef(fit3)[3], coef(fit3)[4]) ; if (ph2<0){ph2 <-
ph2+2*pi}
      ph <- get.sol.max(fit3, tt)
      A <- sqrt( coef(fit3)[4]^2 + coef(fit3)[3]^2)
      names(ph) <- names(b) <- names(c) <- names(A) <- NULL
      lt3[ii] <- cc[2]  ;  amp[ii] <- A  ;  phase[ii] <- ph ;
phase2[ii] <- ph2
      AAA <- SEs(fit3, ph2, A, col.names= c("Int.", "LT", "A", "ph"))
      SE.s[ii, ] <- AAA
      sol.noise[ii, 1] <- cc[5]


      ss2 <- (coef(fit3)[3]*mod$sol.s +  coef(fit3)[4]*mod$sol.c)/A ;
mod$sol.s <- ss2
      fmla4 <- ccat( "X", j, "km ~ t + sol.s")
      fit4 <- lm(fmla4, data=mod)
      pr.sp[ii] <- 1-summary(fit4)$coefficients[3,4]
      ## pr.soln[ii] <- 1-summary(fit4)$coefficients[4,4]
      print(j)

}

mg. <- 2*MgII
mga. <- 2*sol.sine
ph. <- phase # fix.phase.one(phase - th)
A. <- amp*2
alt <- 45:90
mtext.cex <- 0.8

op <- par(no.readonly = TRUE)
par(mfrow = c(1, 3))
par(mgp = c(1.5, 0.5, 0))
par(oma = c(0.5, 0.5, 0.5, 0.5))
par(bg="white")
par(cex.axis=1.5)

rems.Amp  <- c(0.6,1.2,1.6,2.2,2,1.8,1.2,2,3.2,3.2,2.6,2.2) ## Rems
2007   40 N
rems.alt <-
c(48.32491,51.10865,53.68817,57.49702,60.42884,62.46151,65.25389,73.161
14,75.70142,77.47369,79.93315,82.06233)  ## Rems 2007
rems.pha  <- c(10.8,10.4,10.1,9.6,9.2,9,8.4,1.3,0.9,0.8,0.6,0.3)  ##
Rems 2007  40 N

rems09.p <- c(1.5,3,4.5,4.5,3,1.5,0,1.5,1.5)  ## Rems 09 Figure 5
rems09.alt2 <- c(72.1,69.2,68.6,67,66.7,66.3,65.9,65.6,58.5)
rems09.alt1 <- c(77.47369, 75.10888, 73.59735, 72.16893, 71.30992,
69.86640, 65.96773, 65.25389, 52.96298)
rems09.amp <- c(3,3,2.5,2,1.5,1,1,1.5,1.5)

ph2 <- c(9.9,9.9,9.7,9.3,10.7,10.9,10.9,11,10.9) ## Rems 2007 50 N
alt2  <-
c(57.49702,60.42884,62.46151,65.25389,73.16114,75.70142,77.47369,79.933
15,82.06233)
amp2  <- c(1.8,2,2,1.2,2.6,3.2,3.2,3,2.8)  ## Rems 2007 50 N
cex <- 1.4

########## AMPLITUDE ####################
par(mar = c(5, 5, 2, 1))
plot(A.[,1], alt, pch=19, xlim=c(-1, 10), type="l", lwd=2, xlab="",
ylab="", main="", axes=FALSE)
abline(v=0, col="dark gray")
text(8, 46, "(a)", cex=1.5)
```

```
mtext( "Amplitude of the Atmospheric Solar Response
Kelvin (solar max - solar min)", side=1, at=4.8, cex=mtext.cex,
line=3.7)
mtext("Altitude (km)", side=2, at=68, cex=1, line=2.8)
box()
axis(side=2, at=seq(45, 90, by=5), labels=TRUE, cex=0.5, padj=-0.4)
axis(side=1, padj=0.4)
arrows(6.5, 65, 5.4, 70, code=2, length=0.15, lwd=2, col="gray")
text(7, 64, "95% CI", cex=1.3, font=10)
grid()
lines(A.[,1] + SE.s[,3]*1.97, alt, lty=2)
lines(A.[,1] - SE.s[,3]*1.97, alt, lty=2)
points(rems.Amp, rems.alt, pch=19, cex=cex)
points(amp2, alt2, pch=17, cex=cex)
points(rems09.amp, rems09.alt1, pch=2, cex=cex)

lines(A.[,1], alt, lwd=2)
################################################
########  PHASE ###################################
par(mar = c(5, 1, 2, 2))
plot(phase2/w, alt, pch=19, xlim=c(-1, 15), type="l", xlab="", ylab="",
lwd=2, axes=FALSE)
# points(ph., alt, pch=19, col="green")
abline(v=c(0, pi, 2*pi, -pi, -2*pi, 3*pi)/w , col="gray")
abline(v=c(pi/2, 3*pi/2, -pi/2, -3*pi/2, 5*pi/2)/w , col="gray", lty=2)
text(12, 46, "(b)", cex=1.5)
mtext( c(expression(0), expression(pi*"/"*2), expression(pi),
expression(3*pi*"/"*2), expression(2*pi)), side=3, at=c(0, pi/2, pi,
3*pi/2, 2*pi)/w, cex=0.8, line=-1)
mtext( "Phase of atmospheric solar response
Years from solar maximum", side=1, at=7.5, cex=mtext.cex, line=3.7)
axis(side=2, at=seq(45, 90, by=5), labels=FALSE, cex=0.6, padj=-0.4)
axis(side=1, padj=0.4)

points( rems.pha, rems.alt, pch=19, cex=cex) ; points( rems.pha + 11,
rems.alt, pch=19, cex=cex) ; points( rems.pha - 11, rems.alt, pch=19,
cex=cex)
points(ph2 , alt2, pch=17, cex=cex)  ; points(ph2 - 11, alt2, pch=17,
cex=cex)
points(rems09.p, rems09.alt2, pch=2, cex=cex)
points(rems09.p+11, rems09.alt2, pch=2, cex=cex)

off.x <- -8  ;  off.y <- -15
arrows(11+off.x, 68 + off.y, 12.5+off.x, 74+off.y, code=2, length=0.15,
lwd=2, col="gray")
text(11+off.x, 67+off.y, "95% CI", cex=1.3, font=10)
box()

lines( phase2/w + SE.s[,3]*1.97/w, alt, lty=2)
lines( phase2/w - SE.s[,3]*1.97/w, alt, lty=2)

lines( phase2/w + 2*pi/w, alt, lty=1, lwd=2)
lines( phase2/w + 2*pi/w + SE.s[,3]*1.97/w, alt, lty=2)
lines( phase2/w + 2*pi/w - SE.s[,3]*1.97/w, alt, lty=2)

lines( phase2/w - 2*pi/w, alt, lty=1, lwd=2)
lines( phase2/w - 2*pi/w + SE.s[,3]*1.97/w, alt, lty=2)
lines( phase2/w - 2*pi/w - SE.s[,3]*1.97/w, alt, lty=2)
lines(phase2/w, alt, lwd=2)

################################################
########## FIX PROXY ###################################
par(mar = c(5, 0, 2, 3))
```

```
plot(mg.[,1], alt, pch=19, xlim=c(-5,6), xlab="", type="l", ylab="",
lwd=2, axes=FALSE)
lines(mga., alt, lty=4, lwd=2)
abline(v=0,col="dark gray")
lines(SP*1.96, alt, lty=2)
lines(-SP*1.96, alt, lty=2)

legend(-0.5, 73, c("Mg II", expression("sin "*omega * t), "Remsberg
(09)", expression("Remsberg (07) "*40^o*" N"), expression("Remsberg
(07) "*50^o*" N")),
              pch=c(NA,NA,2,19, 17), lty = c(1, 4, NA, NA, NA), bg =
'white' , cex=1, lwd=1.2)

text(4, 46, "(c)", cex=1.5)
mtext( "Solar Proxy Coefficients
Kelvin (solar max - solar min)", side=1, at=0, cex=mtext.cex, line=3.5)
axis(side=2, at=seq(45, 90, by=5), labels=FALSE, cex=0.5, padj=-0.4)
axis(side=1, padj=0.4)
box()
arrows(-3.2, 50, -0.8, 55, code=2, length=0.15, lwd=2, col="gray")
text(-4, 49, "95% CI", cex=1.2, font=10)
lines(mg.[,1], alt, lwd=2)
################################################

#########################################################################
#
######### THIS CODE GENERATES FIGURE 3 FROM CHAPTER 4 ##########
#########################################################################
## This will do a strait regression on the alo temperatures.
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)

coef.corr <- function(A, se2) {
   A <- as.matrix(A)
   v <- solve(t(A)%*%A)*se2
   v.d <- diag(v)
   n <- nrow(v) ; c <- ncol(v)
   cor <- matrix(NA, n, c)
   for (j in 1:n) {
      for (k in 1:c) {
         sjj <- v[j,j] ; skk <- v[k,k]
           cor[j,k] <- v[j,k]/( sqrt(sjj)*sqrt(skk) )
        }
      }
      return(cor)
}



#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
alt <- seq(45, 90, by=1)
n <- length(alt)
df$time <- t - mean(t)

get.lt1 <- matrix(NA, n, 3)
```

```
get.lt2 <- matrix(NA, n, 3)
get.lt3 <- matrix(NA, n, 3)
get.lt4 <- matrix(NA, n, 3)
get.lt5 <- matrix(NA, n, 3)

for (i in 1:n) {
   h.col <- 26 + i
   df$y <- df[, h.col]
   lm1 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n, data=df)
   c <- summary(lm1)$coefficients
   get.lt1[i,1] <- c[2,1]  ;  get.lt1[i,2] <- c[2,1] - c[2,2]*1.96 ;
get.lt1[i,3] <- c[2,1] + c[2,2]*1.96
   xx <- lm1$model[,2] ; yy <- resid(lm1) ; xx2 <- xx*xx ; xx3 <-
xx2*xx ; l <- predict(lm(yy ~ xx + xx2 + xx3))
   #plot(xx, resid(lm1))
   #points(xx, l, col="red")
   #abline(h=0)
   # readline("enter...")

   lm2 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + MgII,
data=df)
   c <- summary(lm2)$coefficients
   get.lt2[i,1] <- c[2,1]  ;  get.lt2[i,2] <- c[2,1] - c[2,2]*1.96 ;
get.lt2[i,3] <- c[2,1] + c[2,2]*1.96

   lm3 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + sol.n,
data=df)
   c <- summary(lm3)$coefficients
   get.lt3[i,1] <- c[2,1]  ;  get.lt3[i,2] <- c[2,1] - c[2,2]*1.96 ;
get.lt3[i,3] <- c[2,1] + c[2,2]*1.96

   lm4 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 +
sol.n, data=df)
   c <- summary(lm4)$coefficients
   get.lt4[i,1] <- c[2,1]  ;  get.lt4[i,2] <- c[2,1] - c[2,2]*1.96 ;
get.lt4[i,3] <- c[2,1] + c[2,2]*1.96

}

v <- lm1$model ; v[,1] <- v[,1]*0+ 1  ;  se2 <- var(resid(lm1))
coef.corr(v, se2)
coef.corr(v[,-1], se2)

par(mfrow=c(1,2))
par(mar = c(4, 4, 1, 0))

plot(get.lt1[,1], alt, type="l", lwd=2, xlim=c(-3,0.5), ylab="",
xlab="", axes=FALSE)
abline(v=0, col="gray")
lines(get.lt1[,1], alt, lwd=2)
lines(get.lt1[,3], alt, type="l", lt=2, lwd=2)
lines(get.lt1[,2], alt, type="l", lt=2, lwd=2)
box()
mtext("Altitude (km)", side=2, at=65, padj=-4, cex=1)
axis(side=2, at=seq(45, 90, by=5), labels=TRUE, cex=0.5, padj=0.5,
hadj=1.5, las=1)
axis(side=1, labels=TRUE, cex=0.2, padj=0, las=1)  #x-axis
text(-2, 70, "95% CI", cex=0.9)
arrows(-1.5, 70, -1, 72.5, lwd=2, length=0.1)
text(-2.5, 60, "(a)")

par(mar=c(4,0,1,1))
```

```
plot(get.lt1[,1], alt, type="n", lwd=2, ylab="", xlab="", axes=FALSE)
abline(v=0, col="grey")
lines(get.lt1[,1], alt, type="l", lwd=2, ylab="", xlab="")
lines(get.lt2[,1], alt, lt=2, lwd=2)
lines(get.lt3[,1], alt, lt=3, lwd=2)
lines(get.lt4[,1], alt, lt=4, lwd=2)
lines(get.lt1[,3], alt, type="l", lt=2, lwd=3, col="gray")
lines(get.lt1[,2], alt, type="l", lt=2, lwd=3, col="gray")
mtext("Linear Cooling Trend (K/year)", side=1, at=-2, padj=3, cex=0.9)
axis(side=1, at=c(0, -0.5, -1, -1.5), cex=0.2, padj=0, las=1) #x-axis
axis(side=1, at=c(-0.25, -0.75, -1.25, -1.75), labels=FALSE, cex=0.2,
padj=0, las=1) #x-axis
box()
text(-1.7, 60, "(b)")
legend(-2, 55, c("sine + cosine","Mg II", "sine and cosine omitted",
"sine only"), lty=1:7, cex = 0.9, lwd=2, bty="n")

#########################################################################
#
######### THIS CODE GENERATES FIGURE 4 FROM CHAPTER 4 ##########
#########################################################################
rm(list=ls(all=TRUE))
library(Hmisc)

# FROM BEIG ET AL. 2003
# Ramaswamy et al. 2001
t.mp <- c(-6.8,0,0.2,0.6,5,-10.5,0.3,0.3,-9,1.2,-1.2,0,-0.24,-0.3,0,-
1,-1.5,0,-1.4,-2.1,-6,-1.93,-2.5)
t.ms <- c(-2.5,-2.5,-4.5,-3.3,-2.2,-1.5,-3,-5.6,-3.5,-8.8,-5.2,-2,-
3.5,-10,-2.5,-0.24,-1.5,-2,-3,-2,-1,-1.6)
t.sp <- c(-0.5,-0.4,-0.3,-0.3,-0.2,-0.5,0,-1,2,-4,0,-0.5,0,-0.5,-0.2,-
0.8,0,-0.9,0,-1,-0.2,-0.4,0,-4.5,-0.38,-0.43,-0.4,-0.4,-1.9,1.67,-
1.85,4.07,-0.27,-0.44,-0.63,-0.73,-1.2,0.8,-0.2,-1.6,1.5,-0.16)

mp <- hist(t.mp, breaks=12, plot=FALSE)
ms <- hist(t.ms, breaks=12, plot=FALSE)

par(mar = c(4, 6, 2, 2))
plot(1:16, -11:4, xlim=c(-6, 7), ylim=c(0, 17), type="n", axes=FALSE,
xlab="", ylab="")
v <- barplot(-mp$counts, col="light gray", add=TRUE, horiz=TRUE,
space=0, axes=FALSE)
barplot(ms$counts, col="light gray", horiz=TRUE, space=0, add=TRUE,
axes=FALSE)

ri <- max(v) - min(v) ; ru <- max(mp$mids) - min(mp$mids) ; x <-
mp$mids
y.new <- ri/ru*(x + 10.5) + 0.5
ylab=mp$mids
axis(side=2, at=y.new, labels=ylab, cex=0.2, padj= 0.25, las=1)
axis(side=1, at=(-6:6), labels=abs(-6:6), cex=0.2, padj=-0.75, las=1)
mtext( "Count", side=1, at=0, cex=0.9, line=1.5)
mtext( "Temperature Trend (K/decade)", side=2, at=8.5, cex=0.9,
line=3.2)
text(3, 16, "(B) Mesosphere", cex=0.9)
text(-3, 16, "(A) Mesopause", cex=0.9)
text(-5, 2, ccat("N=", length(t.mp)), cex=0.8)
text(5, 2, ccat("N=", length(t.ms)), cex=0.8)
box()
```

```
#########################################################################
######### THIS CODE GENERATES FIGURE 5 FROM CHAPTER 4 ##########
#########################################################################

op <- par(no.readonly = TRUE)
par(mfrow = c(1, 2))
par(mar = c(3.5, 4, 0.5, 1))

##    FOR S1 AND S2, WHEN THE DATA IS ABOUT EVENLY DIVIDED
plot(get.ltl4[,1], alt, type="n", xlim=c(-5, 0.6), col="black",
xlab=NA, ylab=NA, axes=FALSE)
abline(v=0, col="gray")
lines(get.ltl4[,1], alt, lwd=2, lt=2)
lines(get.ltg4[,1], alt, lwd=3, lt=3)
lines(get.lt[,1], alt, lwd=2, lt=1)

mtext("Linear Cooling Trend (K/year)", side=1, at=-2.2, padj=3,
cex=0.9)
mtext("Altitude (km)", side=2, at=70, padj=-4, cex=1)
axis(side=1, cex=0.2, padj=-0.5) #x-axis
axis(side=2, cex=0.2, padj=0.4, las=1) #x-axis
box()
e1 <- expression(S[1]*" (t<4 y)")
e2 <- expression(S[2]*" (t>4 y)")
e3 <- "all"
legend(-4, 65, c(e1,e2, e3), lty=c(2,3,1), cex = 0.9, lwd=2, bty="n")
text(0.4, 89, "(a)")

##    FOR  FULL AND  FIRST  YEAR REMOVED
par(mar = c(3.5, 2, 0.5, 1))
plot(get.lt[,1], alt, xlim=c(-2, 0.3), type="n", col="black", xlab=NA,
ylab=NA, axes=FALSE)
abline(v=0, col="gray")
lines(get.lt2[,1], alt, lwd=3, lt=3)
lines(get.lt[,1], alt, lwd=2, lt=1)  ## full data set

mtext("Linear Cooling Trend (K/year)", side=1, at=-0.6, padj=3,
cex=0.9)
axis(side=1, cex=0.2, padj=-0.5) #x-axis
axis(side=2, cex=0.2, padj=0.4, las=1) #x-axis
box()
e2 <- expression(S[3]*" (t>1 y)")
e3 <- "all"
legend(-1.5, 65, c(e2, e3), lty=c(3,1), cex = 0.9, lwd=2, bty="n")
text(0.2, 89, "(b)")
cbind(get.lt2[,1], get.lt[,1])

#########################################################################
######## THIS CODE GENERATES FIGUREs 6 and 7 FROM CHAPTER 4 #######
#########################################################################

rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)

alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
df$time <- t - mean(t)

df$y <- df$X89km
```

```
lm1 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 +
sol.n, data=df)
t <- lm1$model[,2] ; t2 <- t^2 ; t3 <- t^3 ; t4 <- t^4
lm1 <- lm(y ~ sin2pit + cos2pit + sin4pit + cos4pit + sol.n, data=df)
r <- resid(lm1)
lm2 <- lm(r ~ t + t2 + t3 + t4)

par(mar = c(4, 4, 1, 1))
plot(t,r, pch=20, xlab="", ylab="", axes=FALSE)
abline(h=0)
points(t,r, pch=20)
lines(t, predict(lm2), lwd=2, col="dark gray")

mtext("Time (years)", side=1, line=2)
mtext("Residuals from 85 km", side=2, line=2.5)
axis(side=1, cex=0.2, padj=-0.8, las=1)
axis(2)
box()

##########################################################################
######### THIS CODE GENERATES FIGURE 8 FROM CHAPTER 4 ##########
##########################################################################

## This will do a strait regression on the alo temperatures.
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)

coef.corr <- function(A, se2) {
   A <- as.matrix(A)
   v <- solve(t(A)%*%A)*se2
   v.d <- diag(v)
   n <- nrow(v) ; c <- ncol(v)
   cor <- matrix(NA, n, c)
   for (j in 1:n) {
      for (k in 1:c) {
         sjj <- v[j,j] ; skk <- v[k,k]
           cor[j,k] <- v[j,k]/( sqrt(sjj)*sqrt(skk) )
       }
      }
      return(cor)
}

#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
alt <- seq(45, 90, by=1)
n <- length(alt)
df$time <- t - mean(t)

get.lt1 <- matrix(NA, n, 3)
get.lt12 <- matrix(NA, n, 3)
nn <- matrix(NA, n)

for (i in 1:n) {

   h.col <- 26 + i
   df$y <- df[, h.col]
```

```
   lm1 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n, data=df)
   c <- summary(lm1)$coefficients
   get.lt1[i,1] <- c[2,1]  ;  get.lt1[i,2] <-  -c[2,2]*1.96 ;
get.lt1[i,3] <- c[2,2]*1.96
   nn[i] <- length(resid(lm1))

   lm12 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + sol.n,
data=df)
   c <- summary(lm12)$coefficients
   get.lt12[i,1] <- c[2,1]  ;  get.lt12[i,2] <- c[2,2]*1.96 ;
get.lt12[i,3] <- -c[2,2]*1.96

}

par(mar = c(4, 4, 1, 1))
plot(get.lt1[,1], alt, type="l", lwd=2, xlim=c(-2.5,1.5), ylab="",
xlab="", axes=FALSE)
abline(v=0, col="gray")
lines(get.lt1[,1], alt, lwd=2, lt=1)
lines(get.lt1[,3], alt, type="l", lt=2, lwd=2)
lines(get.lt1[,2], alt, type="l", lt=2, lwd=2)
lines(get.lt12[,1], alt, type="l", lt=1, lwd=2, col="dark gray")
lines(get.lt12[,2], alt, type="l", lt=2, lwd=2, col="dark gray")
lines(get.lt12[,3], alt, type="l", lt=2, lwd=2, col="dark gray")
box()
mtext("Altitude (km)", side=2, at=67, padj=-3.4, cex=1)
mtext("Linear trend (K/year)", side=1, at=-0.5, padj=2.5, cex=1)
axis(side=2, at=seq(45, 90, by=5), labels=TRUE, cex=0.5, padj=0.3,
hadj=0.8, las=1)
axis(side=1, labels=TRUE, cex=0.2, padj=-0.5, las=1)  #  x-axis
text(-1, 50, "95% CI", cex=0.9)
arrows(-0.7, 50, -0.25, 52.5, lwd=2, length=0.1)
arrows(-0.7, 50, -0.12, 49, lwd=2, length=0.1)
legend(-2.5, 70, c("Full Model","Solar-like terms
 omitted"), col=c("black","gray"),lty=1, cex = 0.9, lwd=2, bty="n")
```

```
#########################################################################
######### THIS CODE GENERATES FIGURE 11 FROM CHAPTER 4 ##########
#########################################################################

rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(simpleboot)

calc.solar.phase <- function(time, mg2) {

      w <-  0.5752152710
      t <- time - (max(time) + min(time))/2
      sol.s <- sin(w*t)
      sol.c <- cos(w*t)
      df <- data.frame(sol.s, sol.c, mg2=mg2)
    lm.s <- lm( 'mg2 ~ sol.s + sol.c', df)
      ph <- atan2(coef(lm.s)[3], coef(lm.s)[2])  ;  names(ph) <- NULL
      A <- sqrt(coef(lm.s)[3]^2 +  coef(lm.s)[2]^2)
    I <- matrix(1, length(t))
      nl <- nls(mg2 ~ a*I + b*sin(ww*t + p), start=list(a=0, b=A, p=ph,
ww=w))
      #print(A)
      #plot(t, mg2)  ;  points(t, A*solar2(ph, t, noise=FALSE, A = 1),
pch=19, col="red")
      #points(t, predict(lm.s), col="blue")
      cc <- coef(nl)
      w <- cc[4]  ;  names(w)  <- NULL
      ph <- cc[3] ;  names(ph) <- NULL
      int <- cc[1] ; names(int) <- NULL
      A <- cc[2]  ; names(A) <- NULL
      return(list(ph=ph, A=A, w=w, int=int, lm=nl))
}

#####################################
#####################################

A. <- matrix(NA, 46, 3)
ph. <- matrix(NA, 46, 3)
mg. <- matrix(NA, 46, 3)
mga. <- matrix(NA, 46, 3)
soln. <- matrix(NA, 46)

alo <- purify.alo2()

dn <- alo$doy   ;   get <- (dn > 91) & (dn < 273)
alo.sum <- alo[get,]  ;  alo.wint <- alo[!get,]

alo <- alo

t <- alo$time  ;  t <- t - mean(t)
w <-  0.5752152710
MgII <- alo$MgII  ;  MgII <- MgII/sd(MgII) ; MgII <- MgII - mean(MgII)
sol.out <- calc.solar.phase(time=t, mg2=MgII)  ;  w <- sol.out$w
mg.a <- sol.out$A  ;  MgII <- MgII/mg.a  ;  MgII <- MgII - mean(MgII);
sol.out <- calc.solar.phase(time=t, mg2=MgII)
th <- sol.out$ph ;  w <- sol.out$w  ;  alo$MgII <- MgII
sol.sine <- sin(w*t + th)  ;  alo$sol.n <- resid(sol.out$lm)  ;
alo$sol.n <- alo$sol.n/sd(alo$sol.n)

sol.s <- sin(w*t)
sol.c <- cos(w*t)
```

```
sol.s2 <- sin(2*w*t)
sol.c2 <- cos(2*w*t)
rnd <- rnorm(length(t), 0, 20)
SE.s <- matrix(NA, 46, 5)
I <- matrix(1, length(t))
data <- data.frame( alo, I = I, t, sol.s=sol.s, sol.c=sol.c,
sol.sine=sol.sine, y=t*0)
doy <- data$doy
keep <- (doy>79) & (doy<232) ##  THIS IS FOR SUMMER

#  keep <- alo$time > 0 ; data <- data[keep,]

ss.s <- ss.w <- mg.w <- mg.s <- matrix(NA, 46)
A.s <- A.w <- matrix(NA, 46)
alt <- 45:90
lt.s <- lt.w <- matrix(NA, 46)

for (j in 45:90) {

    h.col <- j-45+27
      ii <- j - 44
      data$y <- data[,h.col]
      data.s <- data[keep,]
    data.w <- data[!keep,]
      fmla1 <- "y ~ t + sol.sine + sol.n"
      fit1.s <- fit.alo(fmla=fmla1, kill=FALSE, data=data.s)$lm
      fit1.w <- fit.alo(fmla=fmla1, kill=FALSE, data=data.w)$lm
      ss.w[ii] <- coef(fit1.w)[3]*2
      ss.s[ii] <- coef(fit1.s)[3]*2

      fmla2 <- "y ~ t + MgII"
      fit2.s <- fit.alo(fmla=fmla2, kill=FALSE, data=data.s)$lm
      fit2.w <- fit.alo(fmla=fmla2, kill=FALSE, data=data.w)$lm
      mg.s[ii] <- coef(fit2.s)[3]*2
      mg.w[ii] <- coef(fit2.w)[3]*2

      fmla3 <- "y ~ t + sol.s + sol.c + sol.n"
      fit3.w <- fit.alo(fmla=fmla3, kill=FALSE, data=data.w)$lm  ;
cc.w <- coef(fit3.w)
      fit3.s <- fit.alo(fmla=fmla3, kill=FALSE, data=data.s)$lm  ;
cc.s <- coef(fit3.s)
      A.s[ii] <- sqrt( coef(fit3.s)[4]^2 + coef(fit3.s)[3]^2)*2
      A.w[ii] <- sqrt( coef(fit3.w)[4]^2 + coef(fit3.w)[3]^2)*2
      lt.s[ii] <- coef(fit3.s)[2] ; lt.w[ii] <- coef(fit3.w)[2]

      print(j)

}

alt <- 45:90
op <- par(no.readonly = TRUE)
par(mfrow = c(1, 3))
par(mar = c(5, 5, 2, 1))
par(mgp = c(1.5, 0.5, 0))
par(oma = c(0.5, 0.5, 0.5, 0.5))
par(bg="white")
par(cex.axis=1.5)

#####################################
########## AMPLITUDE ################
#####################################
```

```
plot(A.w, alt, pch=19, xlim=c(-1, 12), type="l", lwd=2, xlab="",
ylab="", main="", axes=FALSE)
lines(A.s, alt, lty=2, lwd=2)
abline(v=0, col="dark gray")
text(10, 46, "(a)", cex=1.5)
mtext( "Atmospheric Solar Response Amplitude
Kelvin (solar max - solar min)", side=1, at=4.8, cex=0.9, line=3.7)
mtext("Altitude (km)", side=2, at=68, cex=1, line=2.8)
box()
axis(side=2, at=seq(45, 90, by=5), labels=TRUE, cex=0.5, padj=-0.4)
axis(side=1, padj=0.4)
legend(5, 55, c("Summer", "Winter"), lty = c(2, 1), bg = 'white' ,
cex=1, lwd=2)
grid()

################################################
########  sin wt ###############################
################################################
par(mar = c(5, 2, 2, 2))
plot(mg.w, alt, pch=19, xlim=c(-10,10), xlab="", type="l", ylab="",
lwd=2, axes=FALSE)
abline(v=0,col="dark gray")
lines(mg.s, alt, lty=2, lwd=2)
text(7, 46, "(b)", cex=1.5)
mtext( "MgII Coefficient
Kelvin (solar max - solar min)", side=1, at=0, cex=0.9, line=3.5)
axis(side=2, at=seq(45, 90, by=5), labels=FALSE, cex=0.5, padj=-0.4)
axis(side=1, padj=0.4)
grid()
box()

################################################
########## FIX PROXY ###########################
################################################

par(mar = c(5, 2, 2, 2))
plot(ss.w, alt, pch=19, xlim=c(-10,10), xlab="", type="l", ylab="",
lwd=2, axes=FALSE)
lines(ss.s, alt, lty=2, lwd=2)
abline(v=0,col="dark gray")
text(7, 46, "(c)", cex=1.5)
mtext( "Solar-sine Coefficient
Kelvin (solar max - solar min)", side=1, at=0, cex=0.9, line=3.5)
axis(side=2, at=seq(45, 90, by=5), labels=FALSE, cex=0.5, padj=-0.4)
axis(side=1, padj=0.4)
grid()
box()
#########################################################################
######### THIS CODE GENERATES FIGURE 12 FROM CHAPTER 4 ##########
#########################################################################

## AO SAO amp and phase
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(fUtilities)
library(fields)

dir.F <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\Summary And Comp\\French\\"
dir.H <- "C:/Documents and Settings/Troy Wynn/My
Documents/dissertation/Dissertation/Summary And Comp/HALOE/process/"
source(ccat(dir.F,"process.txt"))
```

```
source(ccat(dir.H,"process 2.txt"))

SEs <- function(t, ao.p, ao.a, sao.p, sao.a, s.p, s.a, sn, w, noise) {
   n <- length(t)
   se <- sd(noise)
   v.1 <- matrix(1,n)
   s2 <- sin(2*pi*t + ao.p) ; c2 <- cos(2*pi*t + ao.p)
   s4 <- sin(4*pi*t + sao.p) ; c4 <- cos(4*pi*t + sao.p)
   sw <- sin(w*t + s.p) ; cw <- cos(w*t + s.p)
   Y <- cbind(v.1, t, s2, ao.a*c2, s4, sao.a*c4, sw, s.a*cw, sn)
   YY <- sqrt(diag(solve(t(Y)%*%Y)))*se
   return(YY)
}

SE2 <- function(t, ao.p, ao.a, noise) {
   n <- length(t)
   se <- sd(noise)
   v.1 <- matrix(1,n)
   s2 <- sin(2*pi*t + ao.p) ; c2 <- cos(2*pi*t + ao.p)
   Y <- cbind(v.1, t, s2, ao.a*c2)
   YY <- sqrt(diag(solve(t(Y)%*%Y)))*se
   return(YY)
}

get.mean <- function(doy, r, dw=5) {
   doys <- seq(1, 365, by=1)
   doys <- c(320:365, doys, 1:45)  ## day 1 is at 47
   w <- (dw-1)/2
   sds <- matrix(NA, 365) ; j <- 1
   for (i in 1:365) {
     dds <- doys[(46+i-w+2):(46+i+w)]
       get <- matrix(FALSE, length(doy))
       for (p in 1:length(dds)) { get <- doy==dds[p] | get }
       if (sum(get)==0) { sds[j] <- NA } else { sds[j] <- mean(r[get])
}
     j <- j + 1  # ; print(sum(get))
   }
   return(sds)
}


#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
s5 <- sin(5*pi*t) ; c5 <- cos(5*pi*t)
s6 <- sin(6*pi*t) ; c6 <- cos(6*pi*t)
df <- data.frame(alo, s1, s2, y=s2*0, s5, c5, s6, c6)
alt <- 45:90
n <- length(alt)
#df$sin2pit <- sin(2*pi*df$time) ; df$cos2pit <- cos(2*pi*df$time)
#df$sin4pit <- sin(4*pi*df$time) ; df$cos4pit <- cos(4*pi*df$time)
AO.a <- AO.p <- SAO.a <- SAO.p <- matrix(NA, n, 2)
sds <- matrix(NA, n)
sao <- matrix(NA, 46, 365)

for (i in 1:n) {
   h.col <- 26 + i
   df$y <- df[, h.col] ; time <- df$time
```

```
   lm0 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n + doy, data=df)
   doy <- lm0$model$doy
   lm1 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n, data=df)
   cc <- coef(lm1) ; mod <- lm1$model
   deseason <- mod[,1] - (cc[1] + cc[2]*mod[,2] + cc[3]*mod[,3] +
cc[4]*mod[,4] + cc[7]*mod[,7] + cc[8]*mod[,8] + cc[9]*mod[,9] )
   sao[i,] <- get.mean(deseason, doy, dw=30)
   c <- coef(lm1) ; sds[i] <- sd(resid(lm1))
   noise <- resid(lm1) ; tt <- lm1$model$time ; sn <- lm1$model$sol.n
   ao.a <- sqrt(c[3]^2 + c[4]^2) ; ao.p <- atan2(c[3],c[4]) ; B <- c[1]
   sao.a <- sqrt(c[5]^2 + c[6]^2); sao.p <- atan2(c[5],c[6]) ; b <-
c[2]
   s.a <- sqrt(c[7]^2 + c[8]^2) ; s.p <- atan2(c[7],c[8])
   ses <- SEs(tt, ao.p, ao.a, sao.p, sao.a, s.p, s.a, sn, w, noise)

   #alo.data$time <- alo.data$time - mean(alo.data$time)
   #lm2 <- lm(X45km ~ time + sin2pit + cos2pit , data=alo.data) ; c <-
coef(lm2)
   #noise <- resid(lm2) ; tt <- lm2$model$time
   #ao.a <- sqrt(c[3]^2 + c[4]^2) ; ao.p <- atan2(c[3],c[4]) ; B <-
c[1]
   #n2 <- nls(X45km ~ B + b*time + ao.a*sin(2*pi*time + ao.p),
data=alo.data, start=list(B=c[1], b=c[2], ao.a=ao.a, ao.p=ao.p))
   #se2 <- SE2(tt, ao.p, ao.a, noise)

   AO.a[i,1] <- ao.a ; AO.a[i,2] <- ses[3]
   AO.p[i,1] <- ao.p ; AO.p[i,2] <- ses[4]
   SAO.a[i,1] <- sao.a ; SAO.a[i,2] <- ses[5]
   SAO.p[i,1] <- sao.p ; SAO.p[i,2] <- ses[6]
}

op <- par(no.readonly = TRUE)
par(mfrow = c(2, 2))
par(mar = c(2.5, 2.5, 0.5, 0))
ymax <- 90

plot(AO.a[,1], alt, type="l", xlim=c(1,20),lwd=2, ylim=c(45,ymax),
col="red", axes=FALSE) ; box()
lines(AO.a[,1] + AO.a[,2]*1.96, alt, col="red", lt=2)
lines(AO.a[,1] - AO.a[,2]*1.96, alt, col="red", lt=2)
lines(H.aa, lt=1, col="blue",lwd=2)
lines(C.aa, lt=1, col="green",lwd=2)
# lines(S.a[,1], S.alt, pch=19, col="gray") ; points(S.a[,1],S.alt,
pch=19)
lines(H.a[,1], H.alt, col="blue", lt=2,lwd=2)
mtext( "AO Amplitude (K)", side=1, at=15.5, cex=0.8, line=-1.5)
mtext("Altitude (km)", side=2, at=68, cex=1, line=2.8)
axis(side=2, at=seq(45, 90, by=5), labels=TRUE, cex=0.5, hadj=0.7,
padj=0.4, las=1)
axis(side=1, padj=-0.8)
text(2,85, "(a)")
abline(v=0)

par(mar = c(2.5, 0, 0.5, 0.5))
plot(SAO.a[,1], alt, type="l", axes=FALSE, xlim=c(-0.5,7),
ylim=c(45,ymax), col="red",lwd=2) ; box()
abline(v=0, col="gray")
lines(SAO.a[,1], alt, lwd=2, col="red")
lines(SAO.a[,1] + SAO.a[,2]*1.96, alt, col="red", lt=2)
lines(SAO.a[,1] - SAO.a[,2]*1.96, alt, col="red", lt=2)
lines(H.sa, col="blue", lt=1,lwd=2)
```

```
lines(C.sa, col="green", lt=1,lwd=2)
# lines(S.s[,1], S.alt, pch=19, col="gray") ; points(S.s[,1],S.alt,
pch=19)
lines(H.s[,1], H.alt, col="blue", lt=2,lwd=2)
mtext("SAO Amplitude (K)", side=1, at=5, cex=0.8, line=-1.5)
axis(side=1, padj=-0.8)
text(6, 85, "(b)")
a <-
c(42.06052,51.13833,53.87608,56.18156,58.19885,61.22478,69.14986,73.472
63,75.34582,77.07493,81.2536,85,86.44093,87.73775,88.7464,89.46686,90.6
196)
A <- c(1.5,1.5,2,2.5,3,3.5,3,3,4,4.5,5,4.5,4,3.5,3,2.5,2)
lines(A,a, lwd=2)

par(mar = c(2.5, 2.5, 0, 0))
AO.p[,1] <- fixphase(AO.p[,1]) ; AO.p <- AO.p*365/(2*pi)
A.p <- AO.p[,1]
plot(A.p, alt, type="l", xlim=c(-20,190), col="red", ylim=c(45,ymax),
axes=FALSE,lwd=2) ; box()
abline(v=0, col="gray")
lines(A.p + AO.p[,2]*1.96, alt, col="red", lt=2)
lines(A.p - AO.p[,2]*1.96, alt, col="red", lt=2)
lines(H.ap, col="blue", lt=1,lwd=2)
lines(C.ap, col="green", lt=1,lwd=2)
# lines(S.a[,2], S.alt, col="gray") ; points(S.a[,2],S.alt, pch=19)
lines(H.a[,2], H.alt, col="blue", lt=2,lwd=2)
axis(side=2, at=seq(45, 90, by=5), labels=TRUE, cex=0.5, hadj=0.7,
padj=0.4, las=1)
at <- 1:14 ; at[2:13] <- months(FALSE); at[14] <- 365; at[1] <- -31
axis(side=1, at=at, tick=TRUE, labels=NA)
at2 <- 1:13 ; at2[2:13] <- months() ; at2[1] <- -15.5
axis(side=1, tick=FALSE, at=at2, padj=-1,
labels=c("D","J","F","M","A","M","J","J","A","S","O","N","D"))
mtext("AO phase (months)", side=1, at=70, cex=0.8, line=-1.5)
legend(70, 87, legend=c("ALO", "OHP", "CEL", "HALOE", "SABER"),
lty=c(1,1,1,2,1), col=c("red", "blue", "green", "blue", "black"), bg =
"white" , cex=0.8, lwd=1)
text(180, 85, "(c)")

par(mar = c(2.5, 0, 0, 0.5))
SAO.p[,1] <- fixphase(SAO.p[,1]) ; SAO.p <- SAO.p*182.5/(2*pi)
S.p <- SAO.p[,1]
plot(S.p, alt, type="l", axes=FALSE, xlim=c(0,190), ylim=c(45,ymax),
col="red",lwd=2) ; box()
lines(S.p + SAO.p[,2]*1.96, alt, col="red", lt=2)
lines(S.p - SAO.p[,2]*1.96, alt, col="red", lt=2)
lines(H.sp, col="blue", lt=1,lwd=2)
lines(C.sp, col="green", lt=1,lwd=2)
# lines(S.s[,2], S.alt, col="gray") ; points(S.s[,2],S.alt, pch=19)
lines(H.s[,2], H.alt, col="blue", lt=2,lwd=2)
mtext("SAO phase (months)", side=1, at=60, cex=0.8, line=-1.5)
axis(side=1, at=at, tick=TRUE, labels=NA)
axis(side=1, tick=FALSE, at=at2, padj=-1,
labels=c("D","J","F","M","A","M","J","J","A","S","O","N","D"))
text(180, 85, "(d)")
```

```
#########################################################################
##### THIS CODE GENERATES FIGURES 13 and 14  FROM CHAPTER 4 ######
#########################################################################

### HALOE SAO CLIMATOLOGY
library(sfsmisc)
library(fUtilities)
library(fields)

get.mean <- function(doy, r, dw=5) {
   doys <- seq(1, 365, by=1)
   doys <- c(320:365, doys, 1:45)  ## day 1 is at 47
   w <- (dw-1)/2
   sds <- matrix(NA, 365) ; j <- 1
   for (i in 1:365) {
     dds <- doys[(46+i-w+2):(46+i+w)]
       get <- matrix(FALSE, length(doy))
       for (p in 1:length(dds)) { get <- doy==dds[p] | get }
       if (sum(get)==0) { sds[j] <- NA } else { sds[j] <- mean(r[get])
}
     j <- j + 1  # ; print(sum(get))
   }
   return(sds)
}


pt2 <- "C:/Documents and Settings/Troy Wynn/My
Documents/dissertation/Dissertation/Summary And Comp/HALOE/process/"
data <- read.table(ccat(pt2,"haloe.txt"), sep="\t", header=TRUE)
date <- 1:1
for (i in 1:nrow(data)) { date[i] <- ccat(data$m[i], "/", data$d[i],
"/", data$y[i])}
t <- data$time
s2 <- sin(2*pi*t) ; c2 <- cos(2*pi*t)
s3 <- sin(3*pi*t) ; c3 <- cos(3*pi*t)
s4 <- sin(4*pi*t) ; c4 <- cos(4*pi*t)
s5 <- sin(5*pi*t) ; c5 <- cos(5*pi*t)
s6 <- sin(6*pi*t) ; c6 <- cos(6*pi*t)
s7 <- sin(7*pi*t) ; c7 <- cos(7*pi*t)
s8 <- sin(8*pi*t) ; c8 <- cos(8*pi*t)
s9 <- sin(9*pi*t) ; c9 <- cos(9*pi*t)

dat <- data.frame(data, s2, s4, c2, c4, date, s5, c5, s6, c6, s7, c7,
s3, c3, s9, c9)
dat <- dat[,-89] ## the 89th data point seems to be bad

get.sol.max <- function(lm, tt) {
   c <- coef(lm)
   ang <- atan2(c[7],c[8])
   if (ang<0) { ang <- ang + 2*pi}
   w <-  0.5752152710
   # print(ang/w)
   t <- seq(min(tt), max(tt), by=0.025)
   s2 <- c[7]*sin(w*t) ; c2 <- c[8]*cos(w*t)
   # plot(t, s2 + c2) # ; points(tt, c[3]*sin(w*tt) + c[4]*cos(w*tt))
   get <- max(s2 + c2)==(s2 + c2)
   # print(get) ; print(sum(get))
   s.max <- t[get]
   if (s.max<0) { s.max <- s.max + 2*pi/w }
   return(s.max)
}
```

```
solar.max <- "7/31/2001" ; halo.near <- "9/5/2001" ; halo.near2 <-
"2001-9-5"
H.years <- ExcelDate.to(7,31,2001):ExcelDate.to(7,31,2002)
H.years2 <- ExcelDate.from(H.years) ; get2 <- H.years2==halo.near2
H.years <- (H.years - min(H.years))/365
H.yr <- H.years[get2] ; get <- dat$date==halo.near
tt <- data$time ; tt <- tt - tt[get] + H.yr
w <- 2*pi/11
sw <- sin(w*tt) ; cw <- cos(w*tt)


H.sd <- H.a <- H.s <- H.t <- H.sol <- matrix(NA, 150,2)
H.sol2 <- H.sol

H.alt <- colnames(dat)[1:150]
for (i in 1:150) { H.alt[i] <- substr(H.alt[i],2,100) };H.alt<-
as.double(H.alt)
sao <- matrix(NA, 150, 365)

for (i in 1:150) {
   dat$y <- dat[,i]
   lm0 <- lm <- lm(y ~ time + s2 + c2 + s4 + c4 + sw + cw + doy,
data=dat)
   doy <- lm0$model$doy
   lm <- lm(y ~ time + s2 + c2 + s4 + c4 + sw + cw + s5+c5+c6+s6,
data=dat)
   c <- coef(lm)
   fit <- c[1] + c[2]*lm$model[,2] + c[3]*lm$model[,3] +
c[4]*lm$model[,4] +
      c[7]*lm$model[,7] + c[8]*lm$model[,8] + c[9]*lm$model[,9] +
c[10]*lm$model[,10]+
        c[11]*lm$model[,11] + c[12]*lm$model[,12] #+
c[13]*lm$model[,13]# + c[14]*lm$model[,14]
   deseason <- lm$model[,1] - fit
   sao[150-i+1,] <- get.mean(doy, deseason, dw=30)
   H.sd[i,1] <- sd(resid(lm))
   H.t[i,1] <- c[2] ; H.t[i,2] <- summary(lm)$coefficients[2,2]
   H.a[i,1] <- sqrt(c[3]^2+c[4]^2)
   H.a[i,2] <- atan2(c[3],c[4])
   H.s[i,1] <- sqrt(c[5]^2 + c[6]^2)
   H.s[i,2] <- atan2(c[5],c[6])

   H.sol[i,1] <- sqrt(c[7]^2+c[8]^2)
   H.sol[i,2] <- atan2(c[7],c[8])/w
   #H.sol2[i,2] <- get.sol.max(lm, lm$model$time)
   print(i) #; print(summary(lm))
}

doy.x <- c(16,45,75,105,136,166,197,228,258,289,319,349)
doy.l <- c("J", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")
x <- 1:365
x.lab=seq(min(x), max(x), by=15)
mons <- c(1, 32,60,91,121,152,182,213,244,274,305,335,365)
alt <- sort(H.alt)
x <- 1:365
par(mar = c(4, 3.5, 1, 1))
image.plot( 1:365, alt, t(sao), axes=FALSE, xlab="", ylab="",
legend.args=list(text="Amplitude", col="black", cex=1, side=4,
line=2.3))
contour( 1:365, alt, t(sao), levels=seq(-25,15,by=2), col="black",
add=TRUE, lt=0, labcex=1.2, axes=FALSE)
mtext("Altitude (km)", side=2, at=70, padj=-3.2, cex=1.1)
mtext("Month", side=1, at=182, padj=2.5, cex=1.1)
```

```
axis(side=1, at=mons, cex=0.2, padj=-0.8, las=1, labels=FALSE)
axis(side=1, at=doy.x, labels=doy.l, cex=0.2, padj=-0.8, las=1,
tick=FALSE)
axis(2, las=1, hadj=0.8)
box()



## USU SAO climatology
############################################################
#########################################################################
##
## AO SAO amp and phase
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(fUtilities)
library(fields)


get.mean <- function(doy, r, dw=5) {
   doys <- seq(1, 365, by=1)
   doys <- c(320:365, doys, 1:45)  ## day 1 is at 47
   w <- (dw-1)/2
   sds <- matrix(NA, 365) ; j <- 1
   for (i in 1:365) {
     dds <- doys[(46+i-w+2):(46+i+w)]
       get <- matrix(FALSE, length(doy))
       for (p in 1:length(dds)) { get <- doy==dds[p] | get }
       if (sum(get)==0) { sds[j] <- NA } else { sds[j] <- mean(r[get])
}
     j <- j + 1  # ; print(sum(get))
   }
   return(sds)
}


#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
s5 <- sin(5*pi*t) ; c5 <- cos(5*pi*t)
s6 <- sin(6*pi*t) ; c6 <- cos(6*pi*t)
df <- data.frame(alo, s1, s2, y=s2*0, s5, c5, s6, c6)
alt <- 45:90
n <- length(alt)
AO.a <- AO.p <- SAO.a <- SAO.p <- matrix(NA, n, 2)
sds <- matrix(NA, n)
sao <- matrix(NA, 46, 365)

for (i in 1:n) {
   h.col <- 26 + i
   df$y <- df[, h.col] ; time <- df$time
   lm0 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n + doy, data=df)
   doy <- lm0$model$doy
   lm1 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n + s5 + c5 + s6 + c6, data=df)
   cc <- coef(lm1) ; mod <- lm1$model
```

```
   deseason <- mod[,1] - (cc[1] + cc[2]*mod[,2] + cc[3]*mod[,3] +
cc[4]*mod[,4] + cc[7]*mod[,7] + cc[8]*mod[,8] + cc[9]*mod[,9]  +
                          cc[10]*mod[,10] + cc[11]*mod[,11] +
cc[12]*mod[,12] + cc[13]*mod[,13])
   sao[i,] <- get.mean(doy, deseason, dw=30)
   c <- coef(lm1) ; sds[i] <- sd(resid(lm1))
   noise <- resid(lm1) ; tt <- lm1$model$time ; sn <- lm1$model$sol.n
   ao.a <- sqrt(c[3]^2 + c[4]^2) ; ao.p <- atan2(c[3],c[4]) ; B <- c[1]
   sao.a <- sqrt(c[5]^2 + c[6]^2); sao.p <- atan2(c[5],c[6]) ; b <-
c[2]
   s.a <- sqrt(c[7]^2 + c[8]^2) ; s.p <- atan2(c[7],c[8])

}

## plot the semi-annual climatology
doy.x <- c(16,45,75,105,136,166,197,228,258,289,319,349)
doy.l <- c("J", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")
x <- 1:365
x.lab=seq(min(x), max(x), by=15)
mons <- c(1, 32,60,91,121,152,182,213,244,274,305,335,365)

alt <- 45:90
x <- 1:365
par(mar = c(4, 3.5, 1, 1))
image.plot( 1:365, alt, t(sao), axes=FALSE, xlab="", ylab="",
legend.args=list(text="Amplitude", col="black", cex=1, side=4,
line=2.3))
contour( 1:365, alt, t(sao), levels=seq(-25,15,by=2), col="black",
add=TRUE, lt=0, labcex=1.2, axes=FALSE)
mtext("Altitude (km)", side=2, at=70, padj=-3.2, cex=1.1)
mtext("Month", side=1, at=182, padj=2.5, cex=1.1)
axis(side=1, at=mons, cex=0.2, padj=-0.8, las=1, labels=FALSE)
axis(side=1, at=doy.x, labels=doy.l, cex=0.2, padj=-0.8, las=1,
tick=FALSE)
axis(2, las=1, hadj=0.8)
box()

#########################################################################
######### THIS CODE GENERATES FIGURE 1 FROM CHAPTER 5 ###########
#########################################################################

## This will do a strait regression on the alo temperatures.
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)

get.doys <- function(doy) {
#Summer - June 21 ; 172
#Autumn - September 21 ; 265
#Winter - December 21 ; 356
#Spring - March 21 ; 81

SU <- doy>=172 & doy<265
AU <- doy>=265 & doy<356
SP <- doy>=81 & doy<172
WI <- doy>=356 | doy<81

  return(c(sum(SP), sum(SU), sum(WI), sum(AU), length(doy)))

}
```

```
#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
alt <- seq(45, 90, by=1)
n <- length(alt)
df$time <- t - mean(t)
pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\Residuals\\"

get.lt <- matrix(NA, n, 3)
get.res <-
list("45","46","47","48","49","50","51","52","53","54","55","56","57","
58","59","60","61","62","63","64","65","66","67","68","69","70","71","7
2","73","74","75","76","77","78","79","80","81","82","83","84","85","86
","87","88","89","90")
par(mar = c(4, 6, 2, 2))
m <- t(matrix(c(1,1,1,1,1, 1, 2), 7, 1))
np <- matrix(NA, 46)
bp <- matrix(NA, 46)
doys <- matrix(NA, 46, 5)
for (i in 1:n) {

   h.col <- 26 + i
   df$y <- df[, h.col]
   lm2 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n + doy, data=df)
   doy. <- lm2$model$doy ; doys[i,] <- get.doys(doy.)
   lm1 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n, data=df)
   get.res[[i]] <- resid(lm1)
   t <- lm1$model$time ; np[i] <- length(t)
   r <- resid(lm1) ; r2 <- abs(r)
   lmr <- lm(r2 ~ t)
   print(summary(lmr)) ; s <- round(bptest(lm1)$p.value,3)[[1]] ; bp[i]
<- bptest(lm1)$p.value

   layout(m) ; pch=1
   par(mar = c(6, 5, 1, 0), cex.axis=1.5)
   plot(t, r, ylab="", xlab="", axes=FALSE, type="n") ; p <-
predict(lmr)
   lines(t, p , col="black", lwd=2)
   abline(h=0, col="gray")
   points(t, r, pch=pch)
   box()
   mtext("Temperature (K)", side=2, at=0, padj=-3.5, cex=1)
   mtext("Time (years)", side=1, at=0, padj=2.5, cex=1)
   mtext(ccat(44+i, " km"), side=3, at=0, padj=2.5, cex=1)
   axis(side=2, labels=TRUE, cex=1.5, padj=0.3, hadj=0.8, las=1)
   axis(side=1, labels=TRUE, cex=1.5, padj=0, hadj=0.3, las=1)  #  x-
axis
   par(mar=c(6,0,1,1))
   boxplot(r, axes=FALSE, pch=pch, cex=1.2, notch=TRUE)
   mtext(ccat("n: ", length(r)), side=1, at=1, padj=1, cex=0.8)
   mtext(ccat("p-value: ", s), side=1, at=1, padj=3, cex=0.8)

   #dev.copy(pdf, ccat(pth,ccat("X",44+i, "km.pdf")))
   dev.copy2pdf(file=ccat(pth,ccat("X",44+i, "km.pdf")))
   #dev.off()
```

```
   print(i)
   #readline("Enter...")

}
# dev.off()

par(mar = c(3, 3.3, 1, 1))
plot(doys[,5], alt, pch=20, axes=FALSE, xlab="", ylab="",
xlim=c(40,600))
box()
points(doys[,1], alt, pch=20, col="green")
points(doys[,2], alt, pch=20, col="orange")
points(doys[,3], alt, pch=20, col="light blue")
points(doys[,4], alt, pch=20, col="red")
mtext("Altitude (km)", side=2, at=70, padj=-3.5, cex=1)
mtext("Number of Data Points", side=1, at=350, padj=2.5, cex=1)
axis(side=2, labels=TRUE, cex=1.5, padj=0.3, hadj=0.8, las=1)
axis(side=1, labels=TRUE, cex=1.5, padj=-0.8, hadj=0.4, las=1)  #  x-
axis
legend(300, 70, c("SP", "SU", "WI", "AU", "ALL"), pch=20,
col=c("green","orange","light blue", "red", "black"), bg = 'white' ,
cex=1)

##########################################################################
######### THIS CODE GENERATES FIGURE 2 FROM CHAPTER 4 ##########
##########################################################################

rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(fUtilities)
library(fields)
OHP <- function() {

  sd <- matrix(NA, 13, 12)
  sd[1,] <- c(9,6.9,5.4,3.6,2.5,2.2,2.6,2.1,2.2,2.8,6.5,9.4)
  sd[2,] <- c(7.2,6.6,4.4,3.3,2.5,2.3,2.4,2.2,2.3,2.6,6.4,8.2)
  sd[3,] <- c(6.8,6.7,4.1,2.6,2.4,2,2.2,2.5,2,2.5,6.6,7.4)
  sd[4,] <- c(7.3,6.9,4,2.8,2.4,2.1,2.3,2.7,2.4,2.9,6,7.2)
  sd[5,] <- c(7.9,7.5,4.3,2.7,2.7,2.4,2.7,2.9,2.7,3.1,5.9,7.8)
  sd[6,] <- c(8.3,7.6,4.8,2.4,2.7,2.4,3,3.6,2.7,3.9,6.5,8.9)
  sd[7,] <- c(9.1,8,5.1,3.4,3,3,3.3,3.9,3.6,4.4,7.2,9.6)
  sd[8,] <- c(9.1,7.9,5.6,3.9,3.8,4.1,3.8,4.8,4.3,5.6,7.5,10.6)
  sd[9,] <- c(10.7,8.9,6.3,5.4,4.5,4.7,4.7,5.7,5.2,7.3,8.2,11.5)
  sd[10,] <- c(10.3,10.7,7.5,5.8,4.6,5,6.5,6.5,6.2,8.5,9.3,10.4)
  sd[11,] <- c(10.1,11.4,7.7,6.2,6.3,7.2,7.3,7.8,7.1,9.7,10.3,10.9)
  sd[12,] <- c(12.1,11.4,8.9,7.7,9.1,8.2,8.2,8.4,8.8,8.9,11.3,11.4)
  sd[13,] <- c(11.5,9.9,7.5,8.2,9.2,8.7,9.1,7.4,7.9,7.5,11.6,11.2)
  return(sd)
}


ohp <- OHP()

cols <- colorRamp(c("purple", "blue", "green", "yellow", "red"),
space="rgb")
l <- seq(0,22, by=1) ; ng <- length(l)
min <- min(l) ;  max <- max(l-min)
colors.r <- rgb(cols( (l-min)/max ), max=255)
temps <- sort(ohp) ;  temps <- temps[!duplicated(temps)] ; temps <-
(temps - min)/max
colors.r <- rgb(cols(temps), max=255)
```

```
alt <- c(45,48,51,54,57,60,63,66,69,72,75,78,81)
doy.x <- seq(0.5, 11.5, by=1)
doy.l <- c("J", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")
x <- 1:12
par(mar = c(4, 3.5, 1, 1))
image.plot( x, alt, t(ohp), col=colors.r,axes=FALSE, xlab="", ylab="",
            legend.mar=5, legend.args=list(text="sd (K)", col="black",
cex=1, side=4, line=2))
contour( x, alt, t(ohp), levels=seq(0,12,by=1), col="black", add=TRUE,
lt=0, labcex=1.2)
mtext("Altitude (km)", side=2, at=63, padj=-3.5, cex=0.9)
mtext("Month", side=1, at=182, padj=2.5, cex=0.9)
axis(side=1,at=doy.x + 0.5, labels=doy.l, cex=0.2, padj=-0.8, las=1,
tick=FALSE)
axis(side=1,at=0:13 - 0.5, labels=NA)
mtext("Month", side=1, at=6.5, padj=2.5)
axis(2)
box()

##########################################################################
######### THIS CODE GENERATES FIGURE 2a FROM CHAPTER 4 ##########
##########################################################################

rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(fUtilities)
library(fields)

get.sds <- function(doy, r, dw=15) {
   doys <- seq(1, 365, by=1)
   doys <- c(doys, doys, doys)
   w <- (dw-1)/2
   sds <- matrix(NA, 365) ; j <- 1
   for (i in 1:365) {
     dds <- doys[(i-w+366):(i+w+366)]
       get <- matrix(FALSE, length(doy))
       for (p in 1:length(dds)) { get <- doy==dds[p] | get }
       if (sum(get)==0) { sds[j] <- NA } else { sds[j] <- sd(r[get]) }
     j <- j + 1  # ; print(sum(get))
   }
   return(sds)
}

mids <- function(l) {
   l <- sort(l)
   ll <- 1:(length(l)-1)
   for (i in 1:(length(l)-1)) {
      ll[i] <- (l[i] + l[i+1])/2
   }
   return(ll)
}

levelize <- function(z, l) {
   ll <- sort(l) ; n <- length(ll)
   m <- mids(ll)
   for (i in 1:(n-1)) {
      get <- z>=ll[i] & z<ll[i+1]
      z[get] <- m[i]
   }
   i <- i + 1
   get <- z>=ll[i]
   z[get] <- ll[i]
```

```
   return(z)
}

#z <- matrix(rnorm(100)*20, 10,10)
#l <- seq(-53,32, by=4)
#zz <- levelize(z,l)


sds.n <- function(dw) {
   w <- (dw-1)/2
   n <- length((w):(364-w))
   return(n)
}

interp3d <- function(x, y, z, byx, byy) {

   nx <- length(x)
   ny <- length(y)

   # assuming nx > ny ;  also assuming row is x and that col is y
   nny <- seq(min(y), max(y), length=nx)
   nz <- matrix(NA, nx, nx)
   for (i in 1:nx) {
      nz[i,] <- approx(y, z[i,], nny)$y
   }

   sd.new <- linearInterp( x, nny, nz, xo=seq(min(x), max(x),
length=2*nx),yo=seq(min(alt), max(alt), length=2*nx))
image( x, nny, nz, col=terrain.colors(100),axes=FALSE)

}


#  doy <- 0:364
#  r <- rnorm(length(doy), 0, 12)
#  sds <- get.sds(doy,r)

#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
alt <- seq(45, 90, by=1) ; n <- length(alt)
dw <- 31
df$time <- t - mean(t)
pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\Residuals\\"
sds <- matrix(NA, 365, length(alt))
err <- sds

for (i in 1:n) {

   h.col <- 26 + i
   df$y <- df[, h.col]
   lm1 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ doy, data=df)
   mod <- lm1$model ; doy <- mod$doy
   lm2 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 +
s2, data=mod)
   r <- resid(lm2) ; t <- mod$time
```

```
   sdi <- get.sds(doy, r, dw=31)
   sds[,i] = sdi
   print(i)
}


doy.x <- c(16,45,75,105,136,166,197,228,258,289,319,349)
doy.l <- c("J", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")
x <- 1:365
x.lab=seq(min(x), max(x), by=15)
mons <- c(1, 32,60,91,121,152,182,213,244,274,305,335,365)
par(mar = c(4, 3.5, 1, 1))

cols <- colorRamp(c("purple", "blue", "green", "yellow", "red"),
space="rgb")
l <- seq(0,22, by=1) ; ng <- length(l)
min <- min(l) ;  max <- max(l-min)
z <- levelize(sds,l)
colors.r <- rgb(cols( (l-min)/max ), max=255)
image.plot( x, 45:90, z, col=colors.r, axes=FALSE, xlab="", ylab="",
            legend.mar=5, legend.args=list(text="sd (K)", col="black",
cex=1, side=4, line=2))
contour( x, 45:90, z, levels=l, col="black", add=TRUE, lt=0,
labcex=1.2)
mtext("Altitude (km)", side=2, at=69, padj=-3.5, cex=0.9)
mtext("Month", side=1, at=182, padj=2.5, cex=0.9)
axis(side=1,at=doy.x, labels=doy.l, cex=0.2, padj=-0.8, las=1,
tick=FALSE)
axis(side=1,at=mons, labels=NA)
axis(2)
box()
```

```
#########################################################################
######### THIS CODE GENERATES FIGURE 1 FROM CHAPTER 4 ##########
######### AND THE PLOTS IN APPENDIX F  ##############################
#########################################################################

## This will do a strait regression on the alo temperatures.
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(lmtest)

get.doys <- function(doy) {
#Summer - June 21 ; 172
#Autumn - September 21 ; 265
#Winter - December 21 ; 356
#Spring - March 21 ; 81

SU <- doy>=172 & doy<265
AU <- doy>=265 & doy<356
SP <- doy>=81 & doy<172
WI <- doy>=356 | doy<81

  return(c(sum(SP), sum(SU), sum(WI), sum(AU), length(doy)))

}


#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
alt <- seq(45, 90, by=1)
n <- length(alt)
df$time <- t - mean(t)
pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\Residuals\\"

get.lt <- matrix(NA, n, 3)
get.res <-
list("45","46","47","48","49","50","51","52","53","54","55","56","57","
58","59","60","61","62","63","64","65","66","67","68","69","70","71","7
2","73","74","75","76","77","78","79","80","81","82","83","84","85","86
","87","88","89","90")
par(mar = c(4, 6, 2, 2))
m <- t(matrix(c(1,1,1,1,1, 1, 2), 7, 1))
np <- matrix(NA, 46)
bp <- matrix(NA, 46)
doys <- matrix(NA, 46, 5)
for (i in 1:n) {

   h.col <- 26 + i
   df$y <- df[, h.col]
   lm2 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n + doy, data=df)
   doy. <- lm2$model$doy ; doys[i,] <- get.doys(doy.)
   lm1 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n, data=df)
   get.res[[i]] <- resid(lm1)
   t <- lm1$model$time ; np[i] <- length(t)
```

```
   r <- resid(lm1) ; r2 <- abs(r)
   lmr <- lm(r2 ~ t)
   print(summary(lmr)) ; s <- round(bptest(lm1)$p.value,3)[[1]] ; bp[i]
<- bptest(lm1)$p.value

   layout(m) ; pch=1
   par(mar = c(6, 5, 1, 0), cex.axis=1.5)
   plot(t, r, ylab="", xlab="", axes=FALSE, type="n") ; p <-
predict(lmr)
   lines(t, p , col="black", lwd=2)
   abline(h=0, col="gray")
   points(t, r, pch=pch)
   box()
   mtext("Temperature (K)", side=2, at=0, padj=-3.5, cex=1)
   mtext("Time (years)", side=1, at=0, padj=2.5, cex=1)
   mtext(ccat(44+i, " km"), side=3, at=0, padj=2.5, cex=1)
   axis(side=2, labels=TRUE, cex=1.5, padj=0.3, hadj=0.8, las=1)
   axis(side=1, labels=TRUE, cex=1.5, padj=0, hadj=0.3, las=1)  #  x-
axis
   par(mar=c(6,0,1,1))
   boxplot(r, axes=FALSE, pch=pch, cex=1.2, notch=TRUE)
   mtext(ccat("n: ", length(r)), side=1, at=1, padj=1, cex=0.8)
   mtext(ccat("p-value: ", s), side=1, at=1, padj=3, cex=0.8)

   #dev.copy(pdf, ccat(pth,ccat("X",44+i, "km.pdf")))
   dev.copy2pdf(file=ccat(pth,ccat("X",44+i, "km.pdf")))
   #dev.off()
   print(i)
   #readline("Enter...")

}
# dev.off()

par(mar = c(3, 3.3, 1, 1))
plot(doys[,5], alt, pch=20, axes=FALSE, xlab="", ylab="",
xlim=c(40,600))
box()
points(doys[,1], alt, pch=20, col="green")
points(doys[,2], alt, pch=20, col="orange")
points(doys[,3], alt, pch=20, col="light blue")
points(doys[,4], alt, pch=20, col="red")
mtext("Altitude (km)", side=2, at=70, padj=-3.5, cex=1)
mtext("Number of Data Points", side=1, at=350, padj=2.5, cex=1)
axis(side=2, labels=TRUE, cex=1.5, padj=0.3, hadj=0.8, las=1)
axis(side=1, labels=TRUE, cex=1.5, padj=-0.8, hadj=0.4, las=1)  #  x-
axis
legend(300, 70, c("SP", "SU", "WI", "AU", "ALL"), pch=20,
col=c("green","orange","light blue", "red", "black"), bg = 'white' ,
cex=1)

#########################################################################
####### THIS CODE GENERATES FIGURES 5a and b FROM CHAPTER 4 ########
#########################################################################

##  Solar variation of the solar SEs
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(fUtilities)
library(fields)

get.doy <- function(i, doy, dw=121) {
   doys <- c(1:365, 1:365, 1:365)
```

```
   w <- (dw-1)/2  # w is typically going to be 91 days
   sds <- matrix(NA, 365) ; j <- 1
   dds <- doys[(i-w+365):(i+w+365)]
   dL <- dds[1] ; dH <- dds[length(dds)]
   if (dL<dH) { get <- doy>dL & doy<dH
   } else { get <- doy>dL | doy<dH }
   return(get)
}

alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
alt <- 45:90
n <- length(alt)
df$time <- t - mean(t)

get <- matrix(NA, 46, 4)
SEs <- TEs <- matrix(NA, 46, 365)

t0 <- max(t - min(t))/4
dn.min <- 0 ; dn.max <- t0
j <- 1

for (i in 1:n) {
   h.col <- 26 + i
   df$y <- df[, h.col]
   lm2 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + sol.n +
doy, data=df)
   data <- lm2$model ; doy. <- data$doy


   for (k in 1:365) {
      gg <- get.doy(k, doy.)
        dat <- lm2$model[gg,]
        lm.g <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit +
sol.n , data=dat)
        SEs[i,k] <- 1 - summary(lm.g)$coefficients[7,4]
        TEs[i,k] <- summary(lm.g)$coefficients[7,1]
      }

   print(i)

}

pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\phs\\"
write.table(SEs, ccat(pth, "solar_noise_SEs.txt"), quote=FALSE,
sep="\t", col.names=FALSE, row.names=FALSE)
write.table(TEs, ccat(pth, "solar_sig_TEs"), quote=FALSE, sep="\t",
col.names=FALSE, row.names=FALSE)

pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\phs\\"
SEs <- as.matrix(read.table(ccat(pth, "solar_noise_SEs.txt"),
header=FALSE, sep="\t"))
TEs <- as.matrix(read.table(ccat(pth, "solar_sig_TEs"), header=FALSE,
sep="\t"))

par(mar = c(4, 3.5, 1, 1))
```

```
image.plot( 1:365, 45:90, t(SEs), ylab="", xlab="", axes=FALSE,
legend.mar=5, legend.args=list(text="(1 - pvalue)", col="black", cex=1,
side=4, line=2))
contour(1:365, 45:90, t(SEs), col="black", add=TRUE, lt=0, labcex=1.2)
doy.x <- c(16,45,75,105,136,166,197,228,258,289,319,349)
doy.l <- c("J", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")
x <- 1:365
x.lab=seq(min(x), max(x), by=15)
mons <- c(1, 32,60,91,121,152,182,213,244,274,305,335,365)
mtext("Altitude (km)", side=2, at=69, padj=-3.5, cex=1)
mtext("Month", side=1, at=182, padj=2.5, cex=1)
axis(side=1,at=doy.x, labels=doy.l, cex=0.2, padj=-0.8, las=1,
tick=FALSE)
axis(side=1,at=mons, labels=NA)
mtext("(b)", side=1, at=390, padj=-27.5, cex=1.5)
axis(2)
box()



pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\phs\\"
SEs <- as.matrix(read.table(ccat(pth, "solar_noise_SEs.txt"),
header=FALSE, sep="\t"))
TEs <- as.matrix(read.table(ccat(pth, "solar_sig_TEs"), header=FALSE,
sep="\t"))

par(mar = c(4, 3.5, 1, 1))
image.plot( 1:365, 45:90, t(TEs), ylab="", xlab="", axes=FALSE,
legend.mar=5, legend.args=list(text="Kelvin", col="black", cex=1,
side=4, line=2))
contour(1:365, 45:90, t(TEs), col="black", add=TRUE, lt=0, labcex=1.2)
doy.x <- c(16,45,75,105,136,166,197,228,258,289,319,349)
doy.l <- c("J", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")
x <- 1:365
x.lab=seq(min(x), max(x), by=15)
mons <- c(1, 32,60,91,121,152,182,213,244,274,305,335,365)
mtext("Altitude (km)", side=2, at=69, padj=-3.5, cex=1)
mtext("Month", side=1, at=182, padj=2.5, cex=1)
axis(side=1,at=doy.x, labels=doy.l, cex=0.2, padj=-0.8, las=1,
tick=FALSE)
axis(side=1,at=mons, labels=NA)
axis(2)
mtext("(a)", side=1, at=390, padj=-27.5, cex=1.5)
box()

#########################################################################
######### THIS CODE GENERATES FIGURE 7 FROM CHAPTER 4 ###########
#########################################################################

rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(fUtilities)
library(fields)

pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\"

advance <- function(j, doy) {
   i <- j+366 -1
   d <- matrix(1:365, 3*365)
   l <- i-60 ; h <- i + 60
```

```
   dl <- d[l] ; dh <- d[h]
   if (dl>dh) { get <- doy>=dl | doy<dh
   } else { get <- doy>=dl & doy<dh }
   return(get)
}

smooth <- function(z) {
   r <- nrow(z)
   c <- ncol(z)
   copy <- as.matrix(z)
   for (i in 2:(r-1)) {
      for (k in 2:(c-1)) {
           rr <- (i-1):(i+1)
             kk <- (k-1):(k+1)
           vals <- copy[rr, kk]
           z[i,k] <- mean(vals)
         }
      }
      return(z)
}

#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
alt <- seq(45, 90, by=1)
n <- length(alt)
cors <- matrix(NA, 46, 5)
cors2 <- matrix(NA, 46, 365)
n2 <- matrix(NA, 46, 365)

for (i in 1:n) {

   h.col <- 26 + i
   df$y <- df[, h.col]
   lm2 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n + doy + days, data=df)
   doy. <- lm2$model$doy ; days <- lm2$model$days

   SU <- doy.>=172 & doy.<265
   AU <- doy.>=265 & doy.<356
   SP <- doy.>=81 & doy.<172
   WI <- doy.>=356 | doy.<81

   lm1 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + s1 + s2
+ sol.n, data=df)
   r <- resid(lm1) ; t <- lm1$model$time
   for (j in 1:365) {
        WND <- advance(j, doy.)
        rr <- r[WND] ; ddays <- days[WND]
        cors2[i,j] <- peel.it(rr, ddays)$ph ; print(j) ; n2[i,j] <-
length(rr)
   }
   ph <- peel.it(r, days)
   ph.su <- peel.it(r[SU], days[SU])$ph
   ph.au <- peel.it(r[AU], days[AU])$ph
   ph.sp <- peel.it(r[SP], days[SP])$ph
   ph.wi <- peel.it(r[WI], days[WI])$ph
```

```
   cors[i,1] <- ph$ph ; cors[i,2] <- ph.su ; cors[i,3] <- ph.au ;
cors[i,4] <- ph.sp ; cors[i,5] <- ph.wi
   print(i)
}

write.table(cors2, ccat(pth, "cors2.txt"), quote=FALSE, sep="\t",
col.names=FALSE, row.names=FALSE)
write.table(cors, ccat(pth, "cors.txt"), quote=FALSE, sep="\t",
col.names=FALSE, row.names=FALSE)
write.table(n2, ccat(pth, "n2.txt"), quote=FALSE, sep="\t",
col.names=FALSE, row.names=FALSE)

alt <- 45:90
pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\"
cors <- read.table(ccat(pth,"cors.txt"), sep="\t", header=FALSE)
par(mar = c(4, 4, 1, 1))

plot(cors[,1], alt, pch=20, xlim=c(0,0.8), ylab="", xlab="",
axes=FALSE)
abline(v=0, col="gray")
abline(v=c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8, 0.9), h=seq(45,90,by=5),
col="dark gray", lt=2)
points(cors[,1], alt, pch=20)
box()
mtext("Altitude (km)", side=2, at=65, padj=-4, cex=1)
axis(side=2, at=seq(45, 90, by=5), labels=TRUE, cex=0.5, padj=0.5,
las=1)
axis(side=1, labels=TRUE, cex=0.2, padj=-0.3, las=1)
mtext("Correlation Coefficient", side=1, at=0.4, padj=3, cex=1)


rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
library(fUtilities)
library(fields)


smooth <- function(z) {
   r <- nrow(z)
   c <- ncol(z)
   copy <- as.matrix(z)
   for (i in 2:(r-1)) {
      for (k in 2:(c-1)) {
           rr <- (i-1):(i+1)
             kk <- (k-1):(k+1)
           vals <- copy[rr, kk]
           z[i,k] <- mean(vals)
         }
      }
      return(z)
}

pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\"
cols <- colorRamp(c("purple", "blue", "green", "yellow", "red"),
space="rgb")
colors.r <- rgb(cols(1:9.5/10), max=255)


doy.x <- c(16,45,75,105,136,166,197,228,258,289,319,349)
doy.l <- c("J", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")
```

```
x <- 1:365
x.lab=seq(min(x), max(x), by=15)
mons <- c(1, 32,60,91,121,152,182,213,244,274,305,335,365)

alt <- 45:90
x <- 1:365
par(mar = c(4, 3.5, 1, 1))
cors2 <- t(read.table(ccat(pth,"cors2.txt"), sep="\t", header=FALSE))
image.plot( 1:365, alt, cors2, axes=FALSE, xlab="", ylab="",
legend.args=list(text="Correlation", col="black", cex=1, side=4,
line=2.3))
contour( 1:365, alt,cors2, levels=seq(0,0.95,by=0.05), col="black",
add=TRUE, lt=0, labcex=1.2, axes=FALSE)
mtext("Altitude (km)", side=2, at=70, padj=-3.2, cex=1.1)
mtext("Day of year", side=1, at=182, padj=2.5, cex=1.1)
axis(side=1, at=mons, cex=0.2, padj=-0.8, las=1, labels=FALSE)
axis(side=1, at=doy.x, labels=doy.l, cex=0.2, padj=-0.8, las=1,
tick=FALSE)
axis(2, las=1, hadj=0.8)
box()

pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\"
cors2 <- t(read.table(ccat(pth,"cors2.txt"), sep="\t", header=FALSE))

#########################################################################
######### THIS CODE GENERATES FIGURE 8 FROM CHAPTER 4 ###########
#########################################################################

## This will do a strait regression on the alo temperatures.
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)

#alo <- alo.data  # purify.alo()  ## this will remove data points
greater than 3 standard deviation from the mean; up to and including 90
km
alo <- purify.alo()
t <- alo$time  ;  t <- t - (max(t) - min(t))/2
w <-  0.5752152710  # T ~ 10.9 yrs
s1 <- sin(w*t) ; s2 <- cos(w*t)
df <- data.frame(alo, s1, s2, y=s2*0)
alt <- 45:90
n <- length(alt)
df$time <- t - mean(t)

sd1 <- sd2 <- sd3 <- sd4 <- matrix(NA, 46)
t0 <- max(t - min(t))/4
dn.min <- 0 ; dn.max <- t0

j <- 1
for (i in 1:n) {
   h.col <- 26 + i
   df$y <- df[, h.col]
   lm2 <- lm(y ~ time + sin2pit + cos2pit + sin4pit + cos4pit + t2 + t3
+t4 +t5, data=df)
   tt <- lm2$model$time ; tt <- tt - min(tt)

   get1 <- tt<t0 ; get2 <- tt>=t0 & tt<(2*t0) ; get3 <- tt>=(2*t0) &
tt<(3*t0) ; get4 <- tt>=3*t0
   r <- resid(lm2)
   sd1[j] <- sd(r[get1]) ; sd2[j] <- sd(r[get2]) ; sd3[j] <-
sd(r[get3]) ; sd4[j] <- sd(r[get4]) ; j <- j + 1
```

```
   print(i)
}

par(mar=c(5,4,1,1))
plot(sd1, alt, type="l", xlab="",ylab="", lwd=2, xlim=c(3,18),
axes=FALSE)
lines(sd2, alt, lwd=2, lty=2) ; points(sd2, alt, pch=19, cex=0.8)
lines(sd3,alt, lwd=2, lty=2)
lines(sd4, alt, lwd=2) ; points(sd4, alt, pch=22, bg="black", cex=0.8)
box()

mtext("Altitude (km)", side=2, at=67, padj=-3.4, cex=1)
mtext("Standard deviation", side=1, at=10, padj=2.5, cex=1)
axis(side=2, at=seq(45, 90, by=5), labels=TRUE, cex=0.5, padj=0.3,
hadj=0.8, las=1)
axis(side=1, at=seq(4,18, by=2), labels=TRUE, cex=0.2, padj=-0.5,
las=1)  #  x-axis
legend(10, 55, c("0 to 2.5 years", "2.5 to 5 years", "5 to 7.5 years",
"7.5 to 10 years"), pch=c(NA,19,NA,22), pt.bg="black", lty=c(1,2,2, 1),
cex = 0.9, lwd=2, bty="n")

##########################################################################
######### THIS CODE GENERATES FIGURE 9 FROM CHAPTER 4 ##########
##########################################################################

rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
source("C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Local Poly\\fn_localP1.txt")

pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Data\\Solar data\\ALL\\"
dat <- read.table(ccat(pth, "MgII_all.txt"), header=TRUE)
t <- dat$time
gt <- t>-6 & t<13
dat <- dat[gt,]

t <- dat$time
y <- dat$time2
mg <- dat$MgII.s
t0 <- 0
t1 <- 2.480148
t2 <- 2*t1
t3 <- 3*t1
t4 <- 4*t1

par(mar = c(3, 4, 1.5, 4))
plot(t, mg, type="l", lwd=2, xlab="", ylab="", axes=FALSE,
ylim=c(0.255, 0.285))
abline(v=c(t0,t1,t2,t3,t4))
mtext("Mg II values", side=2, at=0.27, cex=0.9, line=2.5)
axis(side=2, labels=TRUE, cex=0.5, padj=0)
mtext("year", side=1, at=4, line=1.7)

g <- ((1:length(y))%%365)==0
yy <- round(y[g],0)
at <- t[g]
axis(side=1, at=at, labels=yy, padj=-0.5)
box()
text(1.24, 0.257, "I")
text(3.7,  0.257, "II")
text(6.2,  0.257, "III")
```

```
text(8.68, 0.257, "IV")

mx <- max(mg)
mn <- min(mg)
md <- (mx+mn)/2 ; d <- mx-mn
x80 <- l80 <- c(10.784889, 14.485293, 16.215352, 10.496835)
t80 <- c(1.24, 3.7, 6.2, 8.68)
mx80 <- max(x80); mn80 <- min(x80)
x80 <- x80 - min(x80) ; x80 <- x80/max(x80)
x80 <- 0.8*x80*(mx-mn) ; x80 <- x80 + mn
lt <- lm(x80 ~ l80) ; cc <- coef(lt)
x <- (4:10)*2 ; y <- cc[2]*x + cc[1]
lines(t80, x80, pch=22, type="b", bg="black", lt=2, lwd=2)
axis(side=4, labels=x, at=y, cex=0.5, padj=-1)
mtext("Standard Deviation
   80 km", side=4, at=0.27, cex=0.8, line=2.5)
```

```
#############################################################################
######### THIS CODE GENERATES FIGURE 2 IN APPENDIX E ##########
#############################################################################

rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)

pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\phs\\"
largest <- function(x,y) { if (max(x)>max(y)) { return(max(x)) } else
{return(max(y))} }

############################
#  READ THE DATA
############################

phs <- seq(0, 0.95, length=20) ; phs <- phs[-c(1,2)]
sds2 <- sds <- sds3 <- matrix(NA, length(phs))
for (j in 1:length(phs)) {
      ph <- phs[j]
      file.name <- ccat(pth,"ng_",ph,"_.txt")
      file.name2 <- ccat(pth,"alo_",ph,"_.txt")
      file.name3 <- ccat(pth,"1dg_",ph,"_.txt")
      data <- read.table(file.name, sep="\t", header=FALSE)[,1]
      data2 <- read.table(file.name2, sep="\t", header=FALSE)[,1]
      data3 <- read.table(file.name3, sep="\t", header=FALSE)[,1]
      sds[j] <- sd(data) ; sds2[j] <- sd(data2) ; sds3[j] <- sd(data3)
}
p2 <- phs^2 ; p3 <- phs^3
lm1 <- lm(sds ~ phs + p2 + p3)
lm2 <- lm(sds2 ~ phs + p2 + p3)
lm3 <- lm(sds3 ~ phs + p2 + p3)

par(mar = c(4, 4.3, 1, 1))
plot(phs, sds, ylim=c(0,0.065), type="n", axes=FALSE, xlab="", ylab="")
points(phs, sds, pch=19)
points(phs, sds3, pch=25)
points(phs, sds2, lt=2, lwd=2, pch=22, col="gray")
lines(phs, predict(lm1))
lines(phs, predict(lm2), col="gray")
lines(phs, predict(lm3), lt=2)
legend(0.2, 0.02, c("No gaps", "USU con", "USU gaps"), pch=c(19,25,22),
col=c("black","black","gray"), bg = 'white' , cex=1)
mtext("Correlation Coefficient", side=1, padj=3, cex=1)
mtext("Standard Deviation", side=2, padj=-4.5, cex=1)
axis(side=2, labels=TRUE, cex=1.5, padj=0.3, hadj=0.8, las=1)
axis(side=1, at=seq(0.1,0.95,by=0.05), labels=TRUE, cex=1.5, padj=-0.8,
hadj=0.4, las=1)  #  x-axis
box()
```

```
#########################################################################
######### THIS CODE GENERATES FIGURES AT THE END OF ##########
############################## APPENDIX E ############################
#########################################################################
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
randomize()



pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\phs\\"

norm <- function(x, m. , sd.) {
   g <- 1/sqrt(2*pi*sd.^2)*exp(-(x-m.)^2/(2*sd.^2))
   return(g)
}

largest <- function(x,y) { if (max(x)>max(y)) { return(max(x)) } else
{return(max(y))} }

mid <- function(x, y=NA) {
   if (is.na(y[1])) {
      d <- density(x)
        x <- d$x ; y <- d$y
   }
   max <- max(y)/1.5
   get <- y>=max
   xx <- x[get] ; yy <- y[get]
   l1 <- xx[1] ; l2 <- xx[length(xx)]
   return( (l1+l2)/2)
}


############################
# WITH  USU DATA GAPS
############################
while (TRUE) {
            phs <- seq(0, 0.95, length=20) ; phs <- phs[-1]
            for (j in 1:length(phs)) {
                  get <- matrix(NA, 20) ; ng <- length(get)
                  ph <- phs[j]
                  for (i in 1:ng) {
                        days <- alo.data$days ; nn <- length(days)
                        n <- 10000 ; sds <- 1
                        r <- rnorm(n, 0, sds)
                        ar <- filter(r, ph, method="recursive")[(n-nn-
4000):n] ## this generates autocorrelated noise. earlier data points
removed to enhance memory in those that remain.
                        ar <- ar[days]
                        get[i] <- peel.it(ar, days)$ph
                  }

                  par(mar = c(4, 3, 1, 1))
                  file.name <- ccat(pth,"alo_",ph,"_.txt")
                  write.table(get, file.name, append=TRUE, quote=FALSE,
sep="\t", col.names=FALSE, row.names=FALSE)
                  if (TRUE) {
                        data <- read.table(file.name, sep="\t",
header=FALSE)[,1]
                        den <- density(data) ; m. <- mean(data) ; sd.
<- sd(data) ; x.m <- den$x[max(den$y)==den$y]
```

```
                        x <- den$x ; y <- den$y ;  nnorm <-norm(x, x.m,
sd.)
                        print(length(data))
                        plot(x, y, type="n", ylab="", axes=FALSE,
xlab="", ylim=c(0,largest(nnorm, den$y)))
                        abline(v=ph, col="dark gray")
                        lines(x, y, lt=2, lwd=2)
                        lines(x, nnorm , lt=1, lwd=2)
                        mtext("Correlation Coefficient", side=1,
padj=3, cex=1)
                        axis(side=2, labels=TRUE, cex=1.5, padj=0.3,
hadj=0.8, las=1)
                        axis(side=1, labels=TRUE, cex=1.5, padj=-0.8,
hadj=0.4, las=1)  #  x-axis
                        box()
                        xx <- (max(x) - min(x))/10 + min(x)
                        yy <- max(nnorm) - (max(nnorm) - min(nnorm))/10
                        text(xx,yy, substitute(phi*" = "*ph,
list(ph=ph)), cex=1.5)
                        fn <- ccat(pth,"alo_",ph,"_.pdf")
                        dev.copy2pdf(file=fn)
                  }
                  print(j)
            }
}


#########################################################
#########################################################
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
randomize()



pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\phs\\"

norm <- function(x, m. , sd.) {
   g <- 1/sqrt(2*pi*sd.^2)*exp(-(x-m.)^2/(2*sd.^2))
   return(g)
}

largest <- function(x,y) { if (max(x)>max(y)) { return(max(x)) } else
{return(max(y))} }

mid <- function(x, y=NA) {
   if (is.na(y[1])) {
      d <- density(x)
        x <- d$x ; y <- d$y
   }
   max <- max(y)/1.5
   get <- y>=max
   xx <- x[get] ; yy <- y[get]
   l1 <- xx[1] ; l2 <- xx[length(xx)]
   return( (l1+l2)/2)
}

while (TRUE) {

            pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\phs\\"
```

```
            norm <- function(x, m. , sd.) {
               g <- 1/sqrt(2*pi*sd.^2)*exp(-(x-m.)^2/(2*sd.^2))
               return(g)
            }

            largest <- function(x,y) { if (max(x)>max(y)) {
return(max(x)) } else {return(max(y))} }

            ############################
            # WITHOUT DATA GAPS
            ############################
            phs <- seq(0, 0.95, length=20) ; phs <- phs[-1]
            for (j in 1:length(phs)) {
                  get <- matrix(NA, 20) ; ng <- length(get)
                  ph <- phs[j]
                  for (i in 1:ng) {
                        days <- alo.data$days ; nn <- length(days)
                        n <- 10000 ; sds <- 1
                        r <- rnorm(n, 0, sds)
                        ar <- filter(r, ph, method="recursive")[(n-
nn+1):n] ## this generates autocorrelated noise. earlier data points
removed to enhance memory in those that remain.
                        get[i] <- peel.it(ar, 1:length(ar))$ph
                  }

                  par(mar = c(4, 3, 1, 1))
                  file.name <- ccat(pth,"ng_",ph,"_.txt")
                  write.table(get, file.name, append=TRUE, quote=FALSE,
sep="\t", col.names=FALSE, row.names=FALSE)
                  if (TRUE) {
                        data <- read.table(file.name, sep="\t",
header=FALSE)[,1]
                        den <- density(data) ; x <- den$x ; y <- den$y
;  m. <- mean(data) ; sd. <- sd(data) ; x.m <- mid(x,y)
                        nnorm <-norm(x, x.m, sd.)
                        print(length(data))
                        plot(x, y, type="n", ylab="", axes=FALSE,
xlab="", ylim=c(0,largest(nnorm, den$y)))
                        abline(v=ph, col="dark gray")
                        lines(x, y, lt=2, lwd=2)
                        lines(x, nnorm , lt=1, lwd=2)
                        mtext("Correlation Coefficient", side=1,
padj=3, cex=1)
                        axis(side=2, labels=TRUE, cex=1.5, padj=0.3,
hadj=0.8, las=1)
                        axis(side=1, labels=TRUE, cex=1.5, padj=-0.8,
hadj=0.4, las=1)  #  x-axis
                        box()
                        xx <- (max(x) - min(x))/10 + min(x)
                        yy <- max(nnorm) - (max(nnorm) - min(nnorm))/10
                        text(xx,yy, substitute(phi*" = "*ph,
list(ph=ph)), cex=1.5)
                        fn <- ccat(pth,"ng_",ph,"_.pdf")
                        dev.copy2pdf(file=fn)
                  }
                  print(j)
            }
}


#########################################################
#########################################################
```

```
rm(list=ls(all=TRUE))
source("dir.txt")
library(sfsmisc)
randomize()

# one day gaps
get.days <- function(dif=1){
   dd <- alo.data$days ; n <- length(dd)
   d1 <- diff(dd)==dif
   g1 <- g2 <- matrix(FALSE, length(dd))
   g2[2:n] <- d1 ;  g1[1:(n-1)] <- d1
   get <- g2|g1
   return(dd[get])
}


pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\phs\\"

norm <- function(x, m. , sd.) {
   g <- 1/sqrt(2*pi*sd.^2)*exp(-(x-m.)^2/(2*sd.^2))
   return(g)
}

largest <- function(x,y) { if (max(x)>max(y)) { return(max(x)) } else
{return(max(y))} }

mid <- function(x, y=NA) {
   if (is.na(y[1])) {
      d <- density(x)
        x <- d$x ; y <- d$y
   }
   max <- max(y)/1.5
   get <- y>=max
   xx <- x[get] ; yy <- y[get]
   l1 <- xx[1] ; l2 <- xx[length(xx)]
   return( (l1+l2)/2)
}


days <- get.days(1) ; days <- days-days[1] + 1  ## max value is 3616
while (TRUE) {

            pth <- "C:\\Documents and Settings\\Troy Wynn\\My
Documents\\dissertation\\Dissertation\\autocorr\\phs\\"

            norm <- function(x, m. , sd.) {
               g <- 1/sqrt(2*pi*sd.^2)*exp(-(x-m.)^2/(2*sd.^2))
               return(g)
            }

            largest <- function(x,y) { if (max(x)>max(y)) {
return(max(x)) } else {return(max(y))} }

            ############################
            #  ONE DAY GAPS USU
            ############################
            phs <- seq(0, 0.95, length=20) ; phs <- phs[-1]
            for (j in 1:length(phs)) {
                  get <- matrix(NA, 20) ; ng <- length(get)
                  ph <- phs[j]
                  for (i in 1:ng) {
                        nn <- length(days)
```

```
                        n <- 10000 ; sds <- 1
                        r <- rnorm(n, 0, sds)
                        ar <- filter(r, ph, method="recursive")[(n-
5000):n] ## this generates autocorrelated noise. earlier data points
removed to enhance memory in those that remain.
                        ar <- ar[days]
                        get[i] <- peel.it(ar, days)$ph
                  }

                  par(mar = c(4, 3, 1, 1))
                  file.name <- ccat(pth,"1dg_",ph,"_.txt")
                  write.table(get, file.name, append=TRUE, quote=FALSE,
sep="\t", col.names=FALSE, row.names=FALSE)
                  if (TRUE) {
                        data <- read.table(file.name, sep="\t",
header=FALSE)[,1]
                        den <- density(data) ; x <- den$x ; y <- den$y
;  m. <- mean(data) ; sd. <- sd(data) ; x.m <- mid(x,y)
                        nnorm <-norm(x, x.m, sd.)
                        print(length(data))
                        plot(x, y, type="n", ylab="", axes=FALSE,
xlab="", ylim=c(0,largest(nnorm, den$y)))
                        abline(v=ph, col="dark gray")
                        lines(x, y, lt=2, lwd=2)
                        lines(x, nnorm , lt=1, lwd=2)
                        mtext("Correlation Coefficient", side=1,
padj=3, cex=1)
                        axis(side=2, labels=TRUE, cex=1.5, padj=0.3,
hadj=0.8, las=1)
                        axis(side=1, labels=TRUE, cex=1.5, padj=-0.8,
hadj=0.4, las=1)  #  x-axis
                        box()
                        xx <- (max(x) - min(x))/10 + min(x)
                        yy <- max(nnorm) - (max(nnorm) - min(nnorm))/10
                        text(xx,yy, substitute(phi*" = "*ph,
list(ph=ph)), cex=1.5)
                        fn <- ccat(pth,"1dg_",ph,"_.pdf")
                        dev.copy2pdf(file=fn)
                  }
                  print(j)
            }
}
```

CURRICULUM VITAE

Troy A. Wynn
Email: troy.wynn@aggiemail.usu.edu

EDUCATION

Ph.D., Physics, Utah State University, October, 2010.
B.S., Physics, Utah State University, August, 2004.
B.S., Mechanical Engineering, Brigham Young University, August, 1999.

TEACHING/LAB EXPERIENCE

Class Instructor ~ USU. Taught second semester calculus based introductory physics for physics and engineering majors during the summers of 2006, 2007, and 2008. Course material covered waves, optics, electro-magnetism, basic electrical circuits, and an introduction to modern physics. Had full responsibility for writing, grading, and administering exams; homework assignments; lectures; demonstrations; simulations; etc. Course administration was set up using the Blackboard Learning System and Mastering Physics online homework system. Instruction included use of video demonstrations, Power Point, and computer simulations. Course material was put together using TurboCAD, Excel, and Power Point. Other software employed: Visual Basic for Applications, Word, Excel, Quick Screen Recorder, R statistical programming language, Maxima and Gimp.

Teachers Assistant ~ USU. For the past several years provided recitation lectures to supplement professors' lectures. Duties included writing and grading quizzes, submitting grades, and grading exams. Also involved student interaction such as answering questions about homework and demonstrating solutions in a lecture format, as well as dealing with student complaints.

Lab Instructor ~ USU. Taught introductory physics labs for both physics and non-physics majors. Instruction topics included optics, electromagnetism, spectroscopy, basic electrical circuits, mechanics, chaos, ideal gases, and interference. Provided an introductory lecture at the beginning of each lab to familiarize the students with equations, concepts, and equipment. During the remaining time checked on students' progress, making sure they were completing their work in a timely manner and were doing it correctly. The labs include heavy use of the Science Workshop computer interface and Data Studio software.

College Teaching Seminar ~ A one semester teacher training course at USU, completed Spring 2007. This course covered 15 lessons on the following topics: teacher/course evaluation analysis; responding to students' requests for accommodations; course preparation plan; helping students learn from reading, discussions, and lectures; course assessment plan; problem-based learning; teaching methods/skills; ways to develop teaching knowledge and skills; and self-assessment.

DISSERTATION

Dissertation Title: *Statistical Analysis of the USU Lidar Data Set with Reference to Mesospheric Solar Response and Cooling Rate Calculation.*

Analyzing 10 years of middle atmosphere temperature data obtained from the USU Rayleigh lidar. Dissertation involves determining the existence of middle atmosphere cooling and the nature and extent of the middle atmosphere solar response. The data analysis is heavily statistical. Model selection methods are considered, in particular the effects of bias due to model misspecification. I have taken classes on principal component analysis and have employed non-parametric statistical methods in my work. Recent work focuses on the effects of collinearity on model regression coefficients and how collinearity can confuse regression results.

ACADEMIC WORK

Wynn, T. and Wickwar, V., "The effects of model misspecification on linear regression coefficients as applicable to solar and linear terms," presented at Rocky Mountain NASA Space Grant Consortium Symposium, Salt Lake City, May 2009.

Wynn, T. and Wickwar, V., "Multicollinearity between the solar proxy and a linear trend in an ordinary least squares regression – Applicable to Mesospheric temperature measurements," presented at Rocky Mountain NASA Space Grant Consortium Symposium, Salt Lake City, May 2008.

Wynn, T. and Wickwar, V., "The effects of large data gaps on estimating linear trend in autocorrelated data," presented at Rocky Mountain NASA Space Grant Consortium Symposium, Salt Lake City, May 2007.

OTHER

- 2007 Teachers Assistant of the Year award from USU Physics Department.
- Passed FE exam, 82nd percentile.
- Received a Rocky Mountain Space Grant NASA Consortium fellowship for 2006, 2007, 2008, and spring 2010.

OTHER WORK EXPERIENCE

Field Physics Assistant (Intern) ~ Space Dynamics Laboratory, May, 2009 thru Dec., 2009. An eight month internship through Space Dynamics Laboratory working at Dugway Proving Grounds, Utah. Duties included operating lidar equipment during field tests at various test sites and troubleshooting, setup, alignment, and transportation of lidar equipment.

Student Employee ~ USU. Responsible for installing telescope control motors and transmissions for the new USU Rayleigh lidar (presently under construction). Also did some basic welding and machining of the transmission supports. Installed electrical control wiring and motor controller; wrote LabVIEW control programs to position telescope; did some basic electrical work such as installing overhead lighting, laying conduit, and pulling wires.