5-2019

Juxtaposition at 45 km of Temperatures from Rayleigh-Scatter Lidar and Reanalysis Models

David K. Moser
Utah State University

JUXTAPOSITION AT 45 KM OF TEMPERATURES FROM RAYLEIGH-SCATTER LIDAR AND REANALYSIS MODELS

by

David K. Moser

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

of

MASTER OF SCIENCE

in

Physics

Approved:

_____ Vincent B. Wickwar, Ph.D. Major Professor	_____ Robert R. Gillies, Ph.D. Committee Member
_____ Maura E. Hagan, Ph.D. Committee Member and Dean of the College of Science	_____ Laurens H. Smith, Ph.D. Interim Vice President for Research and Interim Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2019

ABSTRACT

Juxtaposition at 45 km of Temperatures from Rayleigh-Scatter Lidar and Reanalysis Models

by

David K. Moser, Master of Science

Utah State University, 2019

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

Atmospheric reanalysis models continue to increase in size and scope as we learn more regarding the workings of the complex natural systems involved. They provide for comprehensive global data sets widely regarded as some of the most useful tools both for analysis of historical climate variables and prediction of future trends. With their wide use in such studies, it is understandably important for the models to produce consistently accurate results that can be verified by direct observations. Considering the chaotic and interactive nature of atmospheric physics, small miscalculations in predictions can quickly propagate into larger sets of errors that seem to describe an entirely alternate reality. Reanalysis models curtail this problem through assimilation of millions of observations at each time step, smoothing out inaccuracies by constantly introducing new

constraints and training the models to meet these real data points along the way. However, any systemic issues that remain will inevitably cause a model to fail at times and places where observations are scarce, or nonexistent.

The original Rayleigh-scatter lidar system at the Atmospheric Lidar Observatory (ALO-USU; 41.74° N, 111.81° W), a part of Utah State University's Center for Atmospheric and Space Sciences (CASS), operated over the 45-90 km altitude range from 1993 to 2004. This range not only covers some of the highest altitudes available in most reanalyses, but is also some of the most difficult vertical space in which to obtain consistent temperature measurements of the sort that would be used as inputs and corrections to the models. The nature of the lidar temperature derivations results in uncertainties on average of less than a degree at 45 km. This thesis employs the measurement confidence at this altitude to highlight flaws in co-local and co-temporal temperatures provided by three popular atmospheric reanalysis models: NASA's MERRA-2, the ECMWF's ERA-20C, and the JMA's JRA-55. The comparisons reveal statistically significant cold biases in all three models that persist for the majority of the analysis period.

(169 pages)

PUBLIC ABSTRACT

Juxtaposition at 45 km of Temperatures from Rayleigh-Scatter Lidar and Reanalysis Models

David K. Moser

Weather and climate forecasts are almost exclusively produced by computer models, which use atmospheric measurements as starting points. It is a well-known and joked-about fact that model predictions can be incorrect at times. One of the reasons this happens is due to gaps in our knowledge of atmospheric conditions in areas where measurements don't often taken place, such as the mesosphere, which stretches from roughly 45-90 km altitude.

A lidar is a device that can shoot out short bursts of laser light to measure things such as atmospheric thickness at a distance. From this information one can then derive the temperature in the upper atmosphere. Using temperature measurements taken by lidar at Utah State University (41.74° N, 111.81° W) and temperatures from three popular atmospheric models, a comparison is made covering the period 1993-2004 at 45 km altitude. This comparison demonstrates poor predictive capabilities of the models at the target altitude and suggests the need for integrating datasets such as lidar data into future models. The modeling community depends on real measurement comparisons to bolster the reliability and credibility of their own work, and the comparison done here is intended to highlight an area in need of improvement.

ACKNOWLEDGMENTS

First I want to thank all of the great people involved in the ALO-USU program historically and currently. In particular: Vince Wickwar, for bringing me on to the project, and his friendship and assistance throughout my time at USU; Josh Herron and Durga Kafle, for the vitality of their work on the program's 1993-2004 run; Leda Sox, for providing me with not only bits of knowledge here and there, but also the opportunity for hands-on experience with the instrument; and for the many other undergrads and graduate students who both I and the program have relied on over the years. Of course, this gratitude also extends to my committee, who've been incredibly patient with me, and the unrelenting physics department staff who've done so much work in the background to make this all possible.

I'd also like to express special thanks to: my brother-from-another-mother Taylor, for helping me stay sane; my wonderful parents, who instilled me with the requisite curiosity and the desire to learn; and my siblings, who've always believed in me even when they've had no reason to. The unbroken support of these people is truly a boon.

Finally, I owe the world and more to my dear sweetheart, Nicole, who didn't know what she was getting into when she made me return to school so many years ago. None of this would have happened if not for you. Thank you for "C for Christmas." Thank you for your deep and abiding love. Thank you for our beautiful children. You are truly a gift to me and to them. You are all precious to me. You have my heart forever.

David K. Moser

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Studies of the middle atmosphere (~10-90 km altitude) have revealed significant coupling effects between the upper and lower atmosphere (Gerber et al., 2012). At mid-latitudes, middle atmosphere phenomena include a variety of physical processes such as tides (Sakazaki et al., 2018), sudden stratospheric warmings, or SSW's (Sox, 2016; Yuan et al., 2012), planetary waves (Beissner, 1997; Krüger et al., 2005), gravity waves (Baumgarten et al., 2018; Khaykin et al., 2015), and the occurrence of a double stratopause (Sivakumar et al., 2006). It has been demonstrated that the impacts of tropospheric weather on upper- and middle-atmospheric phenomena can last for days to months (Polvani & Waugh, 2004; see also Krüger et al., 2005), and that improving model physics in the middle atmosphere regime directly improves forecasting skill (Sigmond et al., 2013).

As state-of-the-art models respond to this knowledge by pushing toward the goal of whole-atmosphere simulations (Fujiwara et al., 2017; Molod et al., 2015; Siskind et al., 2015), it is imperative that model output be thoroughly validated by actual observations (Pawson et al., 2000). The middle atmosphere is routinely the most difficult regime within which to obtain constant measurements of atmospheric variables. Most of the ground-based methods used for tropospheric observation, such as sounding balloons and radar, are inoperable or ineffective at middle-atmospheric altitudes. Satellite observations alone are the most temporally consistent source of information, with limb-scanning measurements from the SABER instrument providing long-term temperature and chemical structure profiles of the region (Zhang et al., 2017). While SABER has

proven to be invaluable for middle-atmospheric science (Ortland, 2017; Rezac et al., 2015; Walterscheid & Christensen, 2016), satellite observations typically peak in precision below 35 km altitude (Hoppel et al., 2013), and it has been shown that assimilative model reliance on the satellite dataset produces errors in simulation products (Sakazaki et al., 2018). One of the advantages of satellites, and perhaps the main reason their data are popular for model assimilation, is their ability to observe large swaths of the planet at once, providing a near-global perspective on phenomena that is unparalleled by any other measurement technique. The disadvantage is found in the need to precess the satellite instrument in orbit to improve the temporal resolution of its data. Ground-based techniques encounter the opposite arrangement, with very localized and limited measurements in the spatial domain, but the ability to consistently record at very high temporal precision.

Ground-based lidar studies specifically provide an advantage in the middle atmosphere by continuously collecting scattered laser light as a measurement of relative atmospheric density, from which absolute temperature data can be derived (Hauchecorne & Chanin, 1980). This photocounting method is couched in Poisson statistics (Keckhut et al., 1993), and therefore, it results in high measurement accuracy at the lower bounds of the measured altitudes. Direct comparison of lidar measurements with those from SABER in particular have revealed systematic errors in the latter's dataset (Wing et al., 2018), highlighting the need for precision measurements from other sources as models continue to grow in scale and scope. Since the majority of error in lidar data occurs near the mesopause and below 35 km (where aerosols complicate measurements)

(Hauchecorne et al., 1991), the otherwise high temporal resolution and precise temperature measurements they provide make them ideal candidates for middle-atmospheric model output verification.

The Rayleigh-scatter lidar at Utah State University's Atmospheric Lidar Observatory (ALO-USU; 41.74° N, 111.81° W), a part of the Center for Atmospheric and Space Sciences (CASS), operated with few interruptions between 1993 and 2004, providing an extensive and useful dataset that can be used for model validation. Previous to a modern upgrade that extended the capabilities of the ALO-USU Rayleigh lidar (Sox et al., 2013), the original system consistently measured from 45-90 km altitude over the period of 1993-2004. This is the dataset that will be used throughout this thesis.

There are several other mid-latitude lidar datasets available, including two in southern France at 44° N (Keckhut et al., 2001; Le Pichon et al., 2015), one in Germany at 47.8° N (Eckert et al., 2013), the Jet Propulsion Laboratory Table Mountain Facility in California at 34.5° N (Kirgis et al., 2012), and a sodium lidar originally located in Fort Collins, CO at 40.6° N (Xu et al., 2006) that is now a co-located partner with the Rayleigh lidar at ALO-USU (Cai et al., 2014). There are also a multitude of middle-atmospheric studies at all latitudes that cover at least a portion of the same time frame as this thesis, including McDermid et al. (1998), Schöch et al. (2008), and Batista et al. (2009).

Many intercomparative studies have been done between different middle atmosphere data sets (Randel et al., 2004; Sivakumar et al., 2006), as well as between the data and common reference atmospheres (Hauchecorne et al., 1991; Herron, 2007), and

between measurement data and output from popular reanalysis and predictive models (Fujiwara et al., 2017; Keckhut et al., 1995). The latter type of comparison will be the focus of this thesis. A common approach to comparing measurements with model output is to reduce the information into month-by-month averages, or representative climatologies (Argall & Sica, 2007; Batista et al., 2009; Randel et al., 2004). The composite results can additionally be fitted by a least-squares model to look for trends, such as that predicted due to the effects of climate change (Angell, 1988), or the impacts of the 11-year solar cycle (Labitzke et al., 2002; Remsberg, 2008; Wynn, 2010).

Previous comparisons between lidar and assimilative model output have revealed temperature differences at near-stratopause altitudes (Batista et al., 2009; Finger et al., 1993). Some of the older studies have rather large measurement uncertainties (Argall & Sica, 2007; Hauchecorne et al., 1991), due both to the averaging of data over time and the scarcity of data (McDermid et al., 1998), which make them less useful as feedback for possible model improvements. Only recently has there been a push to compare data with output from the more modern state-of-the-art models (Fujiwara et al., 2017; Sakazaki et al., 2018). These studies indicate that, although many improvements have been made to extend model coverage and physics parameterizations over the middle atmosphere region, significant differences between the models and real measurements persist above 40 km and tend to increase with altitude (Le Pichon et al., 2015). Suggestions have been made regarding the cause of the differences, which include assimilative model overdependence on satellite observations (Hoppel et al., 2013), lack of dynamical variability at the top-end of models (Charlton-Perez et al., 2013), and the inability of

current model resolution to temporally and spatially account for important physical processes such as tides and gravity waves (Kim et al., 2003; Baumgarten et al., 2018).

Several composite climatology studies have been performed that illustrate the robustness of the ALO-USU Rayleigh lidar dataset (Barton et al., 2016; Herron, 2004). The lower altitudes are of special interest, with their very high measurement precision (Herron, 2004; Herron & Wickwar, 2018) near characteristic stratopause altitudes where models tend to have a lack of feedback.

This thesis provides a thorough comparison of highly-precise Rayleigh lidar temperatures with the output from three popular modern assimilative models at a single altitude near the stratopause (45 km). The three reanalysis models compared with ALO-USU Rayleigh lidar measurements are the European Centre for Medium-Range Weather Forecasts' (ECMWF) twentieth century reanalysis (ERA-20C) (Poli et al., 2016), the United States National Aeronautics and Space Administration's (NASA) Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro et al., 2017), and the Japan Meteorological Agency's (JMA) Japanese 55-year Reanalysis (JRA-55) (Kobayashi et al., 2015). Data from all four data sets (the three models and the lidar) covering the period of 1993-2004 were averaged into monthly temperatures on which a variety of statistical comparisons and error analyses were performed.

The outline of this thesis proceeds as follows: first, a description of the lidar instrument and associated equipment used in the collection of data. This is followed by a chapter outlining the derivation of absolute temperature data from relative density lidar measurements. The next chapter describes the uncertainties involved in the temperature

measurements, which are of vital importance to the conclusions of the thesis. A chapter is then dedicated to providing a basic background regarding the three reanalysis models used for comparison with the lidar data. A description of the procedures employed in analyzing and comparing the data is given in the next chapter, with the study results themselves presented in their own chapter immediately following. Lastly, some conclusions are drawn based on the results of this and other work in the field, and some future work is suggested.

THE ALO-USU RAYLEIGH-SCATTER LIDAR

Using the ALO-USU Rayleigh-scatter lidar system, researchers produced an 11-year dataset from 1993-2004 covering the region from 45 km to 90 km; roughly from the stratopause to the mesopause. At different times during this period the lidar made use of two Nd:YAG (neodymium:yttrium aluminum garnet) lasers, a Spectra Physics GCR-5 and GCR-6. The only fundamental difference between the GCR-5 and GCR-6 is that the GCR-5 features one amplifier with two flashlamps while the GCR-6 contains two amplifiers with a single flashlamp each. This means a lower full power for the GCR-5 due to less gain medium: 36 Watts versus the GCR-6's 48 Watts (at 1064 nm).

Nd:YAG lasers are relatively low maintenance, owing to the fact that they are solid-state lasers. This makes them ideal for prolonged use in scientific systems, like atmospheric lidar, that are used for recording lengthy temporal data. The flash lamps used to pump the laser need to be replaced every 30-40 million flashes. Barring any serious mishandling or damage to the laser apparatus, this is the only bit of upkeep required to continue operating the lasers for data collection. Under the regime they were used during this period, on average operating 10 hours each night at 30 Hz for just over a million flashes, replacement only needed to be performed at intervals longer than a month's worth of work.

While the fundamental wavelength of the GCR-5 and GCR-6 is 1064 nm, the backscatter cross- section on atmospheric returns is a factor of 16 higher at 532 nm. The crystal materials in Nd:YAG lasers make them prime candidates for second-harmonic generation, which can then be separated off using a dichroic beam splitter. Consequently,

although splitting off the secondary beam for use in transmission decreased overall laser power by 50%, the frequency doubling resulted in a net improvement of light returns by a factor of 8.

As with most Rayleigh lidar systems, the lasers were pulsed in order to obtain high energy output. This would result in periodic detectable light returns from Rayleigh-scatter off of atmospheric molecules, operating similarly to radar. A pulse length of 7 ns with a rate of 30 Hz was achieved by means of active Q-switching. The application of a Q-switch interferes with resonator feedback using cross-polarization such that lasing cannot occur. Meanwhile, the laser continues to be pumped, amplifying the photon population inversion through stimulated emission. Once the lasing medium is gain-saturated, voltage is supplied to the Q-switch causing a correction of the polarization, and allowing the stored energy to lase and escape in the form of a brief and energy-dense laser pulse. Not only did this technique provide for more laser power, but its cyclic nature also served as a timing mechanism for accurate data acquisition.

A diagram from Herron (2004) (Figure 1) provides a visual representation of the lidar's complete instrumental setup. Light output from the laser was optically redirected by mirror into the vertical. Aligned along the same initial horizontal axis was a 44-cm diameter Newtonian telescope with a 201 cm focal length. The telescope and laser apparatus were both mounted to the same optical table so as to maximize the ability to retain the alignment in case of small accidental bumps to the table. A turning mirror was used to redirect backscattered light from the atmosphere into the collection area of the telescope at a right-angle to the vertical. This mirror had a hole in the center for the laser

signal output to pass through, forming a coaxial alignment system. A field stop was placed at the focal point of the telescope to limit its field-of-view to roughly 3 times the beam divergence of the laser (1.5 mrad). On one hand, this was to account for some wandering of the laser beam due to internal thermal effects as well as variations in the atmosphere's index of refraction. On the other hand, it was made small to keep the signal-to-noise ratio high by limiting the amount of ambient light from the sky entering the detector system. A small collimating lens was placed after the field stop to focus it onto the plane of a mechanical chopper.

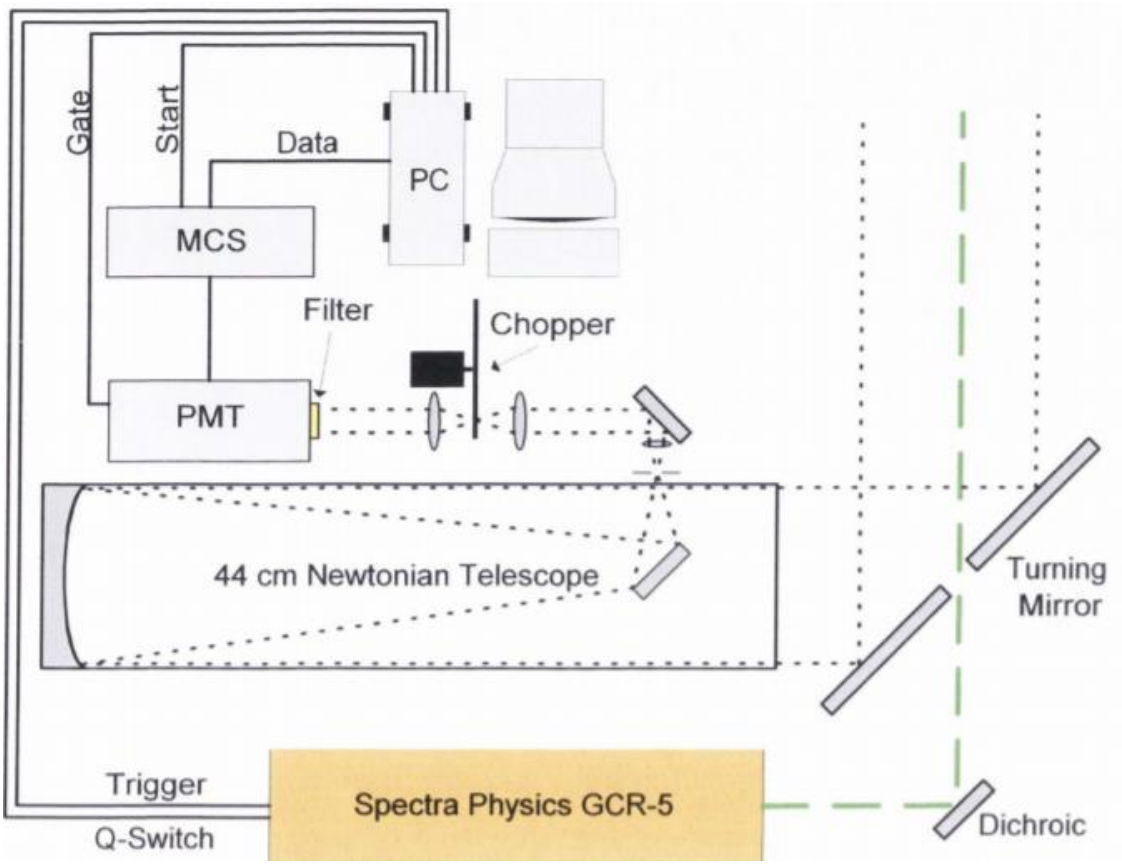


Figure 1 – ALO-USU Rayleigh-scatter instrument diagram (Herron, 2004).

The mechanical chopper was timed to operate in sync with the laser pulse frequency and length. Specifically, it operated at 4500 rpm, or 150 Hz. The phase of the chopper blade was adjusted to open at about 36 km, blocking most of the strongest light returns from below. These high returns could quickly overwhelm the photomultiplier tube (PMT) collector, as well as introduce after-pulsing oscillations, making it impossible to detect the relatively small number of backscattered photons from the higher altitudes of interest. The chopped light was then collimated again and passed through a narrow band-pass interference filter before entering the PMT itself. The filter was centered at 532nm with a full width at half maximum (FWHM) of 1 nm for normally incident light, effectively removing most of the remaining light received from the background sky and limiting collection to the wavelength of the laser signal.

The PMT was a 9954B-series model from Electron Tubes Enterprises, featuring a 46 mm diameter green-sensitive bialkali photocathode and 12 BeCu dynodes designed for good linear response and timing. A PMT converts photons incident on the photocathode into pulses of electrons by means of the photoelectric effect. These pulses are amplified by the dynode chain with each dynode being maintained at a higher potential than the previous one for a gain of 10^6. This makes it possible for the Rayleigh-lidar system to detect single-photon events, although the PMT tops out at ~15% quantum efficiency in reality. However, PMTs can also produce false signals by amplifying their own thermal electrons produced within the photocathode and dynode chain. This is called dark current, or since we were counting pulses from photons, dark counts.

In order to avoid a large number of dark counts, the PMT was placed within a housing providing electromagnetic shielding and dual cooling systems. The first stage of cooling was done by an external water chiller to reach 5° C, and the second stage used an internal Peltier cooler, further lowering the temperature to -25° C. By cooling and shielding the PMT, thermal electron generation was kept to a minimum. Any other thermal electrons generated within the dynode chain were distinguishable because of their weaker than normal electron pulses. This effect exists by simple virtue of having been produced somewhere other than the initial photocathode, and thus not going through the full amplification process as a result. The few remaining thermal electrons that are generated at the photocathode itself are accounted for by measuring and subtracting their effect out in the data reduction process. While this last step is incredibly important at higher altitudes, it makes very little difference at the 45 km altitude that is the subject of this thesis, as will be discussed in a later chapter. Total dark current impact on signal detection should be minimal after taking all precautions described to this point.

One final obstacle to using the PMT for atmospheric lidar returns has already been mentioned briefly, but requires more discussion here. Due to the presence of clouds and aerosols in the lower atmosphere, as well as the fact that the density of the atmosphere drops off exponentially with increasing altitude, there are orders of magnitude more backscattered laser photons from the lower atmosphere than there are from higher altitudes. In fact, the middle atmosphere is ideally suited to lidar study owing to the general lack of variety in particles from which backscatter can be received. A diagram showing the characteristic temperature profile and named regions of the

atmosphere is displayed in Figure 2. Other than the omnipresent molecular constituents from which lidar measurements can derive temperature, several other types of backscattering particles have also been listed on the figure near the region within which they commonly occur. While lower-atmospheric constituents are certainly interesting candidates for specific lidar studies, ALO-USU is most concerned with getting an accurate picture of temperatures in the mesosphere. Although a good portion of lower-atmospheric light returns are blocked by the timing of the mechanical chopper, the PMT would quickly saturate without further gating and results would become nonlinear and useless. Photoelectrons can be prevented from cascading further down the dynode chain if the voltage on successive dynodes is changed to be the same of the previous stage. In the ALO-USU system, gating was employed to lower the effect of photons captured below 35 km. This was done by setting the voltage of the photocathode to equal that of the first dynode, reducing the gain by factor of 1000.

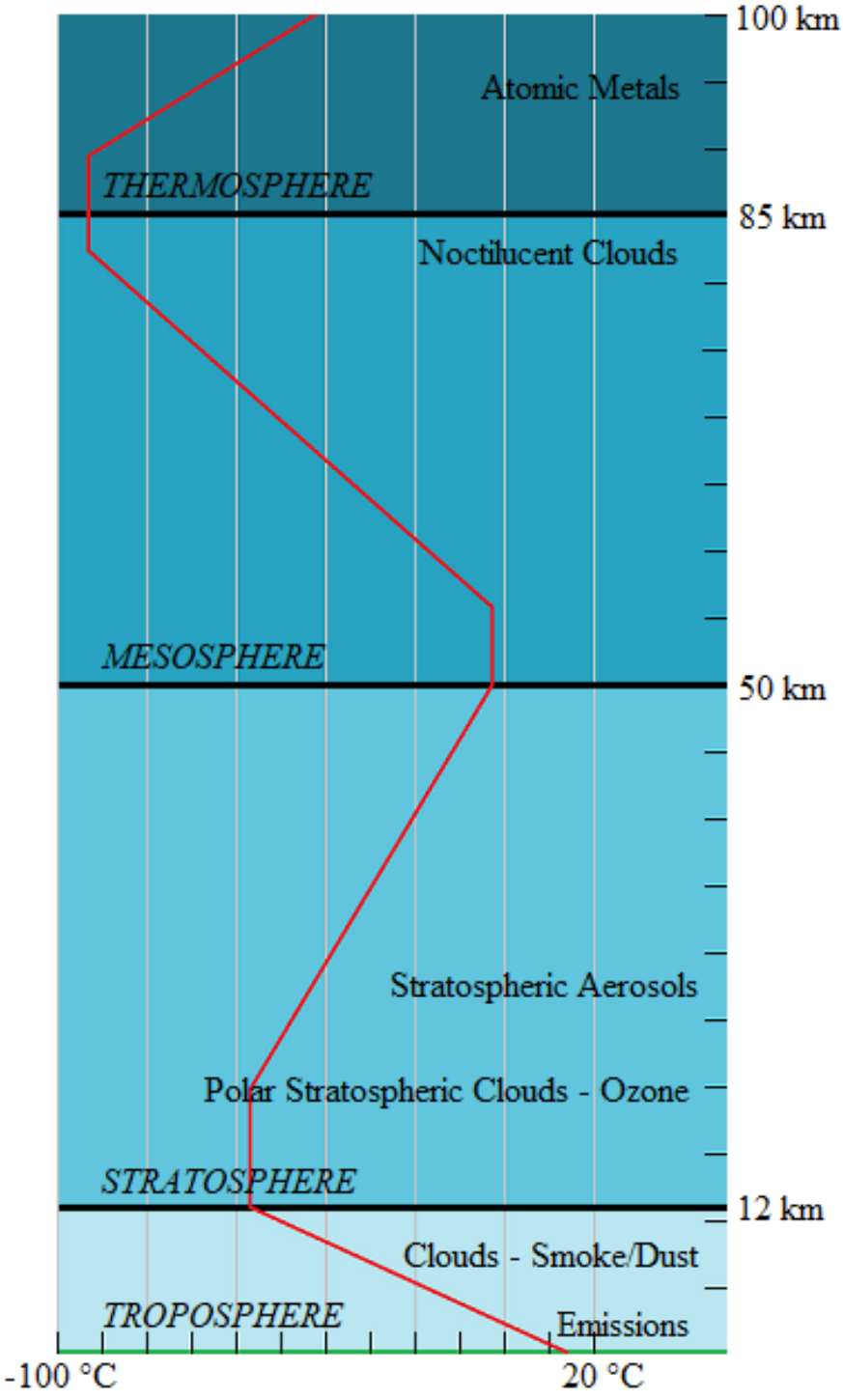


Figure 2 – Generalized map of atmospheric regions with temperature profile.

Finally, the PMT output in the form of a 3 ns FWHM pulse is sent through a 200x-gain fast pre-amplifier before traveling to an MCS (multichannel scaler) unit. Since the lidar system is essentially a time-of-flight photon-counting instrument, an MCS's ability to count at a maximum rate of 150 MHz, intervals of approximately 6.7 ns, makes it ideal for processing returns at a high temporal resolution. The pulses arriving at the MCS were binned at less than the maximum rate in order to spread the available 16,000 memory bins across a larger stretch of altitude. A 250 ns bin, translating to an altitude range of 37.5 m per bin, was chosen as more than sufficient to study the broader physical phenomena of the middle and upper atmosphere. In order to preserve hard drive storage space and allow sufficient time to download and record the data, the return signals were summed over two minutes, containing counts from 3600 laser pulses, before being successively saved to a file on a PC. Figure 3 taken from Herron (2004) shows an example of several hourly averages of lidar returns from 12 June 2003.

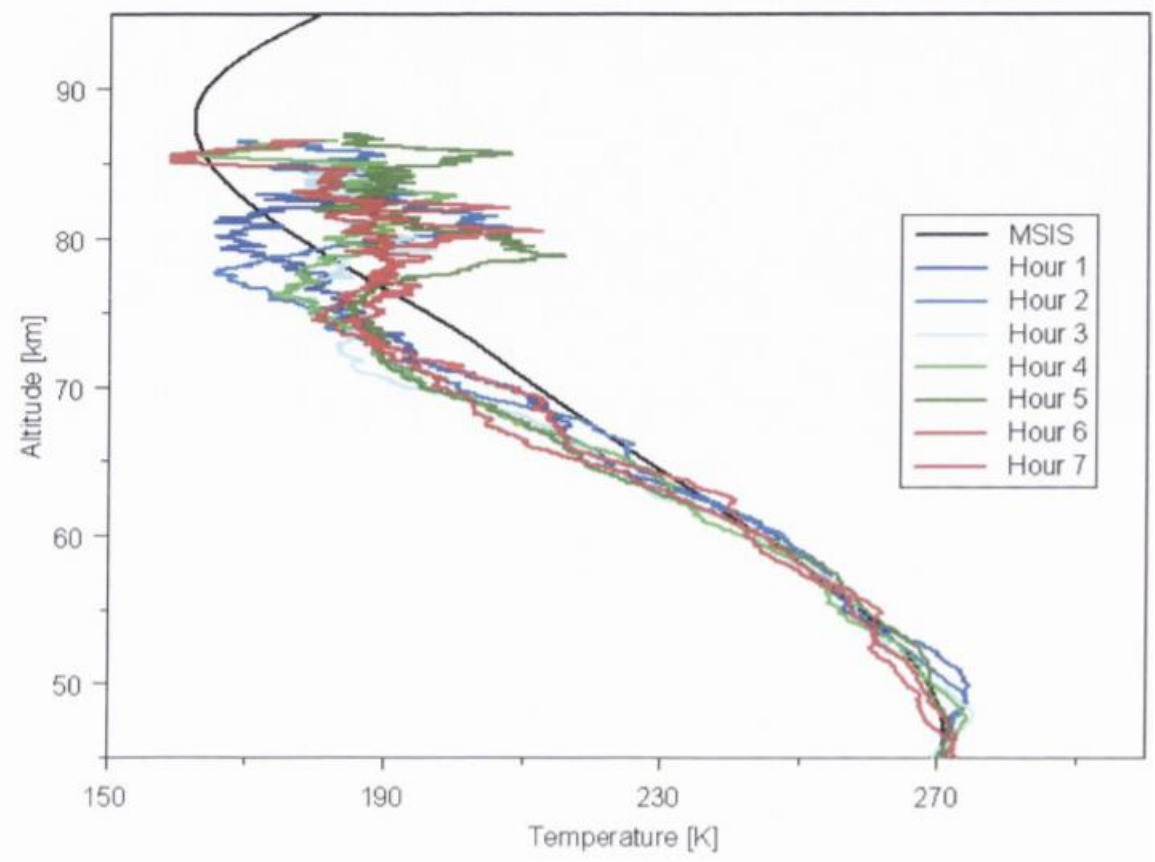


Figure 3 – Hourly average lidar temperatures compared with MSIS-E-90 empirical atmosphere (Herron, 2004).

MEASUREMENT OF ABSOLUTE TEMPERATURES

The number of backscattered photons $N(h)$ from a Rayleigh-lidar system is given by the lidar equation (Measures, 1992) below:

$$N(h) = \frac{N_0 AQT^2(h)}{h^2}\left[n(h)\sigma_\pi^R \right] \tag{1.1}$$

N_0 = Photons in outgoing laser pulse A = Telescope collection area

Q = Optical efficiency of lidar system $n(h)$ = Atmospheric number density at h

h = Height above lidar at backscatter location

σ_π^R = Rayleigh backscatter cross-section

$T(h)$ = Atmospheric transmittance between lidar and h

$T(h)$ is assumed to be constant over the range of the lidar, there being only a 0.4% variation (Hauchecorne & Chanin, 1980) in atmospheric transmittance between the altitudes of 35 and 90 km at a laser wavelength of 532 nm. Similarly, the Rayleigh backscatter cross-section is also dependent on the wavelength of scattered light and the constituents of the atmosphere. Due to the well-mixed nature of the mesosphere, the predominance of N_2, O_2, and Ar in the atmosphere, and the narrow laser light in use, a standardized definition can be used as constant for this study (Bucholtz, 1995):

$$\sigma_\pi^R(\lambda) = 5.45\left[\frac{550}{\lambda(nm)} \right]^4 \times 10^{28} cm^2 sr^{-1} \tag{1.2}$$

Owing to a variety of relentless temporal changes not only in the lidar system but in the atmosphere itself (e.g., smoke from fires, particles from inversion layers, clouds etc.), absolute atmospheric density measurements are extremely difficult to pin down. An attempt has been made to do this with the same lidar data (Barton et al., 2016) by normalizing the vertical profile of relative densities to data from reanalysis models at 45 km altitude. Since the object of this thesis is temperature measurements, however, this process is unnecessary here. If we invert the lidar equation to solve for the atmospheric density as seen here:

$$n(h) = \frac{N_0 AQ}{T^2(h)\sigma_\pi^R} N(h)h^2 \tag{1.3}$$

it then follows that we can solve for the relative atmospheric density at some height h by dividing $n(h)$ by its counterpart at a known altitude, $n(h_0)$, at a known altitude, h_0:

$$n(h) = n(h_0)\frac{N(h)h^2T^2(h_0)}{N(h_0)h_0^2T^2(h)} \tag{1.4}$$

Equation 1.4 no longer suffers from the complications brought on by the inclusion of system calibration measurements (photons-per-laser-pulse, system-wide optical efficiency, and comprehensive understanding of telescope collection-area), nor does it require knowledge of the backscatter cross-section.

To simplify further, over the altitude range being studied with the lidar (45 – 90 km) the atmosphere is so optically thin that much less than 1% of laser energy is scattered away as it travels through, and we can claim that $T(h) \approx T(h_0)$. This holds true anywhere

above 35 km, below which there are problems with ozone absorption, enhanced extinction of Rayleigh returns, and significant additional scattering from atmospheric aerosols otherwise not present at higher altitudes (Hågård & Persson, 1997).

With two additional assumptions and a few more steps we can arrive at absolute temperature measurements from the relative density equation (1.4) above. The first assumption is that the atmosphere is an ideal gas, described by the ideal gas law

$$P(h) = n(h)kT(h) \tag{1.5}$$

where $P(h)$ is the pressure, k is Boltzmann's constant, and $T(h)$ is temperature. The second assumption is that the atmosphere is in hydrostatic equilibrium (Chanin, 1984; Gardner et al., 1989). Hydrostatic equilibrium is defined by the gravitational force being balanced with the pressure gradient force like so:

$$\frac{dP(h)}{dh} + n(h)m(h)g(h) = 0 \tag{1.6}$$

where $m(h)$ is atmospheric mean molecular mass and $g(h)$ is the gravitational acceleration. This is also sometimes referred to as the steady-state diffusion equation. While this latter assumption may not be accurate in areas of strong turbulence, the fact that the return signal is summed for two-minute integrations and binned every 37.5 m lends credence to the idea that there is enough coverage over both space and time to smooth out any error from anomalies. This includes errors due to large amplitude waves, as was shown by Jenkins et al. in 1987.

The right-hand side of the ideal gas equation (1.5) can be substituted in for $P(h)$ in the hydrostatic equilibrium equation (1.6), and then integrated to find the change in temperature between a reference altitude h_0 and some other altitude of interest:

$$k\left[n(h_0)T(h_0) - n(h)T(h)\right] = -\int_{h}^{h_0} n(h')m(h')g(h')dh' \tag{1.7}$$

From here we can solve for the temperature at any altitude so desired:

$$T(h) = T(h_0)\frac{n(h_0)}{n(h)} + \frac{1}{kn(h)}\int_{h}^{h_0} n(h')m(h')g(h')dh' \tag{1.8}$$

Unfortunately, equation 1.8 relies on knowledge of the absolute density at the incremental altitudes in the integral. This may cause a divergence with the ratio of $n(h_0)/n(h)$, and will factor in any errors in density measurements. These issues can be easily avoided by integrating downward from a high reference altitude $h_0 = h_{max}$ to our altitude of interest:

$$T(h) = T(h_{max})\frac{n(h_{max})}{n(h)} + \frac{n(h_{max})}{kn(h)}\int_{h}^{h_{max}} \frac{n(h')}{n(h_{max})} m(h')g(h')dh' \tag{1.9}$$

As can be seen here, a specific selection of a high-altitude h_0 creates a situation where the system and model dependent parameters divide out with $n(h_{max})$, leaving only relative densities behind. The solution is then derived from a ratio of two relative density measurements.

The temperature equation (1.9) is initialized by first choosing h_{max} for each night to be where the average signal is 16-20 times its own standard deviation. This is typically an altitude in the thermosphere, though still within the realm of not disappearing into the background noise. This specific definition for h_{max} was chosen based on experimental trial and error, a process demonstrated using Figure 4 taken from Sox (2016).

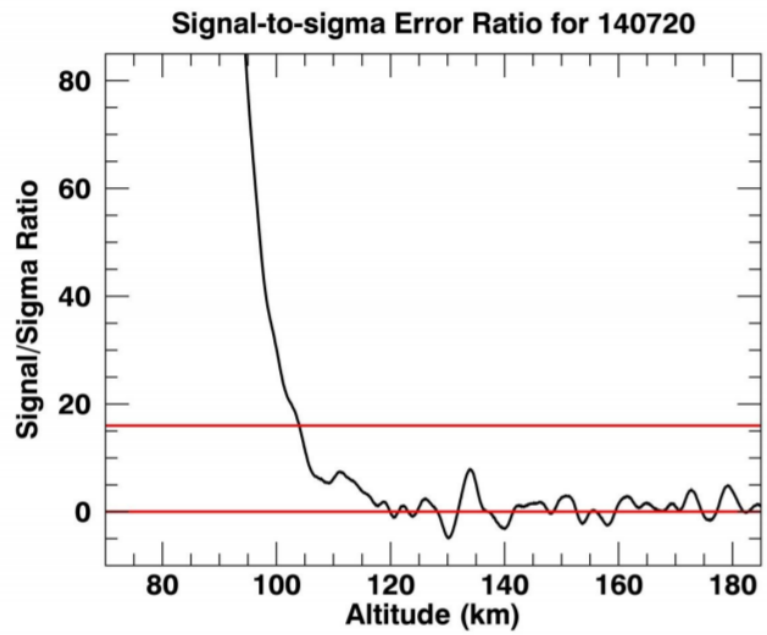


Figure 4 – Ratio of lidar signal-to-standard deviation, with red lines indicatin ratios of zero and sixteen (Sox, 2016).

Next, an initializing temperature at this h_{max} is normally obtained from the observations of an instrument other than the USU Rayleigh lidar, or estimated from an atmospheric model. During the time period of the collection of data used in this thesis,

two different sources were used depending on how high of an h_{max} was obtained for the particular set of measurements: a climatology from the Colorado State University sodium lidar (She et al., 2000), and NASA's MSISe00 atmosphere model (Picone et al., 2002). If h_{max} was above 83 km, only the sodium climatology was used, but if it was a lower altitude then an offset between both sources at midnight local time was applied for initialization. Due to some gaps in the sodium climatology data, the temperature was occasionally interpolated for the correct date and altitude (Herron, 2004). Argall and Sica (2007) have demonstrated that the choice of initialization climatology makes for very little difference in resultant temperatures after the top few kilometers of the data, and almost no difference whatsoever at the 45 km altitude involved in this thesis.

MEASUREMENT UNCERTAINTIES

It is of the utmost importance that measurement errors be accounted for in order to support the claims put forward later in this thesis. Specifically of interest is the amount of error in our temperature measurements at 45 km altitude. It can be shown that the measurement (Poisson) error is less than a fraction of a degree, meaning the total uncertainty at this altitude is driven fundamentally by geophysical variability (Herron & Wickwar, 2018). Such precision has allowed for studies normalizing the relative density measurements to provide for an estimated absolute density climatology of the ALO-USU Rayleigh-scatter lidar data (Barton et al., 2016). The following explanation of measurement errors is largely adapted from previously works using this lidar (Beissner, 1997; Herron, 2004), which have been fundamental in laying the groundwork for the comparisons carried out in later chapters.

The Rayleigh lidar atmospheric return data ultimately recorded by the MCS is effectively the superposition of the background noise, N, and the actual Rayleigh-scatter signal, S. The noise is composed of any PMT dark count and ambient sky light that remains after the application of all efforts to minimize these effects. The previous chapter details these steps if a review is desired. Noise is estimated by averaging the MCS records at altitudes high above the point at which the Rayleigh-scatter signal becomes negligible. The backscatter signal alone is simply the remainder of the total returns if the background noise is removed. Equation 2.1 below establishes this starting point for finding the error, with subscripts indicating variables respective both to altitude intervals

in space (*I* and *J*, corresponding to signal and background) and data integration intervals in time (*K*):

$$S_k = \frac{1}{I}\sum_{i=1}^{I}(S+N)_{ik} - \frac{1}{J}\sum_{j=1}^{J}N_{jk} = \overline{(S+N)}_k - \overline{N}_k \tag{2.1}$$

The background variance can be calculated from this as follows:

$$dN_k = \sum_{j=1}^{J}\frac{\partial N_k}{\partial N_{jk}}dN_{jk} = \frac{1}{J}\sum_{j=1}^{J}dN_{jk} \tag{2.2}$$

If we let $\sigma_x^2 = (dx)^2$ and square equation 2.2, it simplifies to:

$$\sigma_{N_k}^2 = \frac{1}{J^2}\left(\sum_{j=1}^{J}\sigma_{N_{jk}}^2\right) \tag{2.3}$$

Furthermore, if we assume the background noise to be constant with altitude, we have solved for the average variance in the background over J altitudes:

$$\frac{1}{J^2}\left(\sum_{j=1}^{J}\sigma_{N_{jk}}^2\right) = \frac{\sigma_N^2}{J} \tag{2.4}$$

Similarly, the combined signal and noise has a total variance given by:

$$d(S+N)_k = \sum_{i=1}^{I}\frac{\partial(S+N)_k}{\partial(S+N)_{ik}}d(S+N)_{ik} = \frac{1}{I}\sum_{i=1}^{I}d(S+N)_{ik} \tag{2.5}$$

If we assume each $(S+N)_{ik}$ to be independent of one another, we can square this equation and sum, arriving at:

$$\sigma^2_{(S+N)_{ik}} = \frac{1}{I^2}\left(\sum_{i=1}^{I}\sigma^2_{(S+N)_{ik}}\right) \tag{2.6}$$

To find the variance of the signal by itself, we merge the variance of the background from equation 2.3 with the combined signal and noise variance of equation 2.6:

$$\sigma^2_{S_k} = \sigma^2_{(S+N)_k} + \sigma^2_{N_k} = \frac{1}{I^2}\sum_{i=1}^{I}\sigma^2_{(S+N)_{ik}} + \frac{1}{J^2}\sum_{j=1}^{J}\sigma^2_{N_{jk}} \tag{2.7}$$

The lidar return signal adheres to Poisson statistics, since each photon counted is a random, independent event. On paper, this translates into any individual measurement's standard deviation being equal to the square root of that same measurement:

$$\sigma_x = \sqrt{x} \rightarrow \sigma_x^2 = x \tag{2.8}$$

This fact allows us to substitute the measurements in for their corresponding variances in equation 2.7:

$$\sigma^2_{S_k} = \frac{1}{I^2}\sum_{i=1}^{I}(S+N)_{ik} + \frac{1}{J^2}\sum_{j=1}^{J}N_{jk} \tag{2.9}$$

Rewriting equation 2.9 for the spatial averages over altitudes *I* and *J*, it becomes:

$$\sigma^2_{S_k} = \frac{1}{I}(\overline{S+N})_k + \frac{1}{J}(\overline{N})_k \tag{2.10}$$

At high altitudes where the data is initialized for the temperature derivations there isn't enough precision in a single measurement profile to be useful unless we also average temporally. In order to differentiate the temporal averaging from the spatial averages indicated by bars over the variables, the temporal averaging is represented by

angled brackets in all the following equations. The temporal averages are related to the spatial averages of the combined signal and the background signal respectively:

$$\left\langle \overline{S+N} \right\rangle = \frac{1}{K}\sum_{k=1}^{K}(\overline{S+N})_k \tag{2.11}$$

$$\left\langle \overline{N} \right\rangle = \frac{1}{K}\sum_{k=1}^{K}(\overline{N})_k \tag{2.12}$$

The variances corresponding to equations 2.11 and 2.12 are:

$$\sigma^2_{\left\langle \overline{S+N} \right\rangle} = \frac{1}{IK}\left\langle \overline{S+N} \right\rangle \tag{2.13}$$

$$\sigma^2_{\left\langle \overline{N} \right\rangle} = \frac{1}{JK}\left\langle \overline{N} \right\rangle \tag{2.14}$$

Recalling our starting premise in equation 2.1, the signal is represented by:

$$\left\langle \overline{S} \right\rangle = \left\langle \overline{S+N} \right\rangle - \left\langle \overline{N} \right\rangle \tag{2.15}$$

With variance:

$$\sigma^2_{\left\langle \overline{S} \right\rangle} = \sigma^2_{\left\langle \overline{S+N} \right\rangle} + \sigma^2_{\left\langle \overline{N} \right\rangle} \tag{2.16}$$

Therefore, the standard deviation of the photocounting Rayleigh-scatter signal is:

$$\sigma_{\left\langle \overline{S} \right\rangle} = \sqrt{\frac{\left\langle \overline{S+N} \right\rangle}{JI} + \frac{\left\langle \overline{N} \right\rangle}{JK}} \tag{2.17}$$

These uncertainties typically shrink by averaging either temporally or spatially. Whatever the final error in the signal, it must further propagate through the temperature

derivation to provide us with temperature measurement uncertainties. Following from equation 1.9 and the derivation in Gardner et al. (1989), we arrive at:

$$dT = \frac{\partial T}{\partial T_{max}} dT_{max} + \frac{\partial T}{\partial n_{max}} dn_{max} + \frac{\partial T}{\partial n} dn \tag{2.18}$$

and:

$$dT = \frac{n_{max}}{n(h)} dT_{max} + \frac{T_{max}}{n(h)} dn_{max} - \left(\frac{T_{max} n_{max}}{n(h)} + \frac{mg}{kn(h)} \int_{h}^{h_{max}} n(h')dh' \right) \frac{dn}{n(h)} + \frac{mg}{kn} \frac{\partial}{\partial n} \left(\int_{h}^{h_{max}} n(h')dh' \right) dn \tag{2.19}$$

Where the number density n is described by its relationship to scale height H, and decreases exponentially with increasing altitude as expressed here:

$$n(h) = n_{max} e^{-\frac{h_{max} - h}{H}} \tag{2.20}$$

Furthermore, if we let $c = mg / k$ and $\partial / \partial n = d / dn$, then $dn = (dn / dh)dh$ and $dn / dh = -(n / H)$. Substituting these into the final term of equation 2.19 we get:

$$\frac{c}{n} \left[n(h_{\max}) \frac{dh_{\max}}{dn} - n(h) \frac{dh}{dn} \right] = -\frac{c}{n} \left[n(h_{\max}) \frac{H_{\max}}{n(h_{\max})} - n(h) \frac{H}{n(h)} \right] = -\frac{c}{n} \left[H_{\max} - H \right] \tag{2.21}$$

Which is equal to zero if the scale height is held constant. The temperature variance therefore reduces to:

$$\sigma_T^2 = \left[\frac{n_{\max}}{n(h)}\right]^2 \sigma_{T_{\max}}^2 + \left[\frac{T_{\max}}{n(h)}\right]^2 \sigma_{n_{\max}}^2 + \left[\frac{T(h)}{n(h)}\right]^2 \sigma_n^2 \tag{2.22}$$

and substituting in the number density from equation 2.20 gives us the final equation describing the temperature uncertainties:

$$\sigma_T^2 = T^2\left(\frac{\sigma_n}{n}\right)^2 + \left[\sigma_{T_{\max}}^2 + T_{\max}^2\left(\frac{\sigma_{n_{\max}}}{n_{\max}}\right)^2\right] e^{\frac{-2(h_{\max}-h)}{H}} \tag{2.23}$$

Where we assume a neutral-density scale height of 7 km. Although this exact error is difficult to determine with precision, it decreases rapidly with altitude whatever initial errors might be present.

Herron (2004) performed several extensive simulations to verify the data reduction procedure and how it reacted to large amounts of errors in the top-level variables used for initialization. After 10 km of downward integration, the difference between actual and derived temperature decreases by a factor of 4, and after 20 km it decreases by a factor of 17. The initialization altitude for the data in this study is chosen by determining the altitude where the measured signal is 20 times its standard deviation, on average coming to 87 km and translating to an expected uncertainty of 6% in temperature. However, even if the initialization were much more than 6% away from the actual temperature at 87 km, let's say off by 20 K, this error would be reduced to less than 1 K by 66 km, and would be insignificant by the time it reached the 45 km altitude this thesis is concerned with.

As noted by Argall and Sica (2007), the standard deviation of the temperature measurement distribution is simply the combination of the Poisson-derived uncertainty and the geophysical uncertainty:

$$\sigma_{Tot}^2(h) = \sigma_{Geo}^2(h) + \sigma_{Poi}^2(h) \tag{2.24}$$

This in conjunction with a small fraction due to measurement error means that all standard deviations of the results are assumed to represent the geophysical variability or uncertainty alone. For additional background, the ALO-USU Rayleigh lidar composite climatology developed by Herron and Wickwar (2018) provides a thorough analysis of the geophysical variables involved at our location. From their study, a mean temperature deviation of less than 1 K can be expected in the lidar data for most of the year, with sharp peaks in the Northern Hemisphere winter when local geophysical variability is at its highest.

THE REANALYSIS MODELS

Although there are a number of reanalysis models available, three commonly referred to in the literature are the European Re-Analysis of the 20th Century (ERA-20C), the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2), and the Japanese 55-year Reanalysis (JRA-55). These were selected for comparison for three main reasons. First, their generally widespread use and the notable scientific organizations responsible for them provide a larger target audience. Second, they allow for obtaining a variety of assimilative model methodologies to enhance the results of the lidar measurement comparisons. Third, they are publicly available datasets that are obtained in a straightforward manner.

The data output that was used from each of the models is that of the instantaneous atmospheric diagnostic fields variety produced at 3-hour increments beginning at 0000 UT and ending at 2100 UT daily. 0600 UT was the selection used, corresponding to local midnight for ALO-USU.

According to Poli et al. (2016), the value of atmospheric reanalysis models lies in their ability to calculate and/or predict the evolution of the climate and changes in weather conditions. The systems operate by creating simulations on the basis of real measurement data from the surface, lower, and upper atmosphere, including whatever available satellite imagery and data from other types of equipment can be acquired and assessed to be of appropriate quality. This typically translates into meteorological measurements from roughly only the last six decades, mainly because the state of the science, the consistency of regular observation, and the availability of equipment prior to

1950 were insufficient to create the requisite sets of comprehensive meteorological information. Moreover, the majority of older data, when and if it is available, is in a physical written form incompatible with quick-and-easy machine processing, requiring extra work to digitize and assess the veracity of the records. Recent advancements in reanalysis, however, have allowed scientists to create comprehensive datasets for the entire 20th century, even going as far back as 1871. These extensions are based primarily on written surface pressure records and are owed to improved machine vision technologies that allow for quick digitization, while the majority of other data points are derived from a variety of approximations and complex calculations.

The ERA-20C model is the ECMWF's first reanalysis of the entire 20th century (Poli et al., 2016), and a development couched in the European Reanalysis of Global Climate Observations (ERA-CLIM) project. It was extended and improved off of the ERA-40 reanalysis. It features a unique coupled triple-model to examine different features of the geosphere more specifically, and is used mostly to study weather patterns since 1958.

The MERRA-2 system and its predecessor, MERRA, constitute the reanalysis systems with the longest history (Gelaro et al., 2017; Koster et al., 2016). They were developed by NASA as a way to combine satellite data with surface-based weather observations and transform them into readable and scientifically useful datasets. Their early work laid the foundations for the modern meteorological science and reporting that is ubiquitous today. The specific MERRA system itself was founded using the basic principles of the Goddard Earth Observing System Model (GEOS) system jointly

developed by NASA, NOAA, and the NCEP Environmental Modeling Center, which was responsible for the first reanalysis of atmospheric conditions in history, and a powerful tool in its own right (Molod et al., 2015).

The Japanese reanalysis system, JRA-55 (Harada et al., 2016; Kobayashi et al., 2015), has had only one previous iteration, JRA-25. The current model is based on a 55 year analysis of surface and upper-level observations from a wide variety of sources, and is essentially an improved model of JRA-25, which previously covered only a 25 year period. The extension to 55 years puts the newer model starting when regular radiosonde observations became more common worldwide. It is significantly more advanced and alleviates some of the main problems that could be found in the JRA-25 reanalysis, such as a cold bias resulting in the inability to properly represent lower parts of the stratosphere.

According to Stickler et al. (2014), ERA-20C is one of the most commonly referenced reanalysis models. This model is currently in the process of being replaced by the more advanced ERA-5 model, which will feature additional coupling mechanisms between ocean and atmosphere and a higher-resolution data grid. The new model is being produced under a renewal of the ERA-CLIM project called ERA-CLIM2.

The ECMWF operates on the principle of continuous monitoring and quality improvement of climatological observations. In keeping with this principle, the ERA-20C suite is designed to quickly receive real-time data inputs from ECMWF associates to tweak the model's predictive runs. Some of the observations assimilated include surface and mean sea-level pressures from the International Surface Pressure Databank

(ISPDv3.2.6) and the International Comprehensive Ocean-Atmosphere Data Set (ICOADSv2.5.1), as well as marine winds from the latter only. The assimilation of additional data is often the main driver behind the continued improvements to modeling processes, and is demonstrable by the incremental changes to ECMWF's products in particular. ERA40, ERA-Interim, and ERA-20C have each allowed for significant advancements in the quality and reliability of datasets over their predecessors. Poli et al. (2016) argues that one of the contributing factors toward the robustness of ERA-20C over other models comes from the development process, during which they would constantly run a control model without data assimilation alongside a model with the observations added. This would allow for the assessment of the exact impact of assimilated observations once added in. This process enabled them to address some key questions about deteriorating model parameters as a result of adding realistic weather information.

ERA-20C's model is formulated around the ECMWF's Integrated Forecast System (IFS), which is revised into improved "cycles" on a roughly yearly basis. Its methodology is to combine very large volumes of data within each 24-hour period and use variational bias correction of surface pressure observations to produce an accurate model of weather and air fluctuations. It uses three coupled models to complete the analyses, each with its own applicable observational records as input: the ocean wave model, the atmospheric model and the land-surface model (Hersbach et al., 2015). Each of the three contributes to the significant breadth of the analysis, spanning 91 vertical levels up to roughly 80 km altitude with a horizontal grid of 125 kilometers. The temporal resolution is 3-hours for most variables, but data can also be compiled based on

daily, weekly, or monthly time intervals as well. Rather than vertical model levels based on altitude, data is provided on 37 pressure levels, 16 potential temperature levels, and the 2 PVU (Potential Vorticity Unit) potential vorticity level, also known as the dynamic tropopause.

To ensure that the assimilated data are accurate and correctly processed, ERA-20C has a strong quality control check, comprising several methods. The scope can vary from dataset to dataset, but the process includes such things as checking against previous models and background error estimates. One of the most commonly used checks is the EDA, or "ensemble data technique," which basically tests the model against the most improbable events of climate change and irregularity over a 100 year period

MERRA-2 is the second iteration of the Modern-Era Retrospective Analysis for Research and Applications product put together by the National Aeronautics and Space Administration (NASA). It spans the longest time frame, starting in 1871, and is tasked with analyzing temperature observations and other atmospheric data points from the widest number of sources. Given that the main purpose of a reanalysis is to utilize historic data within the modern context of equipment, algorithms and other tools, the sheer amount of data at NASA's disposal appears to put their assimilative model among the most complete. Koster et al. (2016) argues that by simply increasing the number and quality of input observations they limit the opportunities for incorrect outputs. This logically follows from the fact that inaccuracies are often attributed to incorrect or partial observations, and other errors that can stem from the data gathering process.

The original iteration of MERRA used the Community Gridpoint Statistical Interpolation (GSI) system developed by the University Corporation for Atmospheric Research's (UCAR's) DTC (Developmental Testbed Center), and the NASA-built Goddard Earth Observing System Model (GEOS) currently in its fifth iteration. According to Molod et al. (2015), the GEOS-5 used in MERRA-2 can be, and is, applied for several purposes other than reanalysis. However, it's a powerful and versatile product, and the advantages it brings to the MERRA-2 model greatly contribute to overall performance. The system is capable of running simulations on a global scale, with horizontal resolutions of 10 kilometers down to an impressive 1.5 kilometer, and vertical pressure surfaces from the sea level to 0.01 hPa (around 80 km). Weather analyses are extremely accurate and available in near-real-time, while data ranging from ocean tides and surface winds to cloud cover and upper atmospheric temperature are simultaneously gathered from as much of the planet at possible.

MERRA-2 was basically developed to bypass some of the inadequacies of the first iteration, but also to utilize more of GEOS-5's capabilities (Gelaro et al., 2017). For example, unlike MERRA, MERRA-2 is capable of assessing the horizontal discretization within a cubed-sphere grid, analyzing all spaces within the grid evenly. This helps to avoid the kinds of aberrations that can appear in the stratospheric output of other models. Another significant improvement from the first MERRA reanalysis is the ability to account for precipitation and re-evaporation, which allows for a much more reliable and realistic simulation to be created. MERRA-2 was also developed with an eye on aerosol

observations, including black carbon, organic carbon, sulfate, and dust (Molod et al., 2015).

The JRA-55 reanalysis system was developed by the Japan Meteorological Agency (JMA), and is the second iteration of the Japanese atmospheric reanalysis program. The first iteration was called JRA-25, which was permanently terminated in 2014.

JRA-55 spans the time frame between 1958 and the present. The lack of quality and quantity in earlier observations is the main limiting factor, since 1958 marks the beginning of regular radiosonde observations of the atmosphere on a regular basis in many places around the globe. In many ways, the JMA program shares similarities to the ECMWF's, as should be expected with a similar stated goal of incremental improvement on the original reanalysis model. Much of the data assimilated in JRA-55 is shared between the JMA and the ECMWF, and are used in both agency's models. However, the foundational observational data used for JRA-55 is from the ECMWF's previous model, the ERA-40. For datasets after 1979 JRA-55 uses the same datasets as its predecessor, JRA-25, again paralleling the situation with ERA-20C continuing ERA-40's work. For other data prior to 1979, JRA-55 extrapolates from a limited data pool of mostly surface-based atmospheric observations. These records are largely drawn from the data archives of the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) in the United States (Kobayashi et al., 2015).

One of the key advances of JRA-55 over JRA-25 is the use of state-of-the-art mathematical algorithms applied to extrapolated data from meteorological satellites.

Additionally, the system does include some supplementary atmospheric data that are not included in ERA-20C. These include measurements from the Japanese Geostationary Meteorological Satellite (GMS), as well as data products derived from said satellite that were also developed by the JMA such as the CSR (Clear Sky Radiances).

According to Harada et al. (2016), JRA-55 was the first reanalysis model to use 4-dimensional variational analysis (4D-VAR) in its assimilation of the atmospheric conditions from 1958 onwards. Subsequently, the ECMWF upgraded its methodology to include 4D-VAR analysis as well, but NASA has not followed suit despite many comparisons of the methods showing 4D-VAR to be superior (Rabier & Liu, 2003). 4D-VAR runs the model both forward and backward in a small temporal and spatial range around each data point to check for runaway parameters and correct the model toward observations according to their quality. The variational bias correction applied by the JRA-55 is focused on satellite radiances rather than surface pressures as is done with the ERA-20C. The model features a 55 km horizontal grid with 60 vertical pressure levels up to the 0.1 hPa pressure surface (around 65 km), with data available on 6-hourly intervals.

Some of the most prominent differences between ERA-20C and JRA-55 were noticed in the quality of tropical cyclone positional analyses, where JRA-55 appears to outclass ERA-20C on larger geographical scales, and vice-versa (Wang et al., 2016). However, more applicable to the work of this thesis, JRA-55 is less capable of discerning atmospheric differences at higher altitudes, which is not as much of an issue for ERA-20C or MERRA-2. This isn't to say that JRA-55 is terribly inconsistent at measuring upper stratospheric temperatures and detecting real anomalies, however. It simply has a

noticeably lower consistency than the other two models being examined, with a number of artificially created anomalies having been noticed in the results, likely due to some carryover of the same problem from JRA-25, where the issue was significantly worse (Harada et al., 2016).

While it is not within the scope of this thesis to fully measure or express the complete set of features, history, and differences between the three reanalysis models chosen here, some have already been noted, some will be summarized hereafter, and others can be found in the descriptive literature for each respective model (Gelaro et al., 2017; Kobayashi et al., 2015; Poli et al., 2016).

The JRA-55 is a definitively positive evolution of JRA-25, with better quality control of data, increased access to more data points, and a lower propensity for the artificial results seen in stratospheric temperatures. Many gaps in JRA-25's data grid have been filled in and the new version operates much more efficiently due to newer mathematical algorithms, including the important addition of 4D-VAR. Similarly, ERA-20C has improved on the ERA-40 product, extending the reanalysis timeline to cover the entire 20th century. In that fact, it differentiates itself as being more comprehensive than JRA-55, but still doesn't achieve the scope of MERRA-2, which reaches back as far as 1871. Regardless, all three models assimilate plenty of data over the time frame studied in this thesis.

The ERA-20C operates slightly differently than the others, being divided into three distinct branches: 20CM, 20C, and 20CL. 20CM is a forced model integration without any synoptic data assimilation, 20C is a surface and marine model using synoptic

pressure data and winds from ICOADS and ISPD, and 20CL is a land-surface model that is essentially a down-scaling of the 20C (Hersbach et al., 2015). As with JRA-55, ERA-20C utilizes 4D-VAR, which has significantly improved the quality and accuracy of simulations over its predecessors.

Lastly, MERRA-2 has much better precipitation analysis abilities, more congruent aerosol data, and better resistance to background errors through sheer amount of data assimilation than either JRA-55 or ERA-20C. The use of GEOS-5, exclusive to MERRA-2, is also a significant difference from the Japanese and European tools, and allows the model to gather, process, and analyze data from a wide variety of measurement systems in a highly advanced and efficient way. However, MERRA-2 still uses the older 3D-VAR assimilation methodology, albeit with incremental updates and corrections for precipitation forcing and some other variables. This makes it prone to background parameter drift, which it compensates for by assimilating more observations than its competition.

Most relevant to this thesis, however, is performance of all three of these models in the upper stratosphere. According to Harada et al. (2016) and Koster et al. (2016), the overall differences in temperature measurements in the upper stratosphere between the JRA-55 and MERRA-2 models are considered to be relatively small. Overall performance of ERA-20C is described in the literature similar to that of JRA-55. Even though JRA-55 and ERA-20C both utilize advanced 4D-VAR-based analyses, MERRA-2's GEOS-5 system, combined with its access to a much broader pool of atmospheric observations and satellite feeds, makes it appear to be the optimal choice for upper

stratospheric temperature analyses. With that said, any of the three models are described in their literature as fairly reliable even in the upper stratosphere where observations are much less common than at lower altitudes. According to Santer (2003), studies of the upper atmosphere, including but not limited to the stratosphere, have historically been within the purview of NCEP, which often works closely with NASA. This relationship may provide an edge to MERRA-2 over the other models in this regime.

A common acknowledgement from those that work on these or any similar reanalysis models is that there is a lack of adequate objective measurements that could confirm, or alternatively, challenge their results. This problem provides a noticeable space for many questions about the validity of model output. There are more observations being done worldwide at this point in history than ever before, but this still pales in comparison to the amount of data output by models at every location and timestamp. It is highly improbable that a one-to-one comparison and verification of model output using observational data will ever be a possibility. This is bolstered by the fact that models are improving in scale and resolution at least as quickly as new observational data is available. However, the existence of this problem also implies that the development of model-data comparisons and interconnectedness is a must for all future development. This is the main motivation for including discussion of the models in this thesis, and it is hoped that this work can be a tool to assist in the development of future models.

DATA ANALYSIS PROCEDURE

This work was initially inspired by a basic comparison done in private communications within the ALO-USU group in 2006. This original comparison was between the ALO-USU data and the Climate Prediction Center (CPC) model produced by NCEP and NCAR in 1996. The comparison used a simple sinusoidal fit consisting of five parameters: one intercept, two semi-annual terms, and two annual terms. The results merely confirmed that the phases of the largest oscillations were common between the two sets of data, and closer comparisons were not made at the time. The quick study also noted that there were outliers (10 – 25 K in magnitude) from the fit that were often co-temporal between ALO-USU and CPC, indicating that these were real events, but not speculating further.

Before explaining how this work further built on the original discussion, we must first explain the preparation of the data. The methodology for temperature derivations from lidar data has previously been explained, and interpolation on this dataset was not deemed necessary due to the proximity of the altitude bin (44.9787 km) closest to the target altitude (45 km). However, since atmospheric reanalysis models typically offer data at specific pressure surfaces rather than altitudes, and because their horizontal grids are often somewhat large (e.g. ERA-20C only provided whole degree latitude and longitude values), interpolation was required across spatial dimensions to achieve 45 km model temperatures directly above the ALO-USU's location at 41.74° N, 111.81° W. For all three reanalysis models this required calculations from data on geopotential height and

temperature at four pressure levels (1 hPa, 2 hPa, 5 hPa, and 10 hPa) and four horizontal grid locations representing those closest to the lidar. All programming, fitting, analysis, and figure generation was performed in MATLAB, and all original code can all be found in Appendix D.

The derived relationship used to convert the geopotential height to altitude was

$$z = \frac{R_e z_g}{R_e - z_g} \tag{3.1}$$

where R_e = 6368.6 km is the radius of the Earth at ALO-USU's latitude. A linear interpolation was acceptable to collapse the data's horizontal dimensions, but due to the exponential relationship of atmospheric pressure with altitude it was decided that a cubic spline would be more appropriate to interpolate across the four pressure levels with their corresponding altitudes. The result was an array of data representing daily temperatures at 45 km for the 0600 UT (or midnight local time in the summer, Mountain Daylight Time). The ERA-20C and JRA-55 data included every day for the period of January 1993 to December 2004, and the MERRA-2 data began 26 July 1993 instead with the same end date as the others. This small difference has only a very minor potential impact on comparisons at the beginning of the data, since the first day for lidar data is near the end of August 1993 and the fitting method used covers a broad window to smooth out deficiencies in the lidar measurement set specifically. Due to the available data formats, each model required a slightly different approach as far as the MATLAB code was

concerned, but each generally followed the above process and were standardized into a specific format to provide for quick and easy comparisons.

Initial comparisons were done on older reanalysis models, both to verify and expand on the previous work done with CPC, but also because those datasets were already available due to work done by Barton et al. (2016). These included the ECMWF's ERA-Interim model, which is an updated version of ERA-40, and NASA's original MERRA model. Nonlinear regression was performed using a 5-parameter Levenberg-Marquardt least-squares nonlinear fitting algorithm tuned to capture annual, semi-annual, 4-month, and 3-month oscillations (Moser, Wickwar, Navarro, Barton, & Herron 2015). While that work was a good proof-of-concept for this thesis, providing both a relatively good fit and revealing the outliers as in the private discussions aforementioned, there were clearly some deficiencies in the method. Most notably, the fitted curve, seen in Figure 5, only defined a period of a year and repeated itself regardless of expected natural variations such as those caused by the solar cycle, climate change, or the quasi-biennial oscillation (QBO). It was determined that an adaptive fit would be required to adequately identify real outliers, as well as to better understand the differences between the models and the lidar data.

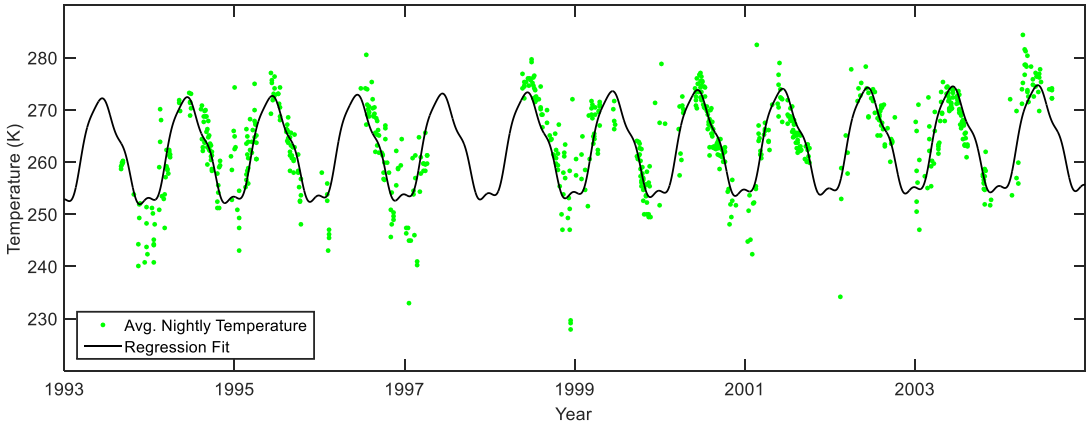


Figure 5 – Preliminary nonlinear model fitted to ALO-USU data.

The adaptive method developed utilizes the same nonlinear fit, but a 10th parameter was added in attempt to capture any linear trend such as that examined by Wynn (2010). The final fit equation is as follows:

$$T(t) = A + B\sin(2\pi t + C) + D\sin(4\pi t + E) + F\sin(6\pi t + G) + H\sin(8\pi t + I) + Kt \tag{3.2}$$

The technique for applying equation 3.2 was the big change. Rather than calculate the least-squares for the Levenberg-Marquardt over the entire 11 years of data and apply a single fit unable to capture unique year-to-year differences, a moving window was used. The window size was chosen to be 3 years for two reasons. First, as the window moves from day-to-day, the linear term would be able to adjust the fit to deal with interannual variations. Notable is the fact that the fit doesn't have a parameter targeted specifically toward phenomena like the QBO, but the nature of the moving window allows the fit to still shift to account for its effects. Second, the 3-year size is large enough to clearly overcome a gap in ALO-USU data from May 1997 to April 1998.

Each day was essentially provided with its own predictive fit calculated from all available data in a 3-year window with that particular day at the center. In order to maximize the amount of data available for fitting, the initial and final 1.5 years all used windows centered 1.5 years away from the data's temporal edge. The calculated parameters for the unique fit were then used to calculate a new predicted temperature value for that day alone. The method was applied to each data set starting on the first day of available data and ending on the last, but repeated for every day in between regardless of extant data on that particular day, which was only an issue with the ALO-USU measurements. This adaptive method resulted in 3998 separate fitted models with unique parameter values for ALO-USU, 4171 for MERRA-2, and 4383 for both ERA-20C and JRA-55. Full dataset fits are displayed in Figures 6 through 9. Rotated, larger versions of these graphs (Figures 31 through 34) are also available in Appendix A, since they include a large amount of data and detail.

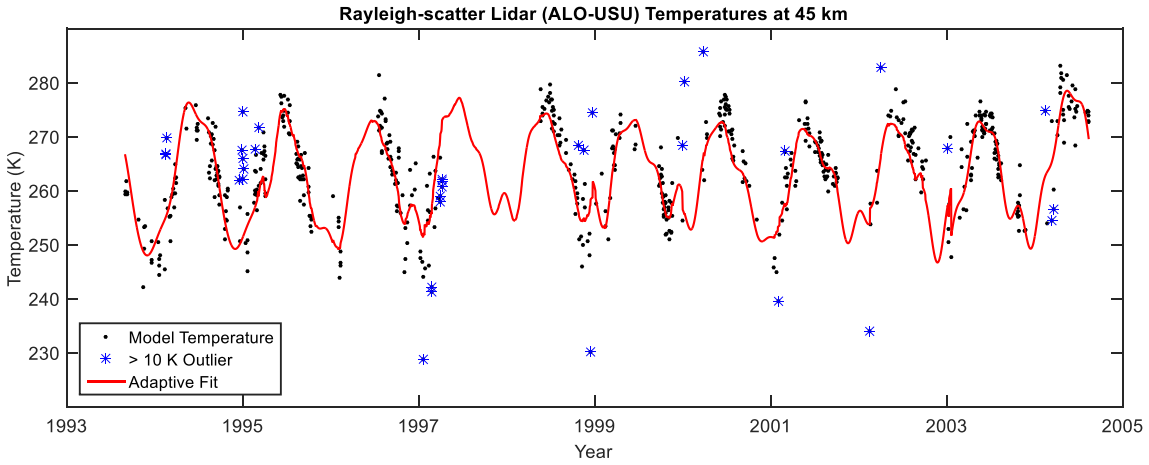


Figure 6 – Adaptive fit to ALO-USU Rayleigh-scatter temperature measurements.

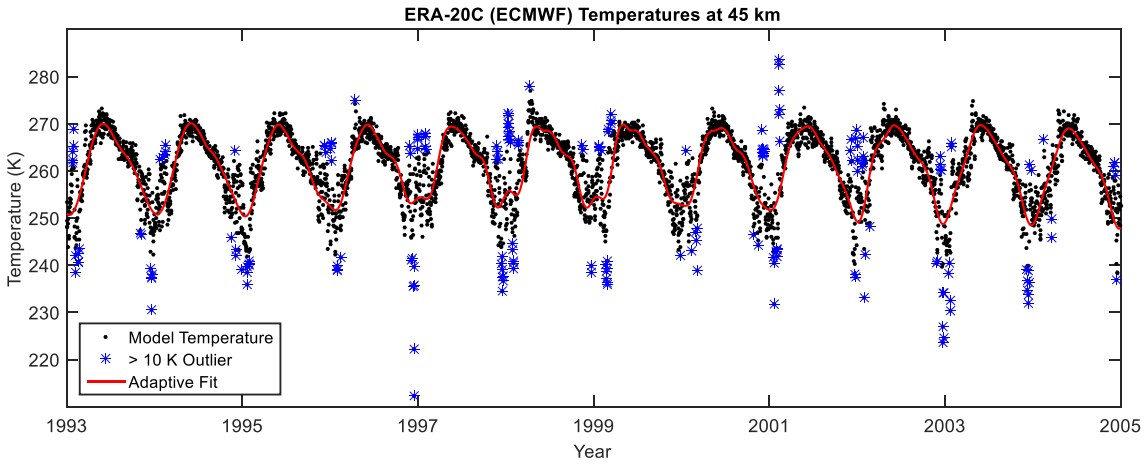


Figure 7 – Adaptive fit to ERA-20C model temperature output.

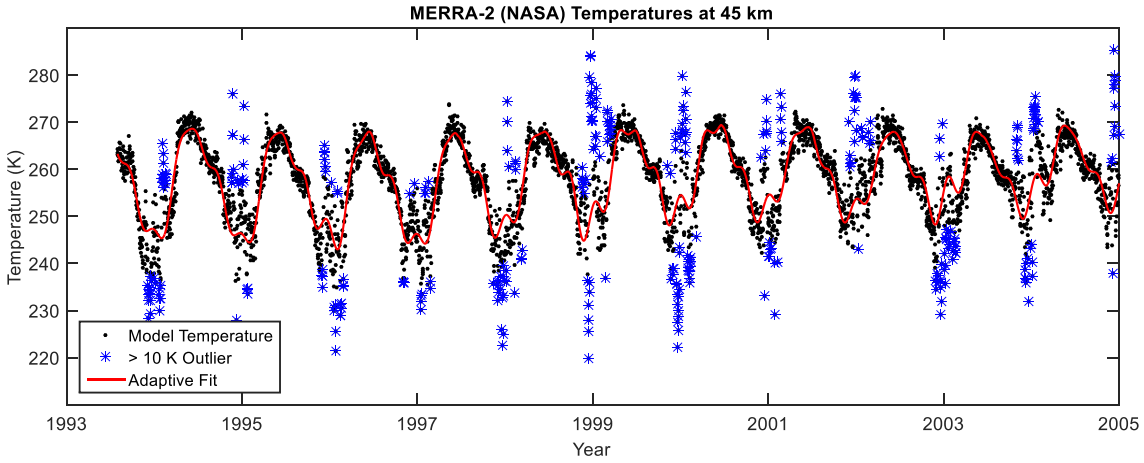


Figure 8 - Adaptive fit to MERRA-2 model temperature output.

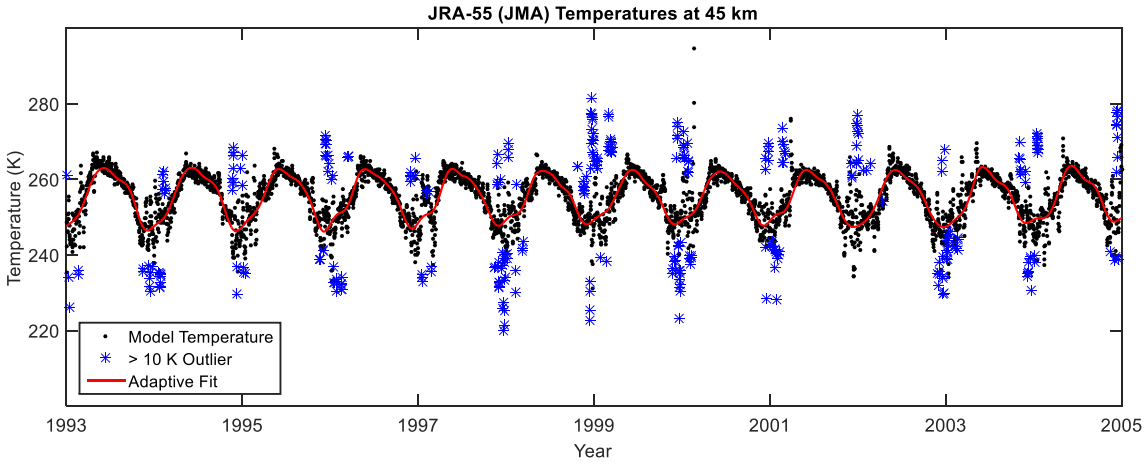


Figure 9 - Adaptive fit to JRA-55 model temperature output.

Because the lidar temperatures are temporally sparse when compared with the model data, one of the purposes of this fitting method was to provide a daily temperature value to compare with the models even on days when the lidar was not operating. More direct comparisons examining differences between the actual data when it exists is a significant part of this thesis as well, but some examination of the fit differences were deemed valuable based on the goodness of the fit. The probability density function histogram showing the residuals of each fit is given in Figure 10. In these plots, the residuals were fitted with the t location-scale distribution, rather than a normal or Gaussian distribution. This location-scale parameterization of the typical Student's t-distribution is more useful for modeling data involving constantly shifting volatility that results in more outliers, and therefore, heavier tails (Meyer, 1987). For all four fits, the t location-scale distribution suits the data, with outliers occurring roughly equally on both the positive and negative ends. Three parameters are used to calculate these distributions, the values of which are provided in Table 1. These parameters are referred to as the

location parameter μ, which represents the mean absent outliers, and the scale parameter σ and shape parameter ν, which are used to calculate the variance and identify outliers for the purpose of computing the mean.

Table 1

Distribution Parameters for Figure 10

Dataset	Parameter values		
	μ	Σ	ν
ALO-USU	-0.68	3.52	3.74
ERA-20C	-0.04	2.09	1.98
MERRA-2	0.16	2.64	1.93
JRA-55	0.09	2.05	2.69

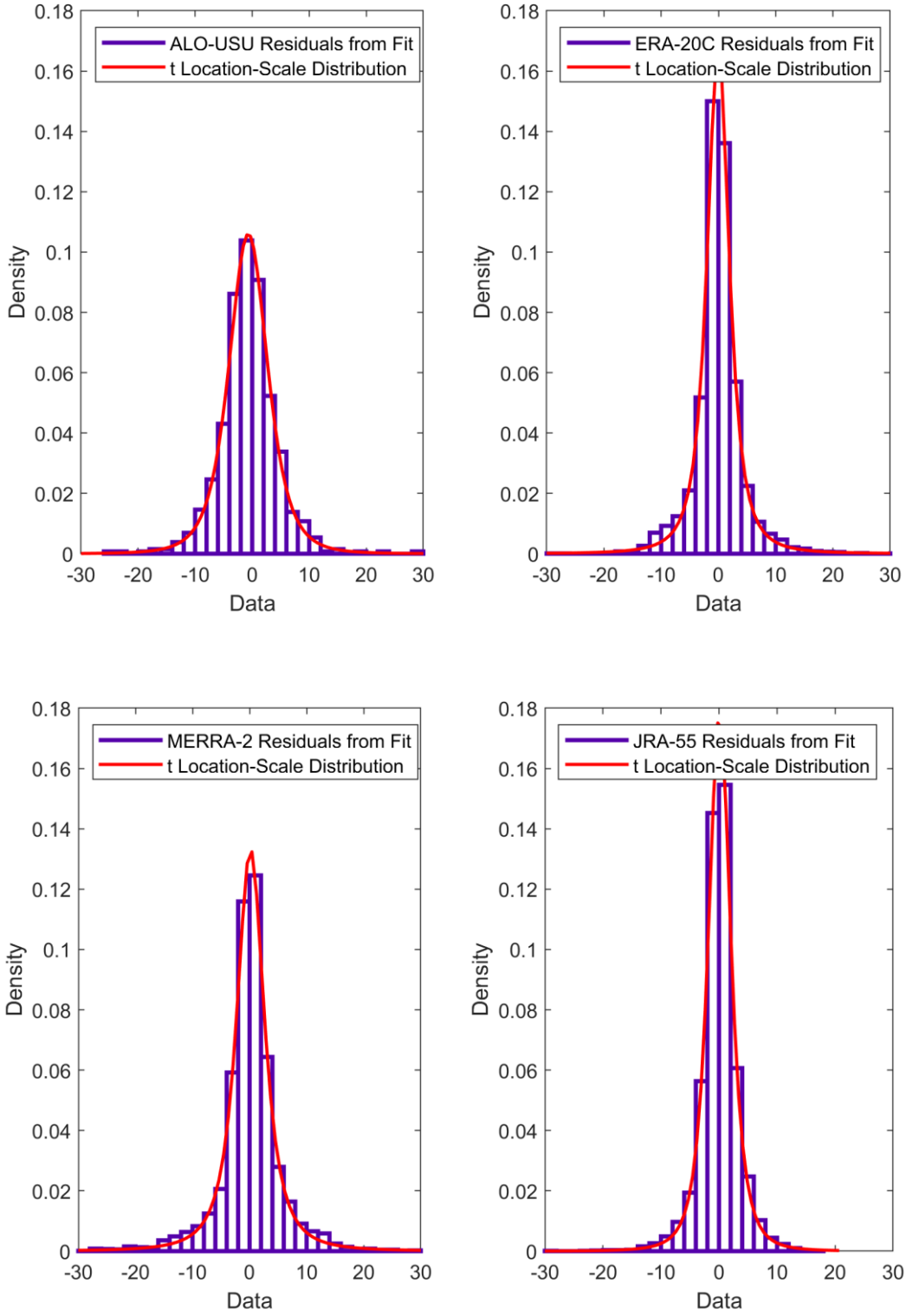


Figure 10 – Fit residual distributions for ALO-USU (top-left), ERA-20C (top-right), MERRA-2 (bottom-left), and JRA-55 (bottom-right).

The other purpose of the 10-parameter fit was to identify real temperature outliers; that is, outliers that can be attributable to natural geophysical causes rather than poor fitting. In order to verify that the adaptive fit was maximizing capture of the actual data, the method was repeated with 12-parameter and 14-parameter fits to see if smaller oscillations were not getting adequately represented. Figure 11 shows the results of these differences plotted against the MERRA-2 dataset, which can be compared against the 10-parameter MERRA-2 plot in Figure 8.

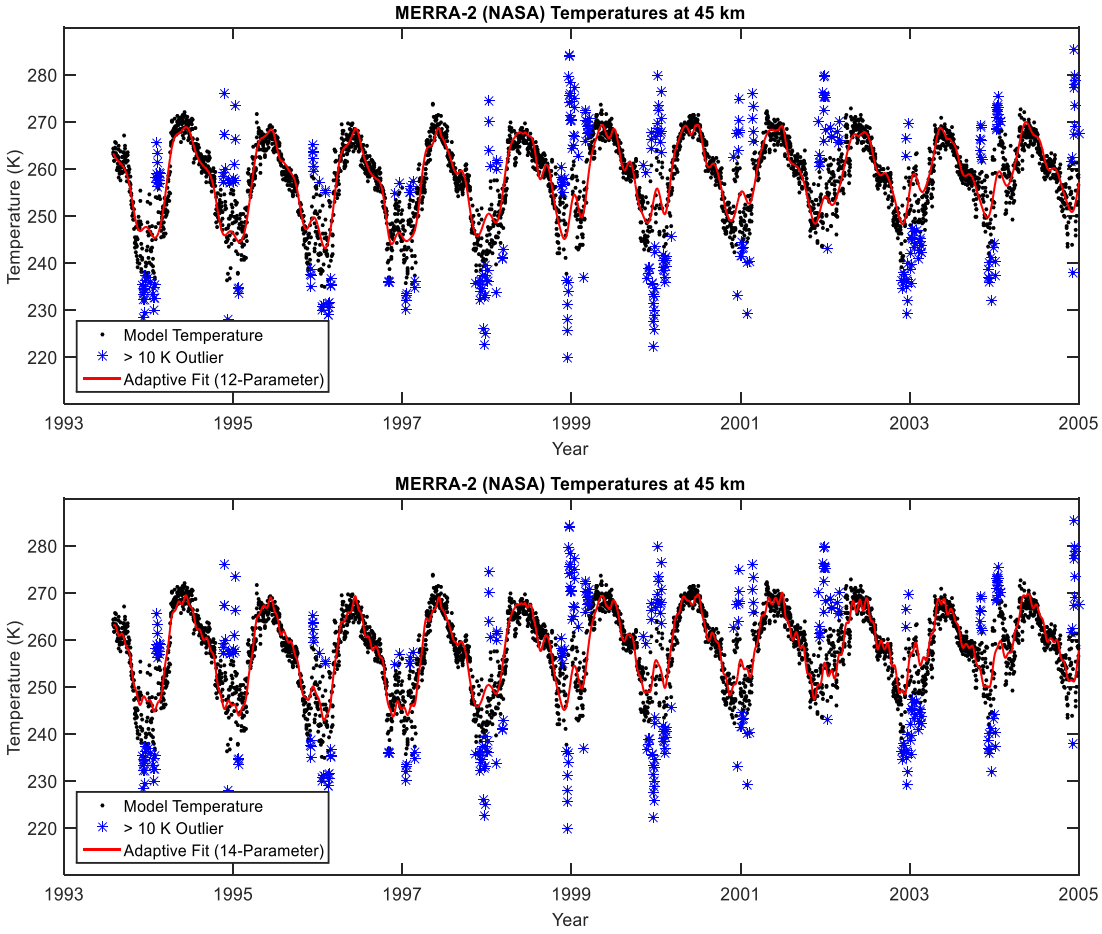


Figure 11 - Adaptive fit to MERRA-2 using 12 (top) and 14 (bottom) parameters.

The higher-order fits reduced the number of significant outliers (defined as having a residual from the fit of greater than 10 K) by less than 5% in most cases, and less than 1% in the case of the lidar data. Figure 12 displays a total monthly accounting of outlier occurrence, where it can be seen that although differences exist, they are minor and do not change any major conclusions that can be drawn from the fit method results. Additionally, the average value of the fitted temperatures changed by less than a thousandth of a degree in all cases except that of the lidar, where it changed by less than a tenth of a degree. Based on this information the 10-parameter fit was deemed to be adequate for the purposes of this study, and the 12-parameter and 14-parameter fits were not used in their current form in order to avoid overfitting the data. Figure 12 also shows that all outliers occur in the colder half of the year, when geophysical variability is higher (Liu et al., 2004) and real outliers should be most expected. There are typically more outliers in the spring than in the fall. With such a seasonal distribution it should come as no surprise that the ALO-USU data has significantly less total outliers, as weather-related reasons more frequently prevent lidar measurements from taking place during the height of outlier occurrence. While the existence of the outliers also makes it more difficult to perform comparisons between the lidar and model data in those months, those dates can still be assessed using the non-fitted data where it exists, and the identification of the outliers is valuable for future research. Tables 5 through 8, found in Appendix B, provide a full month-to-month accounting of the number of outliers in each dataset.

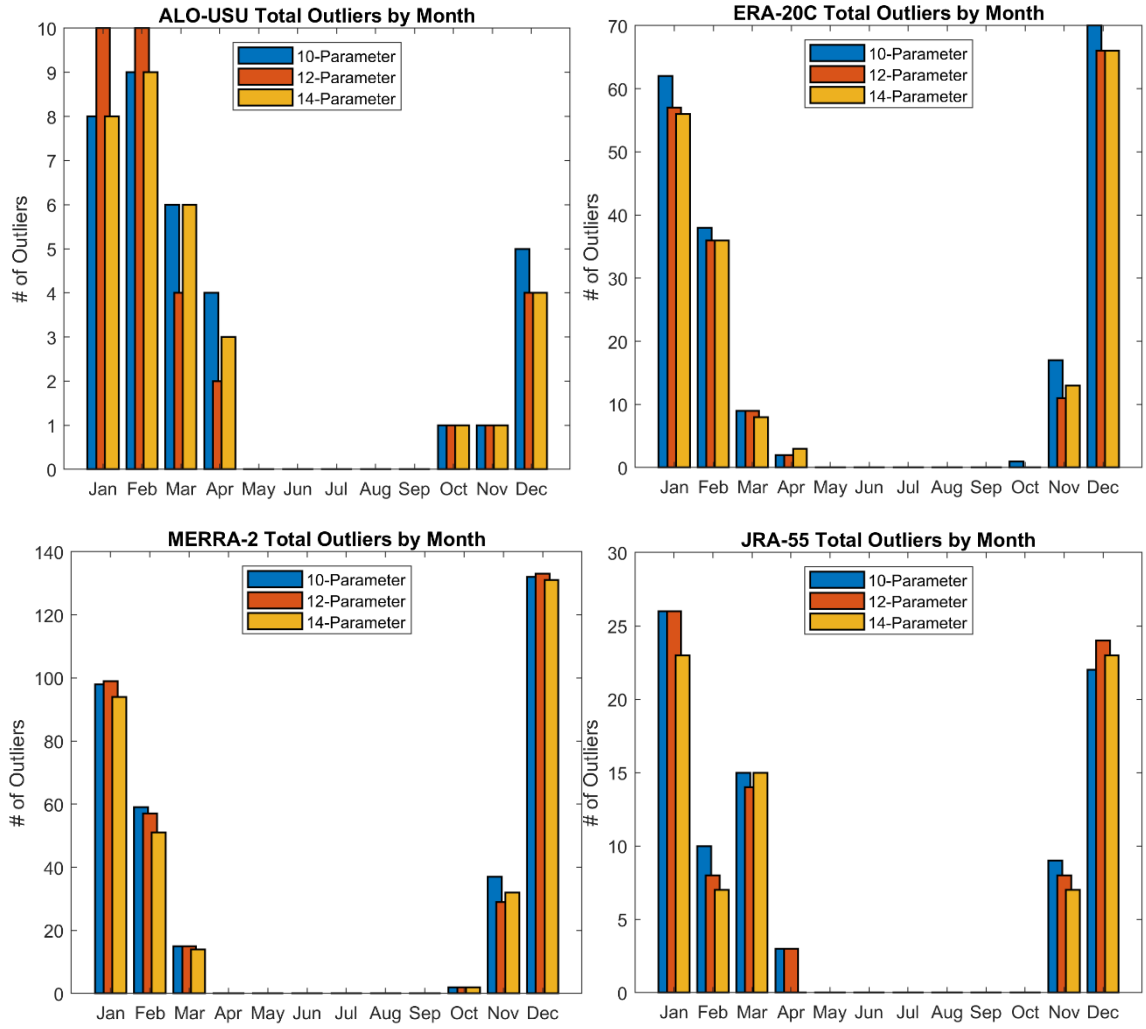


Figure 12 – Comparison of different parameterization regimes using outliers from the fits to ALO-USU (top-left), ERA-20C (top-right), MERRA-2 (bottom-left), and JRA-55 (bottom-right).

A thorough investigation of differences on individual days, as well as an examination of specific outliers, was deemed beyond the scope of this thesis and left mostly for future work, although there are a few exceptions that were identified for minor postulation. For the most valuable comparisons, information was divided into tables representing each month of the 11-year period. Four different varieties of average month

were created for each dataset, along with their standard deviations. It is highly recommended that the reader bookmark this page to use as a reference for continued reading. These data subsets will be referred to by abbreviations (where *###* is the dataset-specific abbreviation: *LID* for ALO-USU, *ERA* for ERA-20C, *MER* for MERRA-2, and *JRA* for JRA-55) to keep this presentation as straightforward as possible:

1. *###-REAL*, an average of all available dataset temperatures.
2. *###-RLID*, an average of dataset temperatures only from days that also have available lidar measurements.
3. *###-FIT*, an average of all temperatures predicted by the fit.
4. *###-FITL*, an average of fit-predicted temperatures only from days that also have available lidar measurements. (Note: FITL is not used in many of the upcoming examples comparing the data. However, the FITL values are all provided in tables along with the other data subset results in Appendix C.)

As an example to check understanding, the subset MER-REAL includes all 4171 temperatures output from the MERRA-2 model used as the basis for any other calculations, fits, or subsets of the MERRA-2 data. The subset MER-FIT also contains 4171 temperatures, but these are the predicted temperature values resulting from the adaptive application of equation 3.2 to the data as was described previously. Both MER-RLID and MER-FITL contain only 650 temperatures, and are the respective subsets of MER-REAL and MER-FIT that transpire on days concurrent with real ALO-USU Rayleigh lidar measurements. Note that for the lidar dataset, LID-REAL and LID-RLID

are equal (the lidar-days-only subset is the same as the total dataset), but LID-FIT and LID-FITL are not, since the fit predicts daily temperature values regardless of the existence of a co-temporal data point. This means LID-FIT provides daily temperatures over the entirety of the measurement period, filling in the gaps when measurements were not taken. This was a major driver behind applying a fit in the first place.

A single average over the entirety of each dataset was also performed, the results of which are in Table 2. The complete averages suggest that the models are all colder overall than the real temperatures measured by the lidar, though the standard deviations are too large to definitively say based on this comparison alone. Regardless, the smallest difference across any two means is found between ERA-FIT and LID-FIT at 1.55 K, which is still larger than all but the maximum ALO-USU uncertainties identified by Herron and Wickwar (2018) at 45 km altitude in their composite year analysis. Notably, the largest total average differences occur in the comparisons between the non-fitted data subsets, suggesting that the fit method may be smoothing over some genuine differences.

Table 2

Whole Dataset Statistics

Dataset	Subset	Number of temperatures	Mean (K)	Standard deviation
ALO-USU	REAL	650	264.61	8.61
	RLID	650	264.61	8.61
	FIT	3998	262.51	8.09
	FITL	650	263.92	7.13
ERA-20C	REAL	4383	260.97	7.67
	RLID	650	262.07	6.77
	FIT	4383	260.95	6.36
	FITL	650	262.33	5.45
MERRA-2	REAL	4171	257.84	9.18
	RLID	650	258.71	7.94
	FIT	4171	257.91	7.33
	FITL	650	258.86	7.11
JRA-55	REAL	4383	255.18	6.30
	RLID	650	256.64	5.50
	FIT	4383	255.19	5.27
	FITL	650	256.30	4.74

Contrasts between the monthly LID data and the model dataset averages were not always straightforward due to the disparity in number of dates containing data. Any given month of LID-RLID, LID-FIT, or LID-FITL could directly be compared with its corresponding model counterpart using a simple difference, demonstrating the value of the fit in making some comparisons. For example, the March 1997 MER-FIT is 254.47 ± 3.36 K, roughly 10 K colder than it's LID-FIT at 264.89 ± 2.96 K. We can say fairly confidently that, according to the fit, and for our location and altitude, MERRA-2 provides significantly colder temperatures than the actual measurements during March

1997. However, in any comparison with LID-REAL, we would always have a lower number of values involved in each average than a corresponding model dataset, compromising the fidelity of a simple difference. Therefore, for all monthly comparisons, including those with a common number of data points, a standard error for the difference between the means was calculated. This calculation was also employed by Le Pichon et al. (2015) to perform similar comparisons. From this statistic a t-test was then performed, providing for a confidence level for each difference.

For a thorough example of the calculations, let us again examine March 1997, but this time we will compare MER-REAL and LID-REAL. MERRA-2 provides temperatures for all 31 days of March 1997, averaging out to 255.33 ± 2.93 K, while the lidar only operated on 9 of those days and provides an average of 260.07 ± 4.41. As discussed in a previous chapter, these standard deviations are the total measurement uncertainty, containing both the Poisson-derived temperature uncertainty and the geophysical variability (Argall & Sica, 2007; Herron & Wickwar, 2018; Leblanc et al., 1998). The simple difference between the means comes to -4.74 K, much less than the difference in the fitted models, and possibly existing within overlapping error bounds. Due to the difference in the number of measurements we can't be confident that the difference isn't attributable to the lack of measurements or the uncertainties. So, using the available information we calculate the standard error of the mean difference, displayed below, where N is the number of measurements and σ is the standard deviation:

$$SE = \left(\sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \right) \left(\sqrt{\frac{(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2}{N_1 + N_2 - 2}} \right) \tag{3.3}$$

This is the value representing the error bars on the difference of the means, which comes to ± 1.25 K in our example. While we could eyeball it at this point and say we are pretty sure that the whole difference cannot be attributable to the lack of measurements, we can go further and perform a t-test, and then use a table to look up the associated p-value. Once we have obtained the standard error, the t-stat is easily computed using the following equation:

$$t = \frac{\left| \bar{x}_1 - \bar{x}_2 \right|}{SE} \tag{3.4}$$

where the numerator is the absolute value of the simple difference between the means we initially calculated. The t-statistic for our example is 3.80, which we can use to look up a p-value from a statistical reference table. For this case, we get a p-value of less than 0.01, which gives us greater than 99% confidence that the difference in the means is real and cannot be explained by the variance in or lack of measurements, whatever the source. Note that this process is referred to as the "two-sided pooled t-procedure," and assumes that the population standard deviations for each dataset are the same, even if the sample standard deviations may differ as they do here. Because the ultimate goal of climate modeling is to show no difference from reality, this appeared to be a safe assumption to make considering the possible existence of differences is precisely the point of the investigation.

Although a simple difference could have been used for comparisons in the cases of the RLID and FITL sets, the method described above comparing MER-REAL and LID-REAL was applied universally in order to obtain a homogenous approach to the results and avoid possible confusion. All conclusions take the associated p-values into account.

RESULTS

There are a multitude of possible ways to present the data and calculated results covered by this thesis. Month-by-month tables for the entire 11-year period and the four data sources can be found in Appendix C, containing all four average subsets (Tables 12 through 15), their standard deviations (Tables 16 through 19), and number of data points (Tables 9 through 11). The standard error of the mean difference and the t-statistic can be calculated from these tables using equations 3.3 and 3.4 respectively, and the t-statistic is used to calculate p-values as has all been explained in the previous chapter.

One set of results for discussion are the average values of the parameters from equation 3.2. Since the fit is unique to each day, these parameters can sometimes vary wildly and there is difficulty in identifying any single value with which one can easily describe the shape and trend of the data as is commonly done in other studies. For example, there are no terms aimed at capturing the QBO, though the moving window allows its effects to be accounted for. Because of these inherent differences between study approaches, the only parameters consistent enough to provide a sensible comparison are the averages of the yearly amplitude parameters (B) and the averages of the linear trend parameters (K), outlined in Table 3. In order to provide a second way of looking at the yearly trend, a simple least-squares linear fit was also performed on each full dataset, the slope of which is referred to in the table as *K alternative*. While this was a less than ideal way of providing yearly parameters that are able to be compared with similar parameters in other studies, it was the best option within the scope of the thesis.

Table 3

Selected Average Fit Parameters

Dataset	Parameter B		Parameter K		K alternative	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
ALO-USU	10.61	1.53	1.62	1.92	0.77	0.93
ERA-20C	8.65	0.40	-0.10	0.48	-0.03	0.49
MERRA-2	9.64	1.55	0.20	1.11	0.42	1.21
JRA-55	7.36	0.44	0.03	0.25	-0.04	0.24

Note. Units for all parameters displayed here are kelvins / year.

Moving on to the temperature results and their comparisons, it is much more valuable to provide a visual context to data such as this rather than discuss the raw numbers, of which there are many. One approach is to form a composite year. By following the procedure outlined by Herron & Wickwar (2018), we can produce the graphs shown in Figures 13 and 14, with σ defined as the standard deviation of the mean (Bevington, 1969) and dotted lines on the plots representing 3σ error bars rather than 2σ or 1σ for visibility's sake. Both of these plots support a few broad conclusions:

1. The ALO-USU lidar generally reports warmer temperatures than the models, with some seasonal exceptions.
2. Among the models, ERA-20C is the warmest, and JRA-55 is the coldest.
3. Only ERA-20C appears to have much non-winter overlap with the lidar.
4. We can be almost entirely certain that the models are underestimating summertime temperatures.

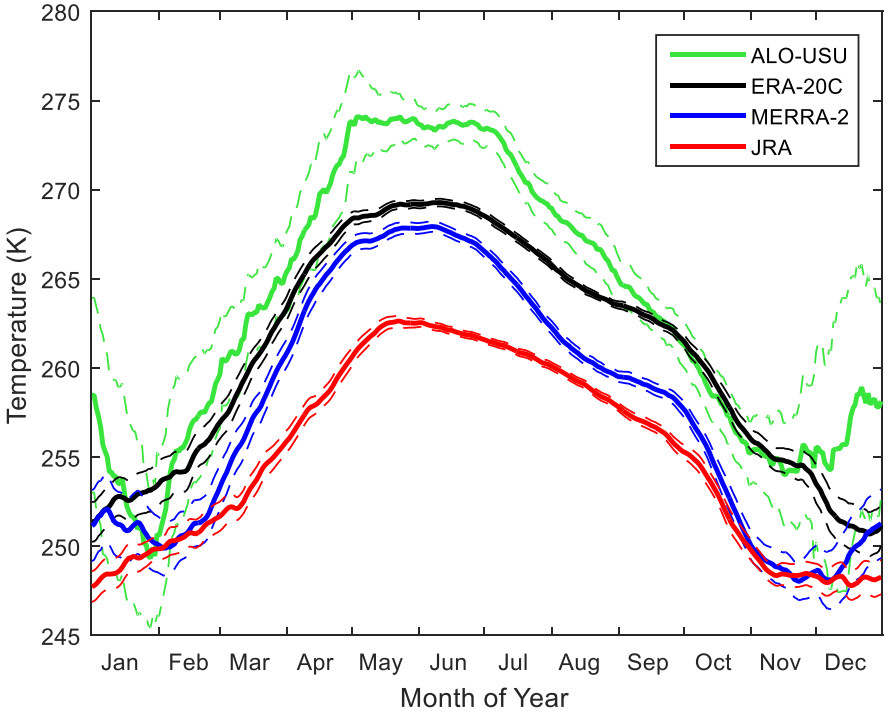


Figure 13 – Composite years (FULL datasets).

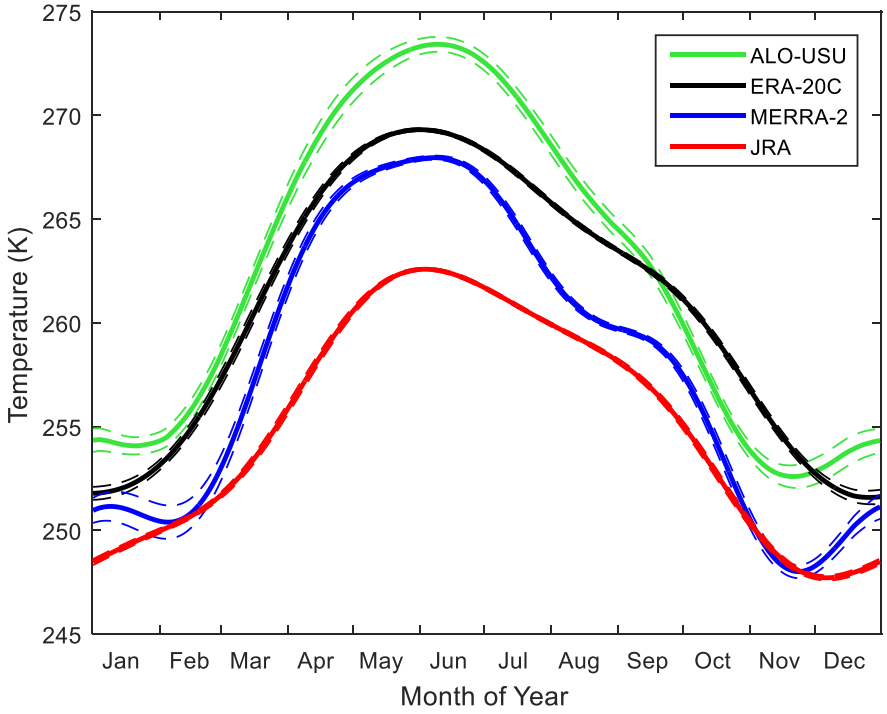


Figure 14 – Composite years (FIT datasets).

These same broad conclusions tend to play out repeatedly in further analyses of the data. However, more temporal granularity is helpful in identifying specific differences and weak points in the models that need to be addressed. This was the main motivation for applying a fit to the data in the first place. A daily difference between the models and the lidar is not possible because the lidar wasn't taking measurements every day. For many cases, the fit provides a decent approximation for what the lidar should have measured, so we can instead compare the model data to the lidar data's fit. Subtracting the REAL dataset for each model from LID-FIT yields Figures 15 through 17, which provide a birds-eye view of daily estimated differences between lidar measurements and each reanalysis model.

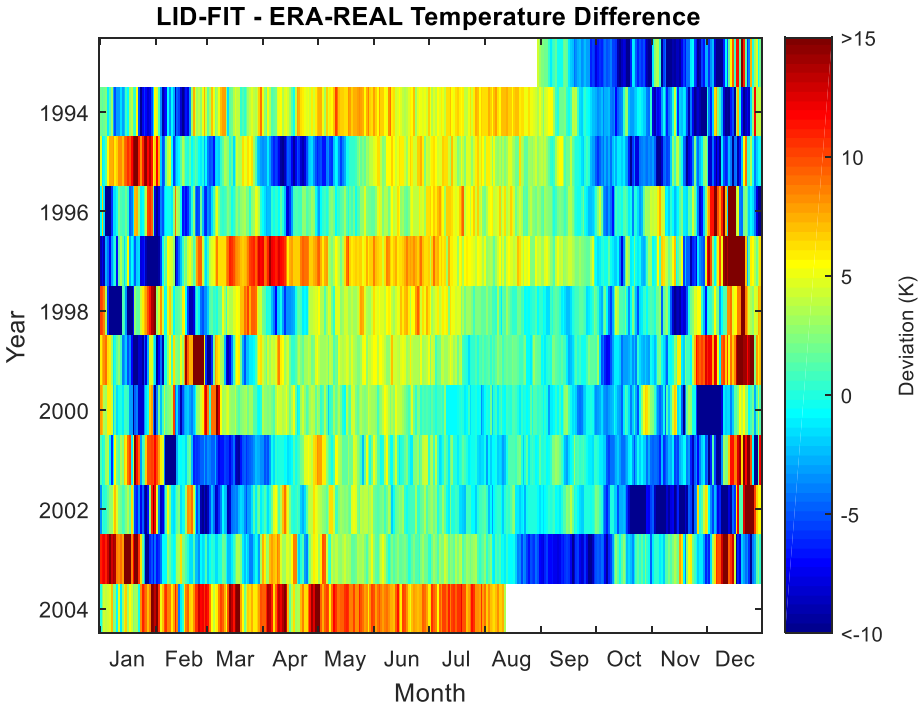


Figure 15 – Difference between fitted lidar temperatures and ERA-20C output.

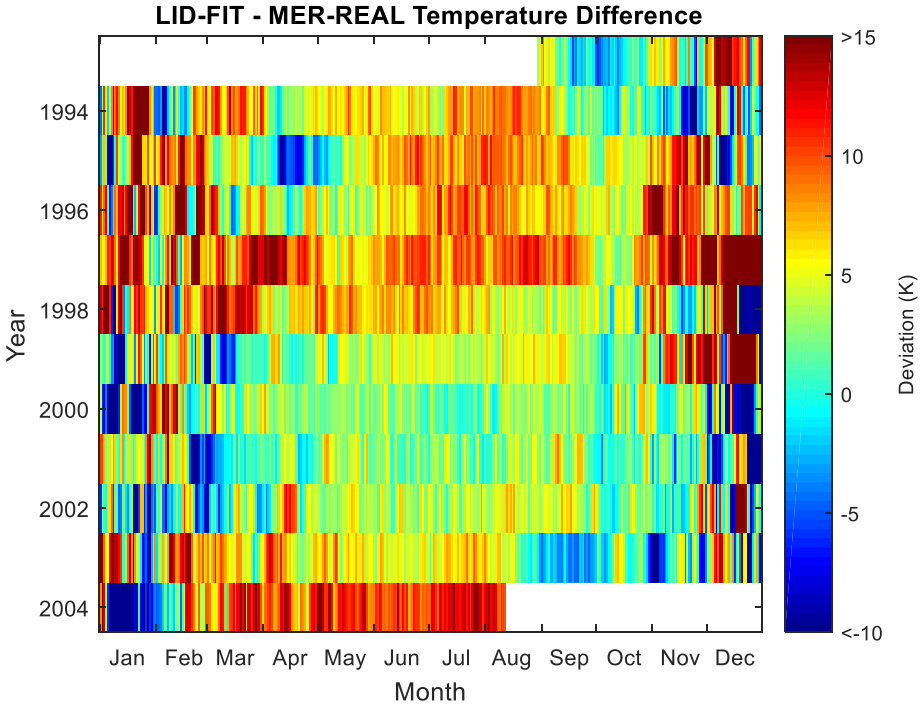


Figure 16 – Difference between fitted lidar temperatures and MERRA-2 output.

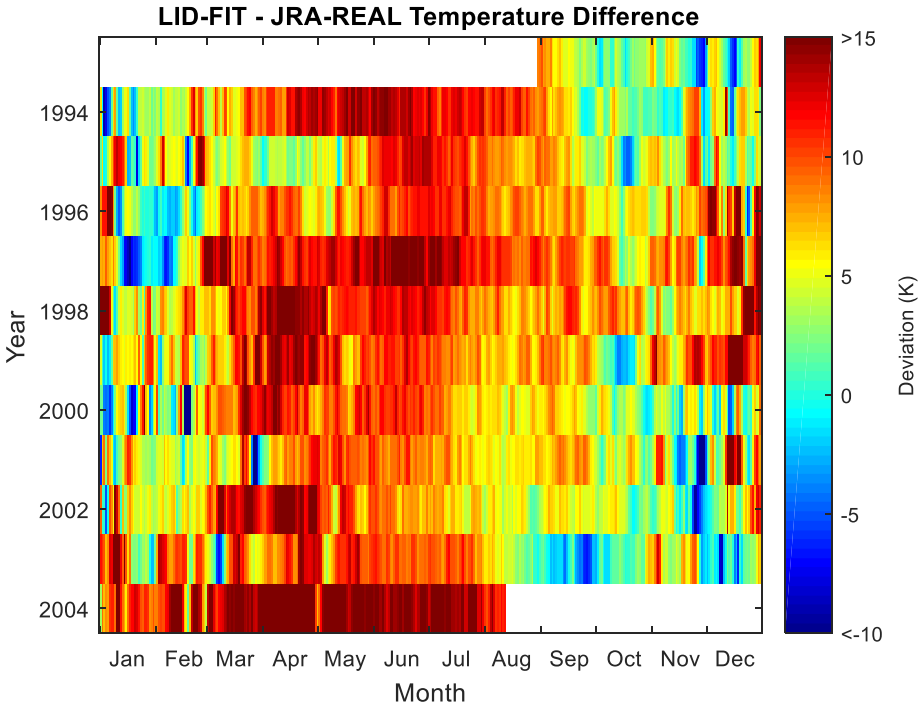


Figure 17 – Difference between fitted lidar temperatures and JRA-55 output.

This finer look at the differences appears to correspond well with the composite years for each respective dataset. As the composite years in Figures 13 and 14 illustrate, LID-FIT typically dips below LID-REAL in the fall, which makes ERA-REAL appear comparatively warmer than the lidar on Figure 15 during that season on most years. This anomaly isn't duplicated in the comparison with the other REAL datasets, however. By breaking down the differences on a day to day basis we have also revealed a significant amount of interannual variability that is lost in the formulation of a composite year. Especially noticeable are the years 2001 and 2004, where the lidar seems to be respectively colder and warmer than appears typical. Aside from the additional detail and an occasional spike or dip here and there, the broad conclusions arrived at from the composite years generally hold true. However, one should be careful to note that the fit is least accurate in the winter months, tending to smooth over real geophysical variability. The winter season also corresponds with the largest temporal gaps in lidar data where LID-FIT is less guided by real measurements. Therefore, many differences present in Figures 15 through 17 during these times could very likely be due to inaccurate fitting rather than an actual difference that the lidar may have otherwise captured had it been operable on those days. It may be more appropriate to compare FIT datasets directly, rather than subtract a REAL from a FIT. This has been done for Figures 18 through 20 on the following pages.

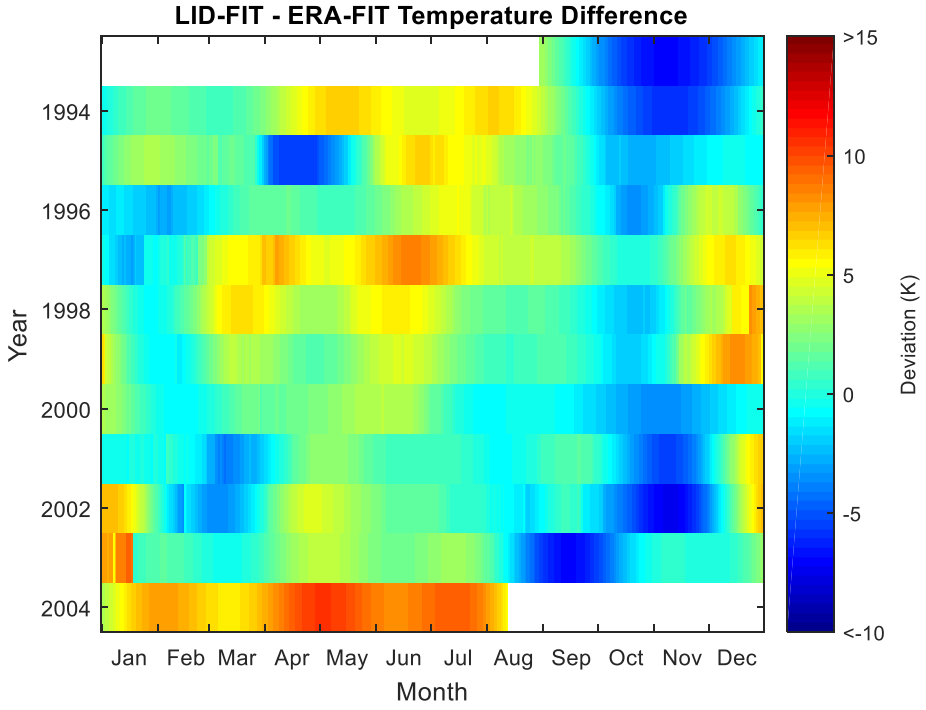


Figure 18 – Difference between fitted lidar temperatures and fitted ERA-20C output.

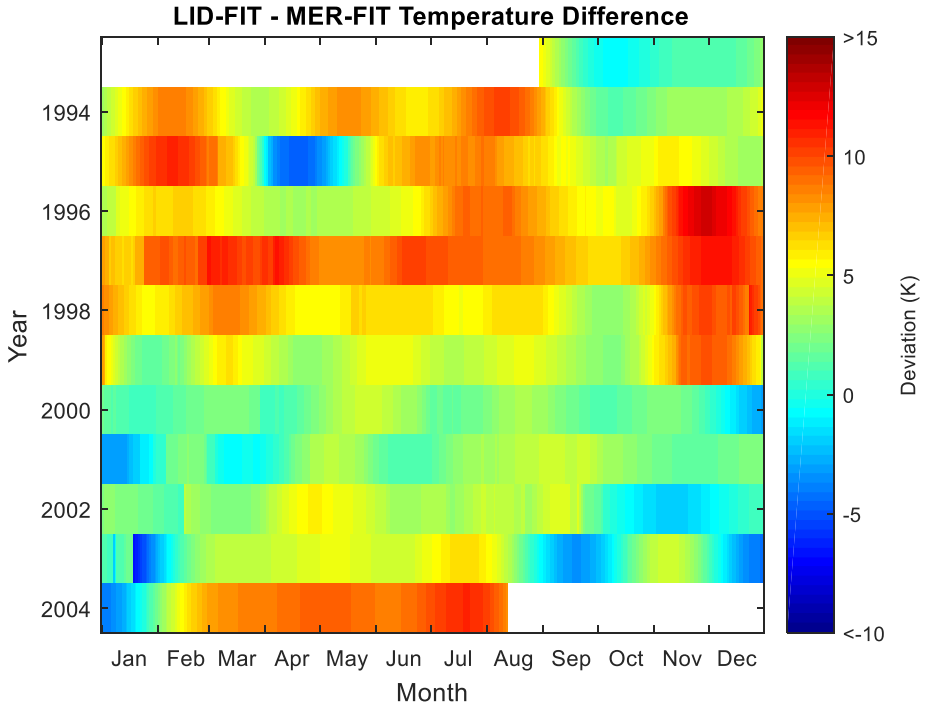


Figure 19 – Difference between fitted lidar temperatures and fitted MERRA-2 output.

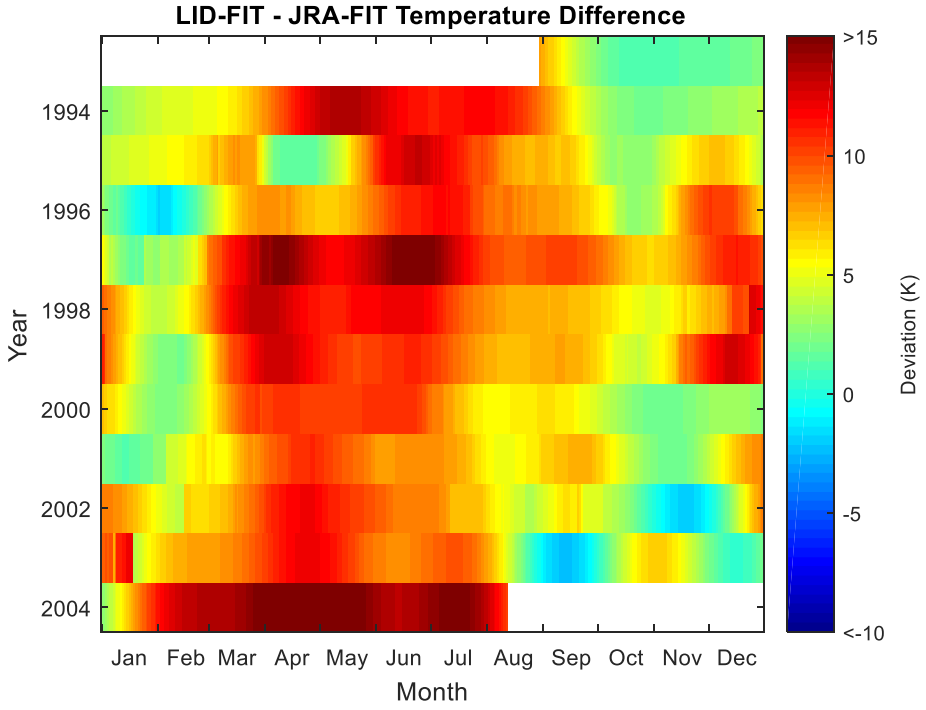


Figure 20 – Difference between fitted lidar temperatures and fitted JRA-55 output.

The data smoothing caused by the fit method is now significantly more noticeable, which turns out to be useful in making the general trends easier to resolve at a glance. There are several temperature oscillations in the fit differences that seem abnormal, and worth more study, such as the cold lidar temperatures from April-May 1995. The years 1999-2002, close to the solar cycle maximum, also appear to have generally smaller differences, possibly indicating larger solar-cycle-related changes than the models are accounting for. This is, of course, still a bird's-eye view, and it serves as a starting point for further research more than leading quickly to any new conclusions.

The most valuable and far-reaching results of this study come from the monthly means of the datasets, for which confidence intervals on the differences have been

calculated as discussed in the previous chapter. These are comparisons of the actual lidar measurements with the output from the models, and no fitting is involved whatsoever. In Figures 21 through 23 a colored grid is displayed. These are arranged like the previous few sets of figures, except there is only one value per month: the REAL model dataset monthly average subtracted from the corresponding LID-REAL monthly average. In addition to the color representing the difference, a hatching pattern has been applied to each monthly box to illustrate the confidence level, i.e. how sure we are that the difference cannot be ascribed to any associated uncertainty. The solid color has been left as-is without any hatching pattern applied whenever the p-value for that difference is less than 0.01, indicating more than 99% confidence in the reality of that difference.

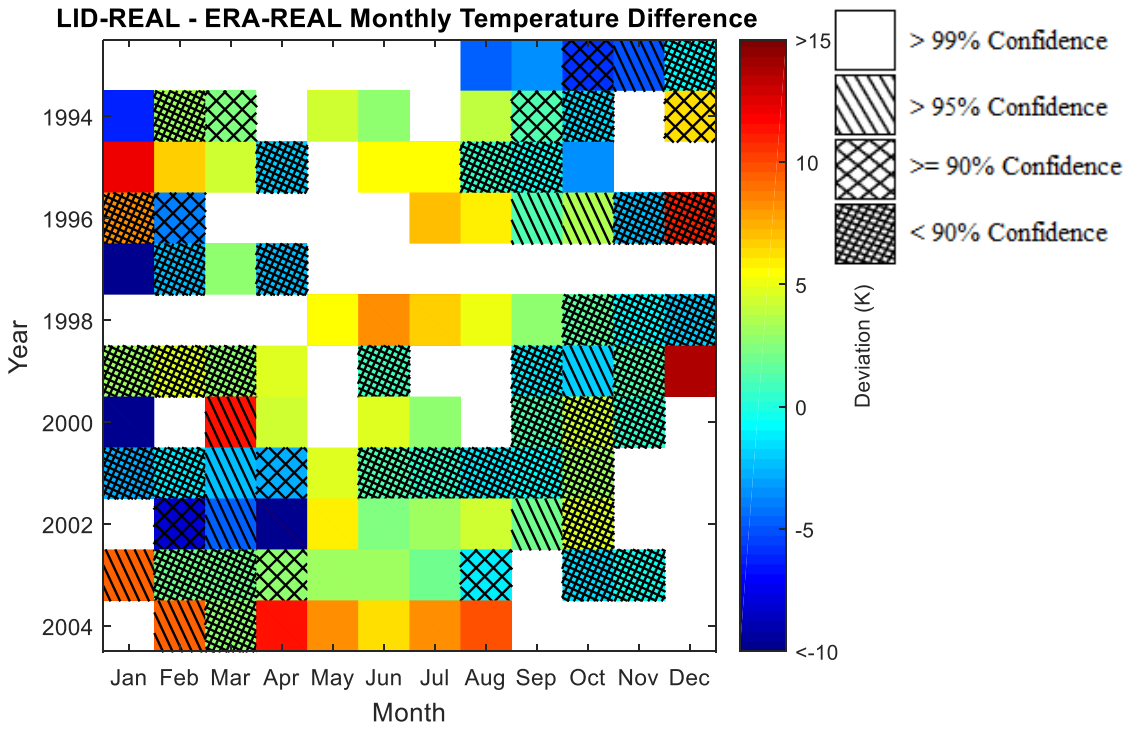


Figure 21 – Monthly mean differences between lidar temperatures and ERA-20C output.

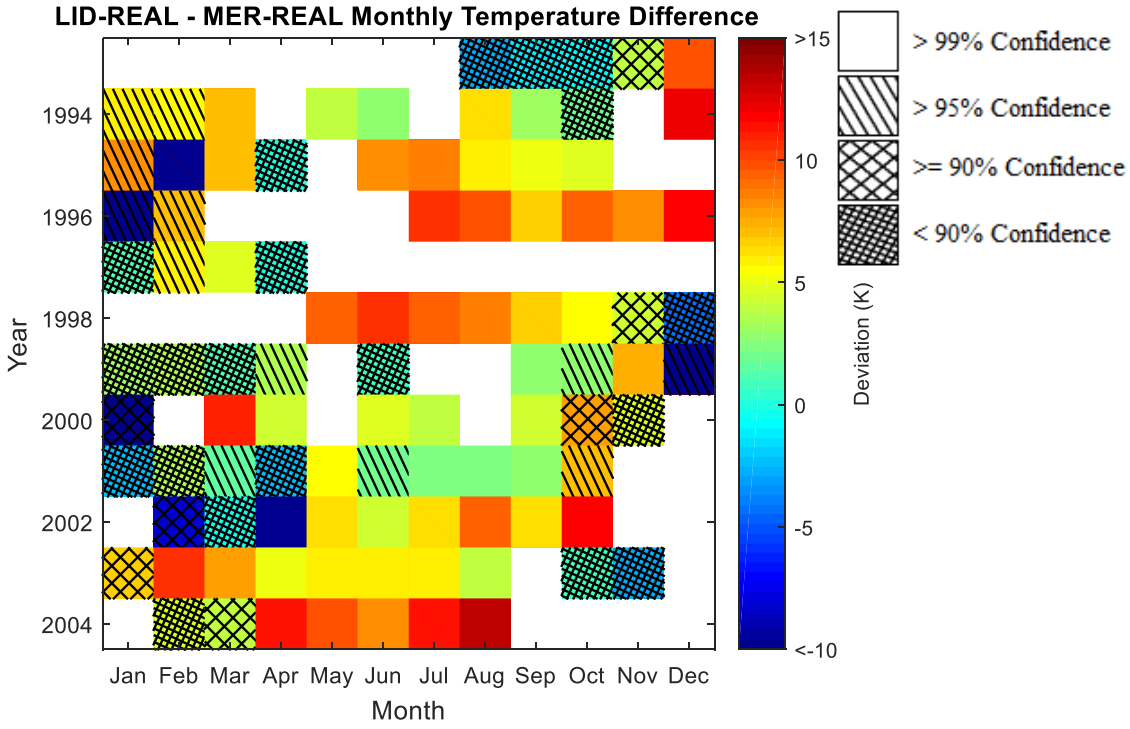


Figure 22 – Monthly mean differences between lidar temperatures and MERRA-2 output.

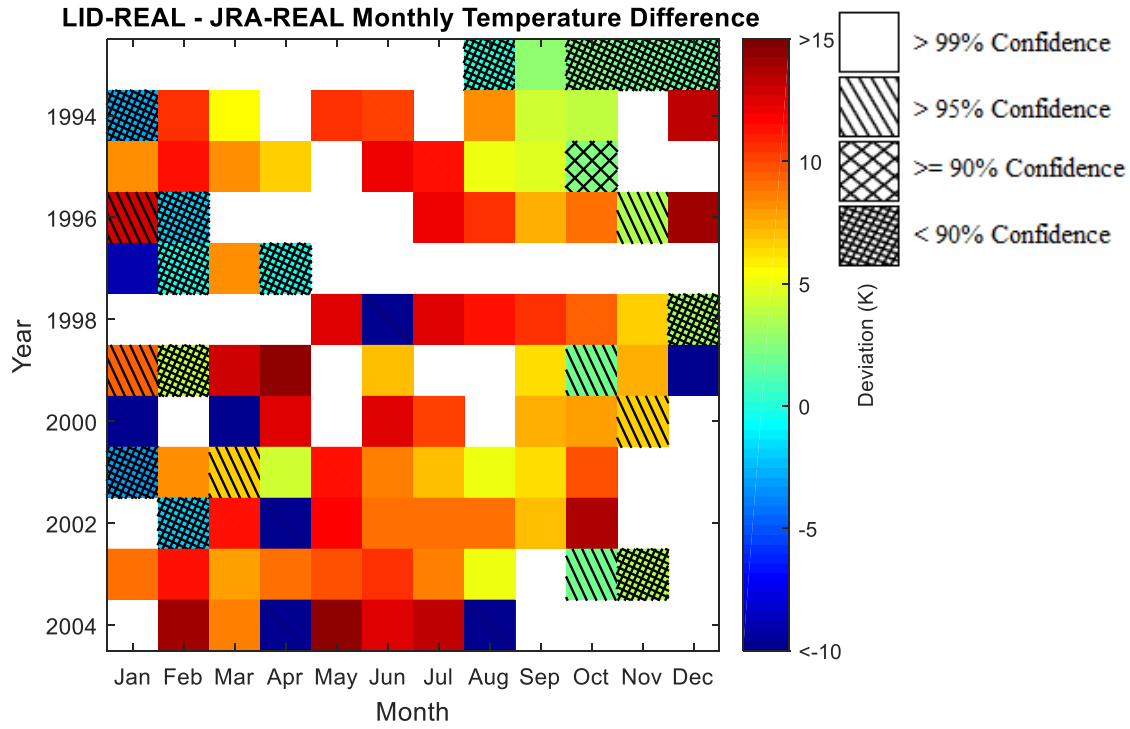


Figure 23 – Monthly mean differences between lidar temperatures and JRA-55 output.

The combination of an increase in geophysical variability and decrease in number of lidar measurements during the winter appears to remain the greatest source of uncertainty, whether a fit is used or not. This fact is indicated by the increase in hatching patterns outside of summer months, and it correlates quite well with the distribution of fit outliers as given in Figure 12. Regardless, there are some impressive and statistically significant deviations from the lidar measurements in all three models, even at these times when lidar measurements are at their most sparse. Table 4 displays the values for all winter months (December, January, & February) with 95% confidence or more in all three model differences, excepting where one of the three model differences only exceeds 90% confidence, in which case those are marked by an asterisk. It is notable that in this example only the February 1994 differences show any model producing temperatures higher than that of the ALO-USU measurements.

Table 4

Winter Temperature Difference Highlights

Month	Model difference from lidar (K)		
	ERA-20C	MERRA-2	JRA-55
February 1994	-6.14*	6.12	-2.78
December 1994	6.63*	12.30	13.65
January 1995	12.47	8.34	8.28
February 1995	6.79	16.01	11.56
December 1999	14.03	19.85	16.17
January 2000	25.61	20.64*	26.98
January 2003	9.61	6.96*	9.34
February 2003	2.17	10.98*	11.55

While working toward these results, one concern was over possible inaccuracies due to the disparate number of measurements between LID-REAL and any model dataset. Although the entire purpose of performing the two-sided pooled t-procedure was to give a statistically robust answer to this question, the RLID datasets also exist to corroborate the accuracy of the comparisons. Since the RLID datasets only contain data co-temporal with lidar measurements there is no discrepancy in measurement number. Performing all the same calculations, and charting the data in the same way as Figures 21 through 23, we can produce Figures 24 through 26.

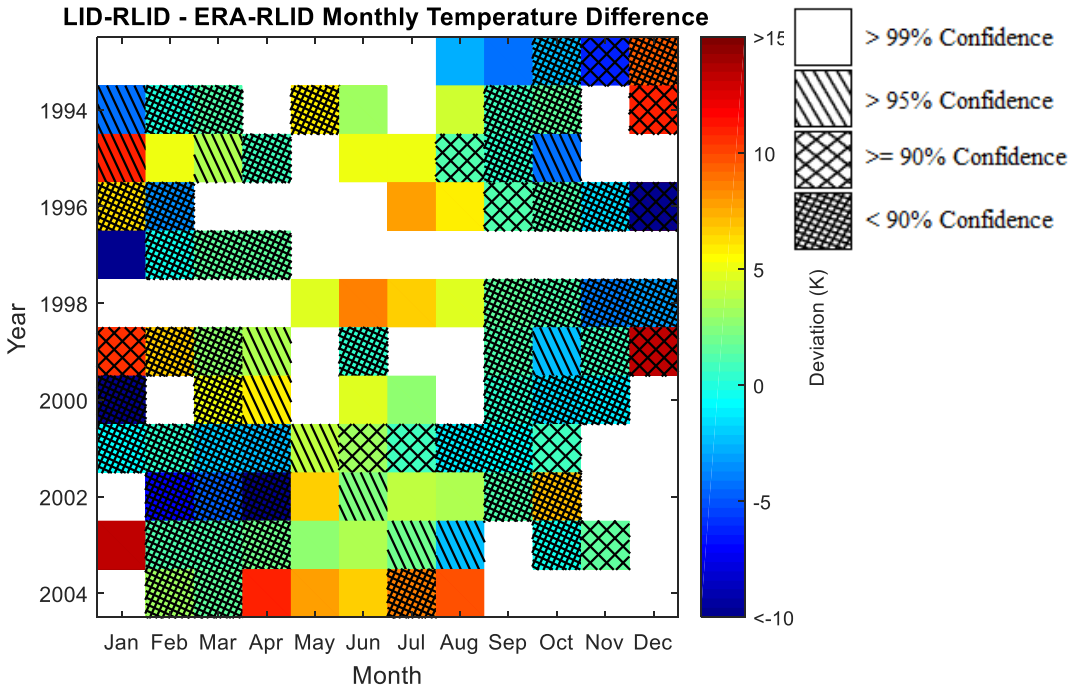


Figure 24 – Monthly mean differences between lidar temperatures and exclusively co-temporal ERA-20C output alone.

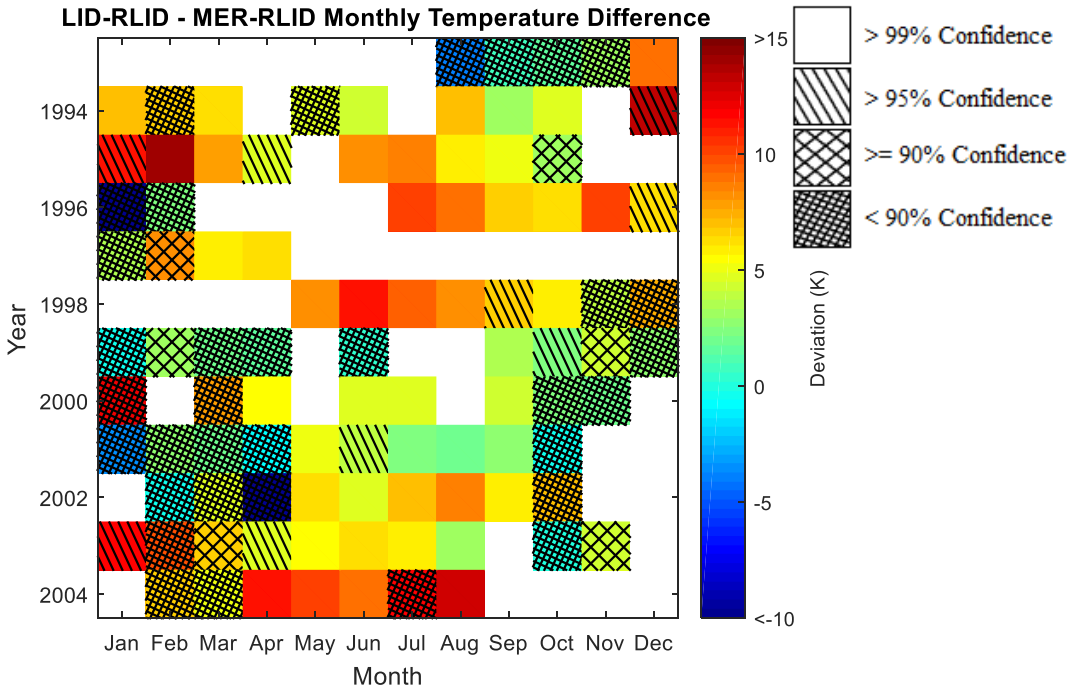


Figure 25 – Monthly mean differences between lidar temperatures and exclusively co-temporal MERRA-2 output alone.

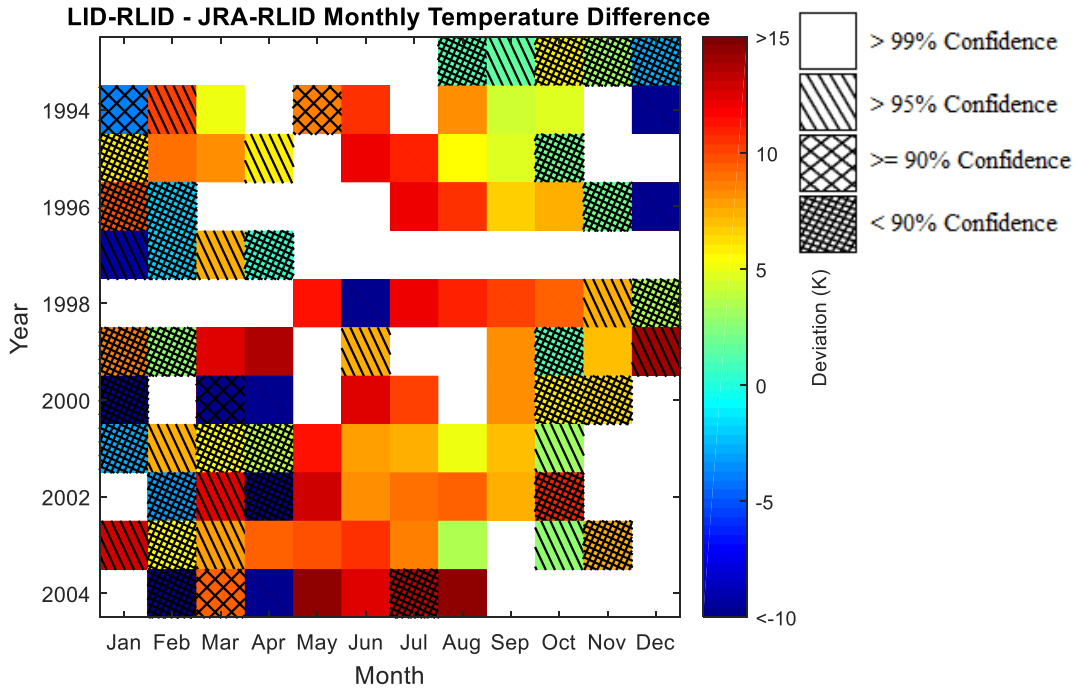


Figure 26 – Monthly mean differences between lidar temperatures and exclusively co-temporal JRA-55 output alone.

Some predictable small dissimilarities exist between the FLID difference and the FULL difference, simply due to content. The most important change between the two is the fact that the FLID comparison shows lower confidence in the differences in general. This is an expected consequence of having fewer total measurements involved. As N increases, the standard deviation decreases. Therefore, the statistical validity of the FULL difference makes it the more accurate comparison. Regardless, the FLID differences confirm much of what has already been said and add further weight to the conclusions.

Lastly, perhaps of interest to some readers is a straightforward presentation of the monthly averages themselves rather than the differences. These have been provided in Figures 27 through 30 below for all four FULL datasets. The most noteworthy thing about these is how they demonstrate a veritable lack of interannual variability in the models when compared with the genuine lidar measurements. Once again, those curious about the RLID, FIT, or FLIT data can find the complete data tables in Appendix C.

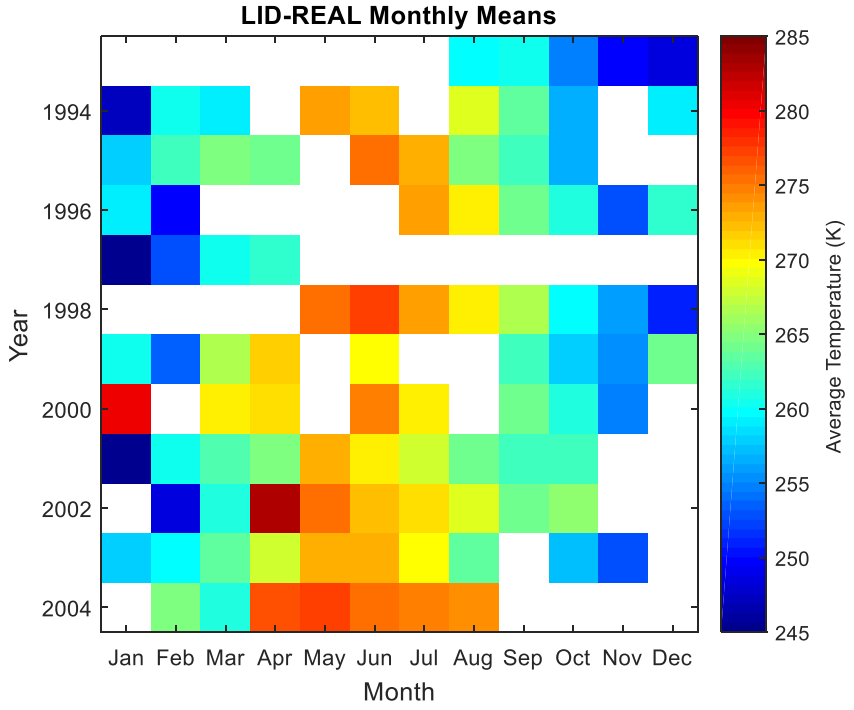


Figure 27 – Lidar measurement monthly averages.

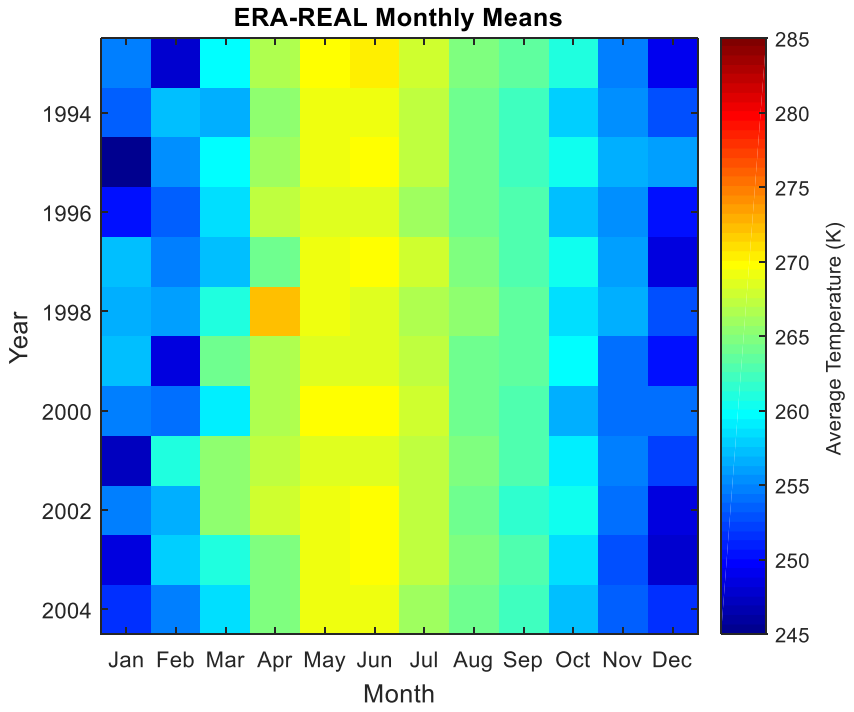


Figure 28 – ERA-20C output monthly averages.

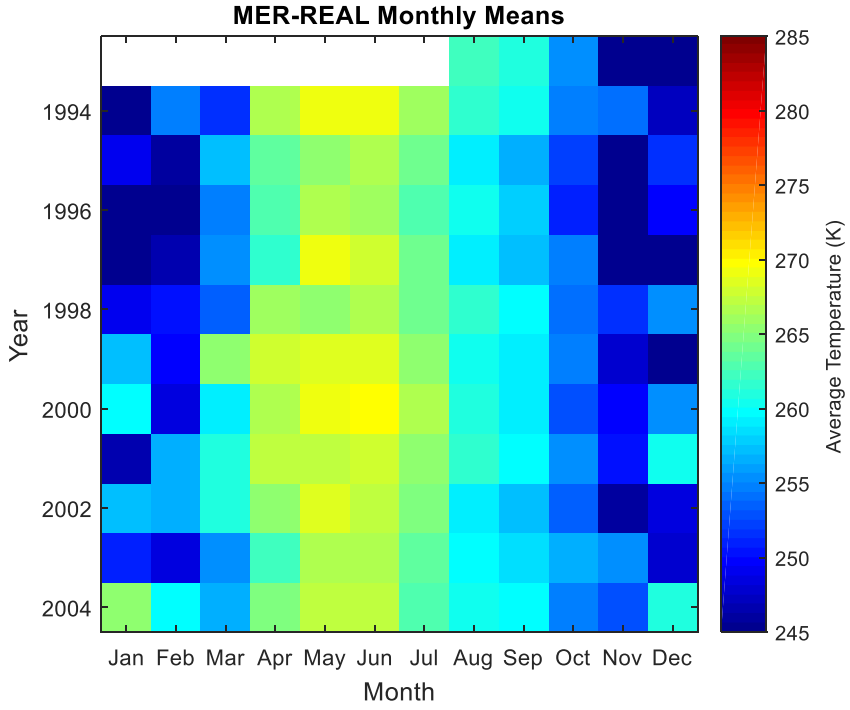


Figure 29 – MERRA-2 output monthly averages.

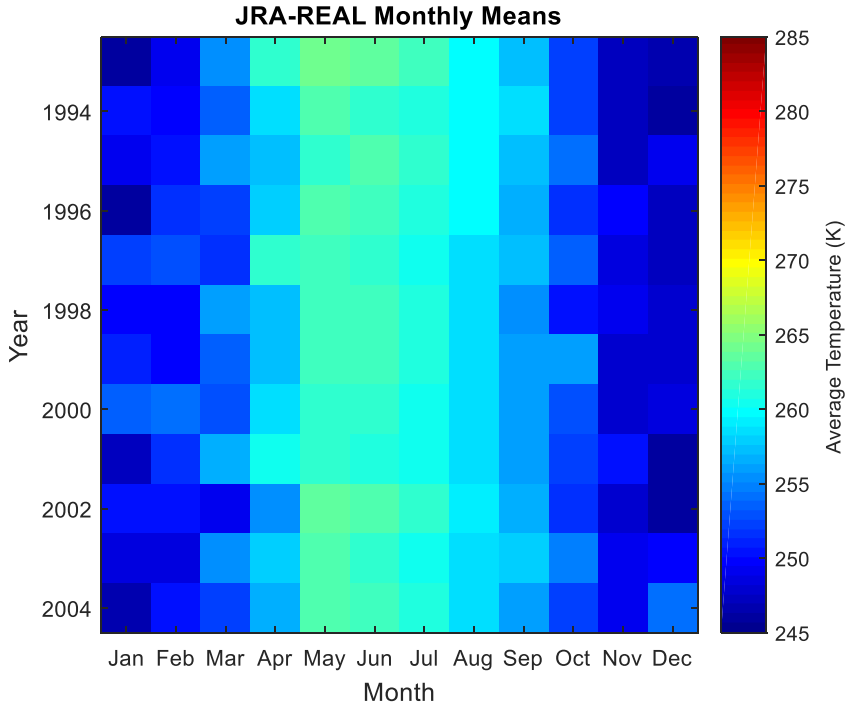


Figure 30 – JRA-55 output monthly averages.

DISCUSSION & CONCLUSION

A common practice in long-term upper-atmospheric studies is to apply spectral or signal analysis techniques such as Fourier transforms or Lomb-Scargle periodograms to examine the phases and amplitudes of a variety of wave phenomena. As explained at the beginning of the previous chapter, the adaptive application of the Levenberg-Marquardt fit using equation 3.2, while an interesting new way to approach the examination of the data, doesn't allow for the simple cross-comparison of the parameters that is so often done in similar studies. Additional work of this nature and for the express purpose of comparing with other studies was not performed at this time.

However, the values may still be comparable for the average parameters B and K from equation 3.2 that were listed in Table 3. These results, representing the yearly amplitude and linear trends respectively, mark a departure from other studies examining similar parameters at near-stratopause altitudes. The average yearly amplitude of 10.61 ± 1.53 is larger than the yearly amplitudes of any other study at this altitude and latitude, with most putting the amplitude in the range of 6-7 K (Leblanc et al., 1998). Indeed, the ALO-USU composite year climatology ranges from 250-275 K, a broader range by a few degrees than both the maximum and minimum temperatures identified by Argall and Sica in 2007. While the parameters for the models are lower, they are still a bit higher than previous studies. This consistency seems to indicate either that a larger yearly amplitude at our location is normal, or that averaging the parameters is not the best approach for comparison. It isn't out of the question to suggest that local geographic features such as the Rocky Mountains cause longitude-dependent wave activities not found to such a

degree elsewhere at this latitude, which could certainly be responsible for pushing temperatures near the stratopause both higher and warmer as the waves oscillate. Such physical activity has also been shown to create a double stratopause, which may lead to abnormal temperature fluctuations around the 45 km altitude examined here (Sivakumar et al., 2006).

As for the linear trend parameter, Batista et al. (2009) found a small negative trend between 40 and 50 km of roughly -1.5 K per decade, although their low-latitude location (23° S) may be responsible for the difference. Unfortunately, other studies looking for a linear trend at 45km have been inconclusive (Hauchecorne et al., 1991). The fit parameter average for the ALO-USU data works out to 1.62 ± 1.98 K per year, indicating the likelihood of a small but fickle increase in temperatures at 45 km over the period of study. On the other hand, this calculation may be artificially inflated by the combination of the nature of the moving fit and the relative lack of lidar data during colder periods. This reasoning is additionally supported by the facts that the same parameter yields very neutral trends for the other datasets, as does the alternative trend approach of applying a simple linear fit to the data. Overall, these parameter averages should probably only be used with some understanding of the source and a decent helping of restraint. Future work could include analyzing the distribution of the parameter values to see if there are any further insights to be found there.

Another prime candidate for future study would be the outliers from the fit. Specifically, determining their causes and examining their impacts in the vertical altitude dimension across a few neighboring days could easily yield several years of work. A very

basic proof-of-concept of a possible procedure for doing this was put together for presentation at the 2015 NCAR Advanced Study Program summer colloquium (Moser, Wickwar, Navarro, & Herron, 2015). To summarize that short study, a cluster of outliers common to all four datasets occurred in December 1998. Closer examination revealed a 50 K amplitude wave taking place over several days around the stratopause. In looking for a source of the wave, surface-level weather maps and observational data indicated the concurrent passage of a strong cold front and winter storm system that saw surface temperatures drop 15 K in less than 24 hours. Additionally, Sox et al. (2016) identified strong SSW activity during the same period in the ALO-USU lidar profiles. While the examination of outliers didn't move beyond this type of associative work, it should be easy to see the potential for studying and identifying atmospheric phenomena, and especially coupling mechanisms between atmospheric layers that are likely to be tied to the presence of outliers.

Finally, the widely-applicable crux of this thesis is that of the temperature differences between reanalysis model output and the lidar temperature measurements. The January & February months for both 1995 and 2003 in Table 4 happen to coincide with SSW events observed to have impacts at our location by Sox et al. (2016). It isn't a stretch to assume that this could be the reason for those particular differences, and that other geophysical phenomena (e.g. gravity waves, Baumgarten et al., 2018; planetary waves, Beissner, 1997; tides, Sakazaki et al., 2018; and double stratopause, Sivakumar et al., 2006) could similarly drive differences at some other times.

As has been suggested by Sakazaki et al. (2018), the relatively short timescales on which some atmospheric tides and gravity wave activity can take place lead to a variety of effects at 45 km that modern models don't have the temporal resolution to accurately simulate. It is very likely that the location of ALO-USU near the Rocky Mountains makes for a much more dynamic atmosphere than at other longitudes, but this would need to be confirmed by studies similar to this one in scope and style at different longitudes but the same latitude across the globe. One candidate for this is the Rayleigh-Mie-Raman lidar at Observatory of Haute Provence (44° N) in France. They have actually done a similar study previously, but across a different and much shorter time frame (2012-2013), and the only model comparison they performed in common with this thesis was using ERA-20C (Le Pichon et al., 2015). Their 95% distribution exceeds 30 K in the 40-60 km range, rendering any conclusive comparison between lidar measurement and model output questionable at best, but the overall trend seen here where the models tend to be systematically colder than the lidar still appears to be somewhat present. They do take the opportunity to single out SSWs as being underrepresented in the models, but they also point to the fact that large-scale planetary waves generally have more variability during winter months as another source of model deviation from measurements.

Schöch et al. (2008) also compared lidar data with an ECMWF product that they don't identify clearly in their paper, although their lidar is also at a much higher latitude (63.9° N) than ALO-USU. The 45 km altitude shows no significant differences between the model and the lidar at their location, suggesting that (assuming their ECMWF product is a modern one) the models could be functioning more accurately closer to the poles than

at mid-latitudes. The most common lidar comparisons are done between lidar measurements and common empirical reference models such as the Committee on Space Research International Reference Atmosphere (CIRA) and the Mass-Spectrometer-Incoherent-Scatter model (MSIS) (Batista et al., 2009; Hauchecorne et al., 1991). Even these studies find their measurements to be mostly in agreement with models. If anything, it is more common for the 45 km temperatures to measure colder than the reference atmospheres (Argall & Sica, 2007).

In contrast to all of these previous studies, we have very high confidence in large systematic differences over a long period of time, mostly on the side of the lidar reading hotter than the models. Surprisingly, the most confident and consistent result of this thesis is that popular modern models are nearly always too cold at 45 km in the summer months at our location. This is also when geophysical variability is at its lowest and simulations should be performing without many issues. Although middle-atmospheric geophysical phenomena are certainly taking place to some degree at all times of the year, these highly significant differences in summer manifest so reliably that they cannot possibly be attributed to the models' inability to resolve such phenomena in the temporal domain. Instead, this result points to either a serious deficiency in the modeling of prevailing physics and chemistry components, or the dearth of assimilated measurements used to validate the model output in the middle atmosphere. Pawson et al. (2000) studied the stratospheric processes involved in 13 different middle atmosphere climate models, and appears to confirm the latter hypothesis by stating:

> "Although all 13 models evaluated represent most major features of the mean atmospheric state, there are deficiencies in the magnitude and location of the features, which cannot easily be traced to the formulation (resolution or the parameterizations included) of the models."

As if to further endorse the results of this thesis, the study goes on to show a cold bias in most models at almost all locations and vertical levels. Despite the fact that the Pawson study took place nearly 20 years ago, this still appears to be a significant problem.

As the major reanalysis models move toward whole-atmosphere simulation (Fujiwara et al., 2017; Molod et al., 2015), they are very likely incorporating the same elements used in the middle atmosphere models examined by Pawson et al. (2000), and suffering the same results at altitudes where the options for data assimilation are relatively sparse. Although most of the observational data assimilated by the models in the middle atmosphere is provided by temporally-reliable satellite measurement, Hoppel et al. (2013) has indicated that the peak of satellite measurement accuracy occurs at altitudes below 35 km. These measurements have been shown to improve considerably when satellite networks are upgraded with newer technology (Keckhut et al., 2001), but even very recent studies have confirmed that modern satellite measurements are not nearly enough to provide the required level of model validation (Sakazaki et al., 2018; Wing et al., 2018). This problem can only grow in size as model resolution continues to improve. In order to achieve the verifiable accuracy the groups behind the models are surely striving for, they will need to greatly expand the amount and variety of measurement data used as inputs to their products.

The ALO-USU Rayleigh-scatter lidar has provided high-precision temperature measurements for many years. The instrument has been upgraded to achieve higher altitude measurements since the data used in this study was recorded, and a sodium lidar has joined ALO-USU as well. The initial stages of an expansion to collect data lower into the stratosphere have also taken place. It is hoped that future work can include an extension of this thesis; examining outliers in detail and fine-tuning the fit method and its applications. Although modern reanalysis models are constantly improving, systematic issues with middle-atmospheric simulation continue to exist. This will be an increasingly important problem to solve as the models continue to expand in scope and incorporate higher altitudes. It is clear that current reanalyses aren't assimilating enough measurements at these altitudes to provide for an adequate validation of model output. In comparing the ALO-USU Rayleigh-scatter lidar data with the reanalysis models it has been demonstrated that large and significant differences exist between the models and measurements at our location. While this includes several anomalous dates that are probably attributable to geophysical variability, there is also a persistent cold bias in the models during summer time when such an attribution is much less plausible. Truly, modelers would do well to begin assimilating datasets like that provided by ALO-USU that cover the gaps in their data to directly address problem-areas. This applies particularly at the lower altitudes of the lidar data integration where the measurement errors become irrelevant.

REFERENCES

Angell, J. K. (1988). Variations and Trends in Tropospheric and Stratospheric Global Temperatures, 1958–87. *Journal of Climate*, 1, 1296–1313, https://doi.org/10.1175/1520-0442(1988)001<1296:VATITA>2.0.CO;2

Argall, P. S., & Sica, R. J. (2007). A comparison of Rayleigh and sodium lidar temperature climatologies. *Annales Geophysicae*, 25(1), 27–35. https://doi.org/10.5194/angeo-25-27-2007

Barton, D. L., Wickwar, V. B., Herron, J. P., Sox, L., & Navarro, L. A. (2016). Variations in Mesospheric Neutral Densities from Rayleigh Lidar Observations at Utah State University. *EPJ Web of Conferences*, 119, 13006. https://doi.org/10.1051/epjconf/201611913006

Batista, P. P., Clemesha, B. R., & Simonich, D. M. (2009). A 14-year monthly climatology and trend in the 35–65km altitude range from Rayleigh Lidar temperature measurements at a low latitude station. *Journal of Atmospheric and Solar-Terrestrial Physics*, 71(13), 1456–1462. https://doi.org/10.1016/j.jastp.2008.03.005

Baumgarten, K., Gerding, M., Baumgarten, G., & Lübken, F.-J. (2018). Temporal variability of tidal and gravity waves during a record long 10-day continuous lidar sounding. *Atmospheric Chemistry and Physics*, 18(1), 371–384. https://doi.org/10.5194/acp-18-371-2018

Beissner, K. C. (1997). *Studies of Mid-lattitude Mesospheric Temperature Variability and Its Relationship to Gravity Waves, Tides, and Planetary Waves* (Doctoral dissertation). Retrieved from Digital Commons (https://digitalcommons.usu.edu/etd/4687). Logan, UT: Utah State University.

Bevington, P. R. (1969). *Data reduction and error analysis for the physical sciences*. New York, NY: McGraw-Hill.

Bucholtz, A. (1995). Rayleigh-scattering calculations for the terrestrial atmosphere. *Applied Optics*, 34(15), 2765. https://doi.org/10.1364/ao.34.002765

Cai, X., Yuan, T., Zhao, Y., Pautet, P.-D., Taylor, M. J., & Pendleton, W. R. (2014). A coordinated investigation of the gravity wave breaking and the associated dynamical instability by a Na lidar and an Advanced Mesosphere Temperature Mapper over Logan, UT (41.7°N, 111.8°W). *Journal of Geophysical Research: Space Physics*, 119(8), 6852–6864. https://doi.org/10.1002/2014ja020131

Chanin, M. L. (1984). Review of lidar contributions to the description and understanding of the middle atmosphere. *Journal of Atmospheric and Terrestrial Physics*, 46(11), 987–993. https://doi.org/10.1016/0021-9169(84)90005-9

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., et al. (2013). On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *Journal of Geophysical Research: Atmospheres*, 118(6), 2494–2505. https://doi.org/10.1002/jgrd.50125

Eckert, E., von Clarmann, T., Kiefer, M., Stiller, G. P., Lossow, S., Glatthor, N., et al. (2013). Drift-corrected trends and periodic variations in MIPAS IMK/IAA ozone measurements. *Atmospheric Chemistry and Physics Discussions*, 13(7), 17849–17900. https://doi.org/10.5194/acpd-13-17849-2013

Finger, F. G., Gelman, M. E., Wild, J. D., Chanin, M. L., Hauchecorne, A., & Miller, A. J. (1993). Evaluation of NMC Upper-Stratospheric Temperature Analyses Using Rocketsonde and Lidar Data. *Bulletin of the American Meteorological Society*, 74, 789–800. https://doi.org/10.1175/1520-0477(1993)074<0789:EONUST>2.0.CO;2

Fujiwara, M., Wright, J. S., Manney, G. L., Gray, L. J., Anstey, J., Birner, T., et al. (2017). Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmospheric Chemistry and Physics*, 17(2), 1417–1452. https://doi.org/10.5194/acp-17-1417-2017

Gardner, C. S., Miller, M. S., & Liu, C. H. (1989). Rayleigh Lidar Observations of Gravity Wave Activity in the Upper Stratosphere at Urbana, Illinois. *Journal of the Atmospheric Sciences*, 46, 1838–1854, https://doi.org/10.1175/1520-0469(1989)046<1838:RLOOGW>2.0.CO;2

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454. https://doi.org/10.1175/jcli-d-16-0758.1

Gerber, E. P., Butler, A., Calvo, N., Charlton-Perez, A., Giorgetta, M., Manzini, E., et al. (2012). Assessing and Understanding the Impact of Stratospheric Dynamics and Variability on the Earth System. *Bulletin of the American Meteorological Society*, 93(6), 845–859. https://doi.org/10.1175/bams-d-11-00145.1

Hågård, A., & Persson, R. (1997). Aerosol Extinction Measurements with Lidar at 10.6 and 1.5 μm. *Advances in Atmospheric Remote Sensing with Lidar*, 51–54. https://doi.org/10.1007/978-3-642-60612-0_13

Harada, Y., Kamahori, H., Kobayashi, C., Endo, H., Kobayashi, S., Ota, Y., et al. (2016). The JRA-55 Reanalysis: Representation of Atmospheric Circulation and Climate Variability. *Journal of the Meteorological Society of Japan*, Ser. II, 94(3), 269–302. doi:10.2151/jmsj.2016-015

Hauchecorne, A., & Chanin, M.-L. (1980). Density and temperature profiles obtained by lidar between 35 and 70 km. *Geophysical Research Letters*, 7(8), 565–568. https://doi.org/10.1029/gl007i008p00565

Hauchecorne, A., Chanin, M.-L., & Keckhut, P. (1991). Climatology and trends of the middle atmospheric temperature (33–87 km) as seen by Rayleigh lidar over the south of France. *Journal of Geophysical Research*, 96(D8), 15297. https://doi.org/10.1029/91jd01213

Hersbach, H., Peubey, C., Simmons, A., Berrisford, P., Poli, P., & Dee, D. (2015). ERA-20CM: a twentieth-century atmospheric model ensemble. *Quarterly Journal of*

the Royal Meteorological Society, 141(691), 2350–2375. https://doi.org/10.1002/qj.2528

Herron, J. P. (2004). *Mesospheric Temperature Climatology Above Utah State University* (Master's thesis). Retrieved from Digital Commons (https://digitalcommons.usu.edu/etd/6877). Logan, UT: Utah State University.

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* (Doctoral dissertation). Retrieved from Digital Commons (https://digitalcommons.usu.edu/etd/4686). Logan, UT: Utah State University.

Herron, J. P., & Wickwar, V. B. (2018). Mid-Latitude Climatologies of Mesospheric Temperature and Geophysical Temperature Variability Determined with the Rayleigh-Scatter Lidar at ALO-USU. *All Physics Faculty Publications*. Paper 2063. https://digitalcommons.usu.edu/physics_facpub/2063

Hoppel, K. W., Eckermann, S. D., Coy, L., Nedoluha, G. E., Allen, D. R., Swadley, S. D., & Baker, N. L. (2013). Evaluation of SSMIS Upper Atmosphere Sounding Channels for High-Altitude Data Assimilation. *Monthly Weather Review*, 141(10), 3314–3330. https://doi.org/10.1175/mwr-d-13-00003.1

Jenkins, D. B., Wareing, D. P., Thomas, L., & Vaughan, G. (1987). Upper stratospheric and mesospheric temperatures derived from lidar observations at Aberystwyth.

Journal of Atmospheric and Terrestrial Physics, 49(3), 287–298. https://doi.org/10.1016/0021-9169(87)90064-x

Keckhut, P., Hauchecorne, A., & Chanin, M. L. (1993). A Critical Review of the Database Acquired for the Long-Term Surveillance of the Middle Atmosphere by the French Rayleigh Lidars. *Journal of Atmospheric and Oceanic Technology*, 10, 850–867, https://doi.org/10.1175/1520-0426(1993)010<0850:ACROTD>2.0.CO;2

Keckhut, P., Hauchecorne, A., & Chanin, M. L. (1995). Midlatitude long-term variability of the middle atmosphere: Trends and cyclic and episodic changes. *Journal of Geophysical Research*, 100(D9), 18887. https://doi.org/10.1029/95jd01387

Keckhut, P., Wild, J. D., Gelman, M., Miller, A. J., & Hauchecorne, A. (2001). Investigations on long-term temperature changes in the upper stratosphere using lidar data and NCEP analyses. *Journal of Geophysical Research: Atmospheres*, 106(D8), 7937–7944. https://doi.org/10.1029/2000jd900845

Khaykin, S. M., Hauchecorne, A., Mzé, N., & Keckhut, P. (2015). Seasonal variation of gravity wave activity at midlatitudes from 7 years of COSMIC GPS and Rayleigh lidar temperature observations. *Geophysical Research Letters*, 42(4), 1251–1258. https://doi.org/10.1002/2014gl062891

Kim, Y., Eckermann, S. D., & Chun, H. (2003). An overview of the past, present and future of gravity-wave drag parametrization for numerical climate and weather

prediction models. *Atmosphere-Ocean*, 41(1), 65–98. https://doi.org/10.3137/ao.410105

Kirgis, G., Leblanc, T., McDermid, I. S., & Walsh, T. D. (2012). Stratospheric ozone interannual variability (1995–2011) as observed by Lidar and Satellite at Mauna Loa Observatory, HI and Table Mountain Facility, CA. *Atmospheric Chemistry and Physics Discussions*, 12(11), 30825–30867. https://doi.org/10.5194/acpd-12-30825-2012

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan*, Ser. II, 93(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Koster, R. D., McCarty, W., Coy, L., Gelaro, R., Huang, A., Merkova, D., et al. (2016). *MERRA-2 Input observations: Summary and assessment*. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160014544.pdf

Krüger, K., Naujokat, B., & Labitzke, K. (2005). The Unusual Midwinter Warming in the Southern Hemisphere Stratosphere 2002: A Comparison to Northern Hemisphere Phenomena. *Journal of the Atmospheric Sciences*, 62(3), 603–613. https://doi.org/10.1175/jas-3316.1

Labitzke, K., Austin, J., Butchart, N., Knight, J., Takahashi, M., Nakamoto, M., et al. (2002). The global signal of the 11-year solar cycle in the stratosphere:

observations and models. *Journal of Atmospheric and Solar-Terrestrial Physics*, 64(2), 203–210. https://doi.org/10.1016/s1364-6826(01)00084-0

Leblanc, T., McDermid, I. S., Keckhut, P., Hauchecorne, A., She, C. Y., & Krueger, D. A. (1998). Temperature climatology of the middle atmosphere from long-term lidar measurements at middle and low latitudes. *Journal of Geophysical Research: Atmospheres*, 103(D14), 17191–17204. https://doi.org/10.1029/98jd01347

Le Pichon, A., Assink, J. D., Heinrich, P., Blanc, E., Charlton-Perez, A., Lee, C. F., et al. (2015). Comparison of co-located independent ground-based middle atmospheric wind and temperature measurements with numerical weather prediction models. *Journal of Geophysical Research: Atmospheres*, 120(16), 8318–8331. https://doi.org/10.1002/2015jd023273

Liu, H.-L., Talaat, E. R., Roble, R. G., Lieberman, R. S., Riggin, D. M., & Yee, J.-H. (2004). The 6.5-day wave and its seasonal variability in the middle and upper atmosphere. *Journal of Geophysical Research: Atmospheres*, 109(D21), n/a–n/a. https://doi.org/10.1029/2004jd004795

McDermid, I. S., Leblanc, T., Keckhut, P., Hauchecorne, A., She, C. Y., & Krueger, D. A. (1998). Climatology of the middle-atmosphere temperature from long-term lidar measurements at mid and low latitudes. *Optical Remote Sensing for Industry and Environmental Monitoring*. https://doi.org/10.1117/12.319557

Measures, R. M. (1992). *Laser remote sensing: Fundamentals and applications*. Malabar, FL: Krieger Publishing.

Meyer, J. (1987). Two-Moment Decision Models and Expected Utility. *American Economic Review*. 77. 421-30.

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geoscientific Model Development*, 8(5), 1339–1356. https://doi.org/10.5194/gmd-8-1339-2015

Moser, D. K., Wickwar, V. B., Navarro, L. A., Barton, D. L., & Herron, J. P. (2015). *Comparing Rayleigh Lidar and Assimilative Model Temperatures at 45 km*. Poster presented at 30th CEDAR Workshop, University of Washington, Seattle, WA.

Moser, D. K., Wickwar, V. B., Navarro, L. A., & Herron, J. P. (2015). *USU Rayleigh Lidar and Assimilative Models to Examine Coupling between the Mesosphere and Lower Atmosphere*. Poster presented at the 2015 NCAR ASP Summer Colloquium, NCAR & University of Colorado, Boulder, CO.

Ortland, D. A. (2017). Daily estimates of the migrating tide and zonal mean temperature in the mesosphere and lower thermosphere derived from SABER data. *Journal of Geophysical Research: Atmospheres*, 122(7), 3754–3785. https://doi.org/10.1002/2016jd025573

Pawson, S., Kodera, K., Hamilton, K., Shepherd, T. G., Beagley, S. R., Boville, B. A., et al. (2000). The GCM-Reality Intercomparison Project for SPARC (GRIPS): Scientific Issues and Initial Results. *Bulletin of the American Meteorological Society*, 81, 781–796. https://doi.org/10.1175/1520-0477(2000)081<0781:TGIPFS>2.3.CO;2

Picone, J. M., Hedin, A. E., Drob, D. P., & Aikin, A. C. (2002). NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal of Geophysical Research: Space Physics*, 107(A12), SIA 15–1–SIA 15–16. https://doi.org/10.1029/2002ja009430

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., et al. (2016). ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, 29(11), 4083–4097. https://doi.org/10.1175/jcli-d-15-0556.1

Polvani, L. M., & Waugh, D.W. (2004). Upward Wave Activity Flux as a Precursor to Extreme Stratospheric Events and Subsequent Anomalous Surface Weather Regimes. *Journal of Climate*, 17, 3548–3554. https://doi.org/10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2

Rabier, F., & Liu, Z. (2003). Variational Data Assimilation: Theory and Overview. *Seminar on Recent developments in data assimilation for atmosphere and ocean.* Paper presented at ECMWF, Shinfield Park, Reading RG29AX, United Kingdom.

Randel, W., Udelhofen, P., Fleming, E., Geller, M., Gelman, M., Hamilton, K., et al. (2004). The SPARC Intercomparison of Middle-Atmosphere Climatologies.

Journal of Climate, 17, 986–1003. https://doi.org/10.1175/1520-0442(2004)017<0986:tsiomc>2.0.co;2

Remsberg, E. (2008). On the observed changes in upper stratospheric and mesospheric temperatures from UARS HALOE. *Annales Geophysicae*, 26(5), 1287–1297. https://doi.org/10.5194/angeo-26-1287-2008

Rezac, L., Jian, Y., Yue, J., Russell, J. M., Kutepov, A., Garcia, R., et al. (2015). Validation of the global distribution of CO2volume mixing ratio in the mesosphere and lower thermosphere from SABER. *Journal of Geophysical Research: Atmospheres*, 120(23), 12,067–12,081. https://doi.org/10.1002/2015jd023955

Sakazaki, T., Fujiwara, M., & Shiotani, M. (2018). Representation of solar tides in the stratosphere and lower mesosphere in state-of-the-art reanalyses and in satellite observations. *Atmospheric Chemistry and Physics*, 18(2), 1437–1456. https://doi.org/10.5194/acp-18-1437-2018

Santer, B. D. (2003). Behavior of tropopause height and atmospheric temperature in models, reanalyses, and observations: Decadal changes. *Journal of Geophysical Research*, 108(D1). https://doi.org/10.1029/2002jd002258

Schöch, A., Baumgarten, G., & Fiedler, J. (2008). Polar middle atmosphere temperature climatology from Rayleigh lidar measurements at ALOMAR (69° N). *Annales Geophysicae*, 26(7), 1681–1698. https://doi.org/10.5194/angeo-26-1681-2008

She, C. Y., Chen, S., Hu, Z., Sherman, J., Vance, J. D., Vasoli, V., et al. (2000). Eight-year climatology of nocturnal temperature and sodium density in the mesopause region (80 to 105 km) over Fort Collins, Co (41°N, 105°W). *Geophysical Research Letters*, 27(20), 3289–3292. https://doi.org/10.1029/2000gl003825

Sigmond, M., Scinocca, J. F., Kharin, V. V., & Shepherd, T. G. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, 6(2), 98–102. https://doi.org/10.1038/ngeo1698

Siskind, D. E., Sassi, F., Randall, C. E., Harvey, V. L., Hervig, M. E., & Bailey, S. M. (2015). Is a high-altitude meteorological analysis necessary to simulate thermosphere-stratosphere coupling? *Geophysical Research Letters*, 42(19), 8225–8230. https://doi.org/10.1002/2015gl065838

Sivakumar, V., Bencherif, H., Hauchecorne, A., Keckhut, P., Rao, D. N., Sharma, S., et al. (2006). Rayleigh lidar observations of double stratopause structure over three different northern hemisphere stations. *Atmospheric Chemistry and Physics Discussions*, 6(4), 6933–6956. https://doi.org/10.5194/acpd-6-6933-2006

Sox, L. (2016). *Rayleigh-Scatter Lidar Measurements of the Mesosphere and Thermosphere and their Connections to Sudden Stratospheric Warmings* (Doctoral dissertation). Retrieved from Digital Commons (https://digitalcommons.usu.edu/etd/5227). Logan, UT: Utah State University.

Sox, L., Wickwar, V. B., Fish, C. S., & Herron, J. P. (2016). Connection between the midlatitude mesosphere and sudden stratospheric warmings as measured by

Rayleigh-scatter lidar. *Journal of Geophysical Research: Atmospheres*, 121(9), 4627–4636. https://doi.org/10.1002/2015jd024374

Sox, L., Wickwar, V. B., & Herron, J. P. (2013). *Middle Atmosphere Temperature Results from a New, High-powered, Large-Aperture Rayleigh Lidar*. Paper presented at the 2013 Utah Space Grant Consortium Meeting, Utah Space Grant Consortium, Salt Lake City, UT.

Stickler, A., Brönnimann, S., Valente, M. A., Bethke, J., Sterin, A., Jourdain, S., et al. (2014). ERA-CLIM: Historical Surface and Upper-Air Data for Future Reanalyses. *Bulletin of the American Meteorological Society*, 95(9), 1419–1430. https://doi.org/10.1175/bams-d-13-00147.1

Walterscheid, R. L., & Christensen, A. B. (2016). Low-latitude gravity wave variances in the mesosphere and lower thermosphere derived from SABER temperature observation and compared with model simulation of waves generated by deep tropical convection. *Journal of Geophysical Research: Atmospheres*, 121(20), 11,900–11,912. https://doi.org/10.1002/2016jd024843

Wang, X. L., Feng, Y., Chan, R., & Isaac, V. (2016). Inter-comparison of extra-tropical cyclone activity in nine reanalysis datasets. *Atmospheric Research*, 181, 133-153. https://doi.org/10.1016/j.atmosres.2016.06.010

Wing, R., Hauchecorne, A., Keckhut, P., Godin-Beekmann, S., Khaykin, S., & McCullough, E. M. (2018). Lidar temperature series in the middle atmosphere as a reference data set. Part B: Assessment of temperature observations from

MLS/Aura and SABER/TIMED satellites. *Atmospheric Measurement Techniques Discussions*, 1–23. https://doi.org/10.5194/amt-2018-139

Wynn, T. A. (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* (Doctoral dissertation). Retrieved from Digital Commons (https://digitalcommons.usu.edu/etd/797). Logan, UT: Utah State University.

Xu, J., She, C. Y., Yuan, W., Mertens, C., Mlynczak, M., & Russell, J. (2006). Comparison between the temperature measurements by TIMED/SABER and lidar in the midlatitude. *Journal of Geophysical Research*, 111(A10). https://doi.org/10.1029/2005ja011439

Yuan, T., Thurairajah, B., She, C.-Y., Chandran, A., Collins, R. L., & Krueger, D. A. (2012). Wind and temperature response of midlatitude mesopause region to the 2009 Sudden Stratospheric Warming. *Journal of Geophysical Research: Atmospheres*, 117(D9), n/a–n/a. https://doi.org/10.1029/2011jd017142

Zhang, Y., Sheng, Z., Shi, H., Zhou, S., Shi, W., Du, H., & Fan, Z. (2017). Properties of the Long-Term Oscillations in the Middle Atmosphere Based on Observations from TIMED/SABER Instrument and FPI over Kelan. *Atmosphere*, 8(12), 7. https://doi.org/10.3390/atmos8010007

APPENDICES

APPENDIX A. ENLARGED ADAPTIVE FIT IMAGES

Figure 31 – Adaptive fit to ALO-USU Rayleigh-scatter temperature measurements (enlarged).

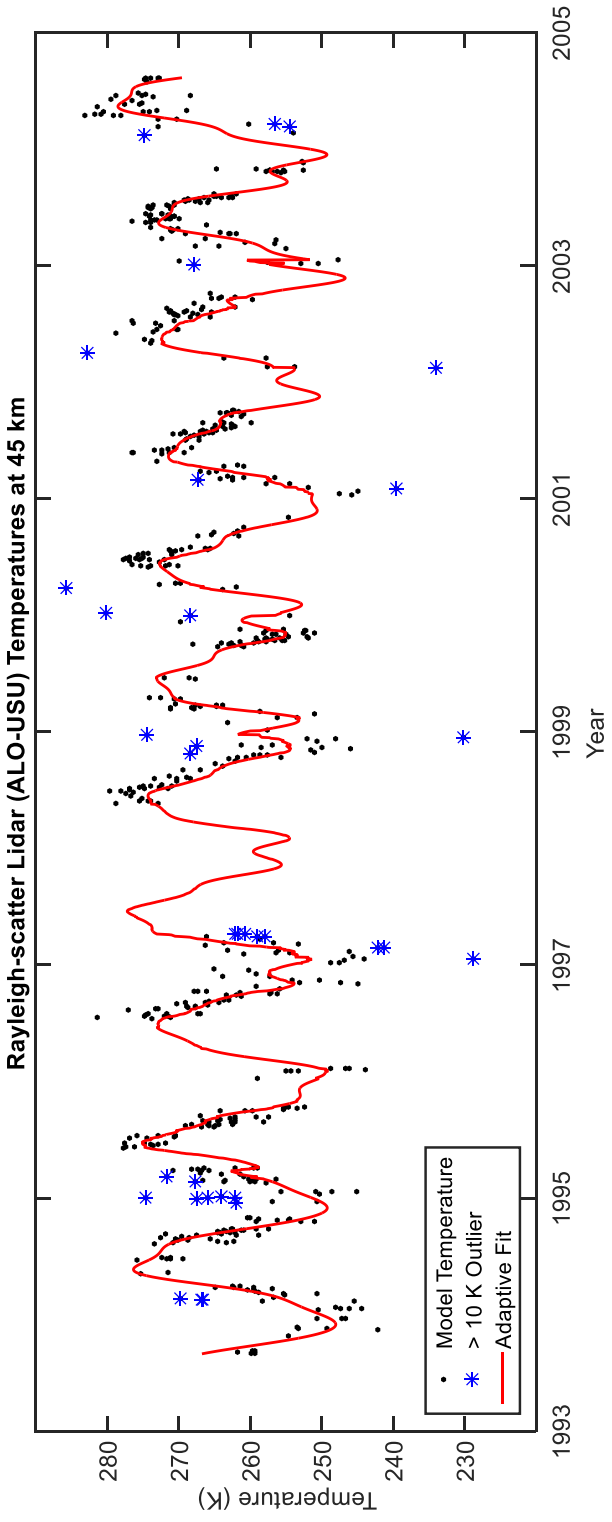


Figure 32 – Adaptive fit to ERA-20C model temperature output (enlarged).

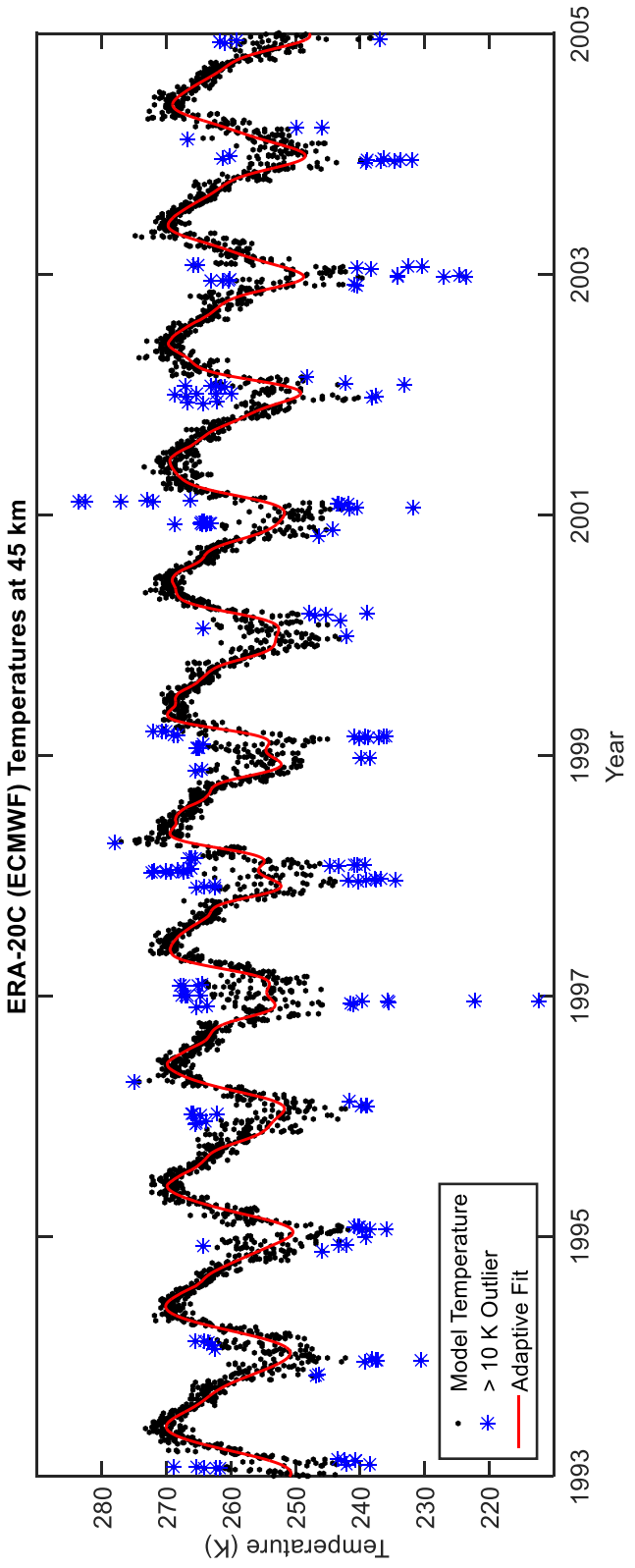


Figure 33 – Adaptive fit to MERRA-2 model temperature output (enlarged).

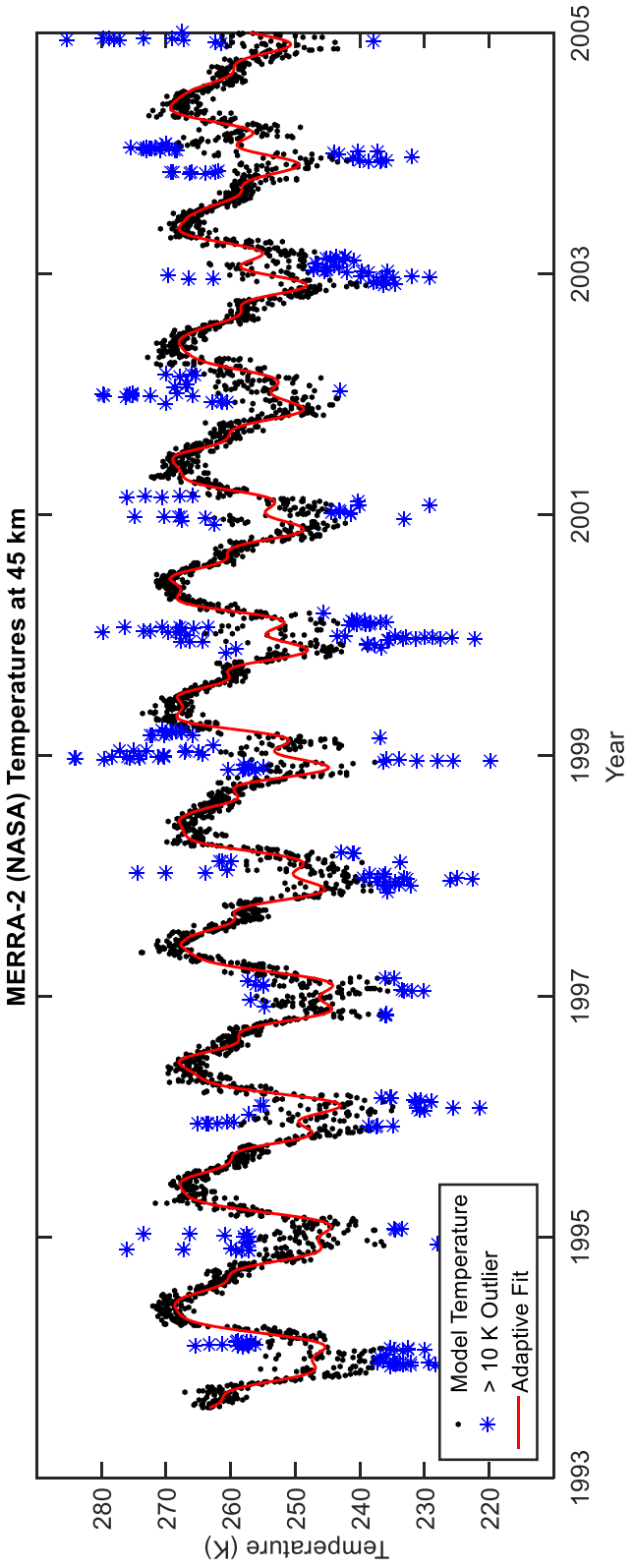
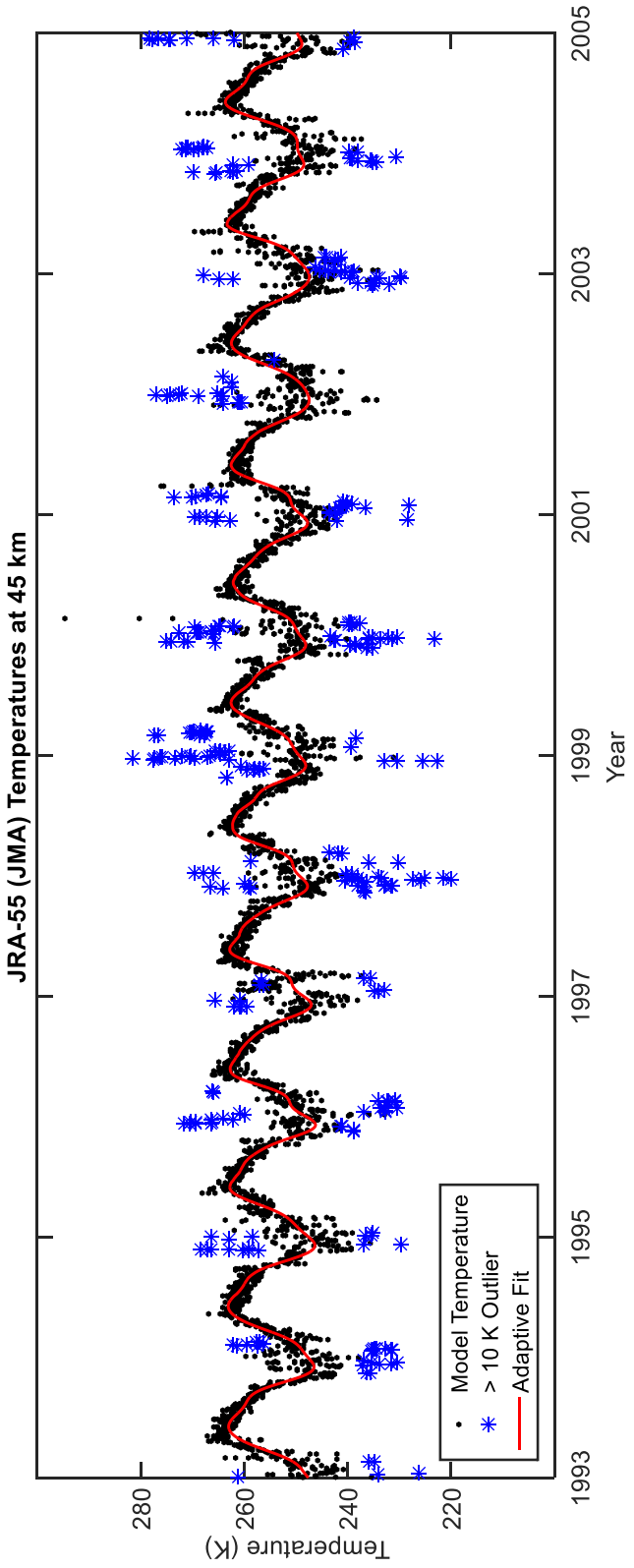


Figure 34 – Adaptive fit to JRA-55 model temperature output (enlarged).



APPENDIX B. OUTLIER TABLES

Table 5

Lidar Dataset Monthly Outliers per Number of Fit Parameters

Year	January			February			March			April			May			June			July			August			September			October			November			December		
	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14
1993	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1994	0	0	0	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2
1995	4	4	3	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	1	1	1	2	2	2	2	1	0	3	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1998	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2
1999	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
2000	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2002	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2003	1	2	2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2004	0	0	0	1	2	2	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 6

ERA-20C Dataset Monthly Outliers per Number of Fit Parameters

Year	January			February			March			April			May			June			July			August			September			October			November			December		
	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14
1993	5	5	5	7	7	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	5	5	5
1994	1	1	1	4	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	4	4	4
1995	6	6	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	4	3
1996	7	7	6	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	10	10	10
1997	4	4	3	2	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	2	2	7	7	7
1998	14	12	12	4	4	4	0	0	0	1	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	2	2	2
1999	3	0	0	8	8	8	5	5	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
2000	1	1	1	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3	3	3	7	5	5
2001	7	8	8	7	7	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	8	8
2002	8	7	9	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	9	9	10
2003	6	6	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	8	8
2004	0	0	0	1	1	1	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	4	4	4

Table 7

MERRA-2 Dataset Monthly Outliers per Number of Fit Parameters

Year	January			February			March			April			May			June			July			August			September			October			November			December		
	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14
1993	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	13	13
1994	8	8	7	13	12	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	8	8	3	4	4
1995	9	8	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	9	8	8
1996	6	6	6	10	10	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	1	2	2
1997	5	5	5	5	5	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	3	14	14	14
1998	6	6	6	4	4	4	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	5	5	21	20	20
1999	7	8	8	2	2	2	10	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	5	19	19	20
2000	15	14	13	8	6	5	1	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	7	7	7
2001	7	9	7	6	6	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	14	14	14
2002	5	5	4	3	3	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	3	12	12	12
2003	12	13	13	8	8	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	5	3	3	8	7	6
2004	18	17	18	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	12	13	11

Table 5

JRA-55 Dataset Monthly Outliers per Number of Fit Parameters

Year	January			February			March			April			May			June			July			August			September			October			November			December		
	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14	10	12	14
1993	1	1	1	0	0	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1994	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
1995	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2
1997	2	2	1	2	0	0	3	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	0	1	1
1998	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	3
1999	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2000	9	9	8	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	0	0	0	0	0	0	5	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	3	6	7	5
2002	1	1	1	0	0	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
2003	3	3	2	0	0	0	3	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2004	2	2	2	3	3	2	0	0	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	3	9	9	9

APPENDIX C. COMPLETE RESULT TABLES

Table 9

Number of Temperatures per Month per Dataset (REAL Subset)

Year	January				February				March				April				May				June			
	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55
1993	0	31	0	31	0	28	0	28	0	31	0	31	0	30	0	30	0	31	0	31	0	30	0	30
1994	5	31	31	31	6	28	28	28	13	31	31	31	0	30	30	30	2	31	31	31	6	30	30	30
1995	9	31	31	31	9	28	28	28	10	31	31	31	6	30	30	30	0	31	31	31	9	30	30	30
1996	1	31	31	31	7	29	29	29	0	31	31	31	0	30	30	30	0	31	31	31	0	30	30	30
1997	6	31	31	31	8	28	28	28	9	31	31	31	3	30	30	30	0	31	31	31	0	30	30	30
1998	0	31	31	31	0	28	28	28	0	31	31	31	0	30	30	30	5	31	31	31	6	30	30	30
1999	2	31	31	31	3	28	28	28	11	31	31	31	4	30	30	30	0	31	31	31	3	30	30	30
2000	1	31	31	31	0	29	29	29	3	31	31	31	3	30	30	30	0	31	31	31	18	30	30	30
2001	4	31	31	31	6	28	28	28	8	31	31	31	4	30	30	30	7	31	31	31	3	30	30	30
2002	0	31	31	31	3	28	28	28	2	31	31	31	1	30	30	30	4	31	31	31	5	30	30	30
2003	6	31	31	31	2	28	28	28	6	31	31	31	10	30	30	30	12	31	31	31	12	30	30	30
2004	0	31	31	31	2	29	29	29	4	31	31	31	12	30	30	30	4	31	31	31	8	30	30	30
Year	July				August				September				October				November				December			
1993	0	31	0	31	1	31	31	31	4	30	30	30	1	31	31	31	4	30	30	30	3	31	31	31
1994	0	31	31	31	11	31	31	31	19	30	30	30	9	31	31	31	0	30	30	30	4	31	31	31
1995	7	31	31	31	15	31	31	31	7	30	30	30	9	31	31	31	0	30	30	30	0	31	31	31
1996	9	31	31	31	11	31	31	31	9	30	30	30	9	31	31	31	7	30	30	30	3	31	31	31
1997	0	31	31	31	0	31	31	31	0	30	30	30	0	31	31	31	0	30	30	30	0	31	31	31
1998	11	31	31	31	5	31	31	31	4	30	30	30	10	31	31	31	7	30	30	30	5	31	31	31
1999	0	31	31	31	0	31	31	31	6	30	30	30	21	31	31	31	10	30	30	30	3	31	31	31
2000	13	31	31	31	0	31	31	31	6	30	30	30	1	31	31	31	1	30	30	30	0	31	31	31
2001	15	31	31	31	14	31	31	31	7	30	30	30	2	31	31	31	0	30	30	30	0	31	31	31
2002	10	31	31	31	12	31	31	31	6	30	30	30	1	31	31	31	0	30	30	30	0	31	31	31
2003	19	31	31	31	7	31	31	31	0	30	30	30	11	31	31	31	2	30	30	30	0	31	31	31
2004	1	31	31	31	5	31	31	31	0	30	30	30	0	31	31	31	0	30	30	30	0	31	31	31

Table 10

Number of Temperatures per Month (All Datasets, RLID & FITL Subsets)

Year	January	February	March	April	May	June	July	August	September	October	November	December
1993	0	0	0	0	0	0	0	1	4	1	4	3
1994	5	6	13	0	2	6	0	11	19	9	0	4
1995	9	9	10	6	0	9	7	15	7	9	0	0
1996	1	7	0	0	0	0	9	11	9	9	7	3
1997	6	8	9	3	0	0	0	0	0	0	0	0
1998	0	0	0	0	5	6	11	5	4	10	7	5
1999	2	3	11	4	0	3	0	0	6	21	10	3
2000	1	0	3	3	0	18	13	0	6	1	1	0
2001	4	6	8	4	7	3	15	14	7	2	0	0
2002	0	3	2	1	4	5	10	12	6	1	0	0
2003	6	2	6	10	12	12	19	7	0	11	2	0
2004	0	2	4	12	4	8	1	5	0	0	0	0

Table 11

Number of Temperatures per Month per Dataset (FIT Subset)

Year	January				February				March				April				May				June			
	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55	ALO-USU	ERA-20C	MER-RA-2	JRA-55
1993	0	31	0	31	0	28	0	28	0	31	0	31	0	30	0	30	0	31	0	31	0	30	0	30
1994	31	31	31	31	28	28	28	28	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
1995	31	31	31	31	28	28	28	28	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
1996	31	31	31	31	29	29	29	29	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
1997	31	31	31	31	28	28	28	28	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
1998	31	31	31	31	28	28	28	28	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
1999	31	31	31	31	28	28	28	28	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
2000	31	31	31	31	29	29	29	29	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
2001	31	31	31	31	28	28	28	28	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
2002	31	31	31	31	28	28	28	28	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
2003	31	31	31	31	28	28	28	28	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
2004	31	31	31	31	29	29	29	29	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30
Year	July				August				September				October				November				December			
1993	0	31	0	31	1	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
1994	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
1995	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
1996	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
1997	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
1998	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
1999	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
2000	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
2001	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
2002	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
2003	31	31	31	31	31	31	31	31	30	30	30	30	31	31	31	31	30	30	30	30	31	31	31	31
2004	31	31	31	31	10	31	31	31	0	30	30	30	0	31	31	31	0	30	30	30	0	31	31	31

Table 12

Mean Monthly Temperatures (K) per Dataset (REAL Subset)

Year	January				February				March				April				May				June			
	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55
1993		254.57		246.05		248.02		249.26		259.92		255.08		266.3		261.29		269.61		264.16		270.19		263.55
1994	247.28	253.42	241.16	250.06	260.61	257.15	254.89	249.77	259.01	256.67	251.75	253.21		265.45	266.34	258.51	273.44	268.89	269.3	262.61	271.96	269.13	269.05	261.38
1995	257.5	245.03	249.16	249.22	262.09	255.3	246.08	250.53	264.4	259.85	257.07	255.84	263.79	265.72	263.53	257.11		269.05	265.62	261.65	275.12	269.53	266.51	262.72
1996	258.98	250.43	240.84	245.9	249.69	253.53	242.31	251.42		258.42	254.91	252.05		267.06	262.75	257.99		268.66	266.32	262.68		268.71	266.15	262.13
1997	243.59	257.33	242.39	252.45	252.58	254.8	246.7	253.02	260.07	257.22	255.33	251.75	261.49	263.92	261.45	261.82		268.82	268.75	262.35		269.75	267.74	261.56
1998		256.81	249.26	249.47		255.67	250.23	249.65		261.12	253.16	256.1		271.94	265.75	256.94	275.05	269.33	265.58	262.28	277.17	268.52	266.46	261.88
1999	260.41	257.18	257.22	250.69	253.54	248.69	249.66	249.53	266.85	264.01	265.58	253.68	271.75	266.83	267.96	257		268.18	268.58	262.05	269.42	268.68	268.35	262.25
2000	280.17	254.56	259.53	253.19		254.18	248.32	254.07	270.58	259.01	259.26	252.84	271.03	266.42	266.46	258.17		269.5	269.14	261.81	274.59	269.62	269.74	261.7
2001	244.47	247.22	246.6	247.47	260.18	261.07	256.34	251.8	263.1	265.44	261.22	256.36	264.89	267.47	266.96	260.29	273.11	268.36	267.48	261.32	270.05	268.58	267.99	261.11
2002		254.42	256.94	250.54	248.47	256.52	256.49	250.31	260.69	265.14	261.01	249.03	282.78	267.55	265.31	255.61	275.33	269.12	268.74	263.29	271.92	269.41	267.25	262.82
2003	257.75	248.14	250.79	248.41	259.71	257.54	248.73	248.16	263.34	261.12	255.17	255.19	267.53	264.64	262.11	258.09	272.51	269.07	266.62	262.61	272.56	269.38	266.4	261.73
2004		251.32	265.38	246.65	264.38	254.61	259.44	250.16	261.01	258.23	256.78	252.09	276.47	264.86	264.77	256.38	277.41	269.09	267.26	262.6	275.27	268.96	266.94	262.41
Year	July				August				September				October				November				December			
1993		267.84		261.93	259.79	264.61	262.39	259.81	260.05	263.58	261.21	257.23	254.7	260.63	255.34	252.14	249.5	254.6	245.42	247.35	248.16	248.86	238.27	246.45
1994		267.2	265.88	260.89	268.21	264.28	261.69	259.91	263.32	262.07	260	258.74	256.42	257.87	254.49	252.44		255.11	254.32	247.39	259.28	252.65	246.98	245.63
1995	273.06	267.4	264.1	261.41	264.82	264.13	258.9	259.66	262.12	262.24	256.81	257.38	256.78	260.14	251.99	254.23		256.61	243.51	247.23		256.12	251.31	249.23
1996	273.41	266	262.72	260.91	270.15	263.93	260.11	259.51	264.32	263.04	257.65	256.63	260.93	257.21	251.24	251.51	252.97	255.26	244.32	249.4	261.51	250.18	249.62	247.01
1997		267.66	264.25	260.22		264.48	258.77	258.35		262.85	257.36	257.19		260.01	254.68	253.33		256.18	243.91	248.6		248.68	238.4	247.29
1998	273.69	266.67	264.04	260.82	270.45	265.1	261.73	258.64	266.31	263.59	259.43	255.48	259.84	258.27	254.25	250.27	255.93	256.43	251.27	248.94	251.08	253.1	255.5	247.62
1999		267.24	265.4	260.9		264.22	260.57	258.54	262.18	263.2	259.12	255.65	257.96	259.88	254.99	255.7	255.42	254.21	247.67	247.79	264.2	250.17	244.35	248.03
2000	270.44	267.69	266.35	260.12		263.87	261.2	258.26	263.91	262.7	259.25	256.17	260.87	256.42	252.66	252.72	254.58	254.22	249.54	247.76		253.77	255.6	248.52
2001	267.86	267.5	265.37	260.57	263.79	264.71	261.44	258.53	262.41	262.95	259.7	255.91	262.36	259	255.08	252.21		254.76	250.06	250.59		251.94	260.03	245.73
2002	270.93	267.44	264.42	261.55	268.57	264.09	259.08	259.19	263.9	261.84	257.48	256.61	265.51	260.6	253.34	251.67		254.23	246.12	247.59		248.59	248.23	246.16
2003	269.43	267.42	263.45	260.45	263.59	264.56	259.49	258.51		262.56	258.31	257.74	257.09	258.37	256.33	254.91	252.66	253.11	255.51	248.86		247.61	247.57	249.62
2004	274.54	266.04	263.01	260.75	273.79	263.89	260.32	258.62		261.92	259.8	256.19		257.13	254.79	251.93		253.39	253.09	248.85		251.5	260.99	253.95

Table 13

Mean Monthly Temperatures (K) per Dataset (RLID Subset)

Year	January				February				March				April				May				June			
	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55
1993																								
1994	247.28	251.59	240.2	251	260.61	260.49	253.86	250.1	259.01	257.75	252.47	253.58					273.44	267.28	268.74	264.47	271.96	268.59	267.48	261.32
1995	257.5	246.22	245.87	251.55	262.09	256.86	247.88	252.76	264.4	260.7	256.47	255.9	263.79	262.71	259.01	257.68	NaN				275.12	269.66	266.85	262.56
1996	258.98	252.5	242.09	248.95	249.69	253.3	247.04	251.63																
1997	243.59	255.93	239.97	253.01	252.58	252.74	244.24	254.49	260.07	258.27	254.11	252.54	261.49	259.63	255.22	260.81								
1998																	275.05	270.25	266.42	263.49	277.17	268.45	265.52	262.01
1999	260.41	249.6	260.57	251.69	253.54	246.7	250.05	250.76	266.85	263.96	264.68	254.14	271.75	267.88	270	257.82					269.42	268.93	268.92	261.9
2000	280.17	251.34	267.82	259.72					270.58	265.44	262.6	253.86	271.03	265.03	265.49	253.9					274.59	269.7	269.76	261.72
2001	244.47	244.96	247.96	247.14	260.18	259.84	257.12	252.65	263.1	264.85	261.1	257.26	264.89	267.31	265.58	260.8	273.11	269.06	268.03	261.29	270.05	266.81	265.99	261.98
2002					248.47	255.56	248.69	251.53	260.69	265.29	255.68	247.89	282.78	265.37	266.7	257.09	275.33	268.59	269	262.1	271.92	269.39	266.85	263.51
2003	257.75	244.09	245.71	244.4	259.71	258.42	249.86	254.23	263.34	261.98	256.48	255.21	267.53	265.49	262.59	257.97	272.51	269.64	266.78	262.35	272.56	268.96	265.92	261.9
2004					264.38	261.02	257.57	241.87	261.01	259.17	255.85	251.31	276.47	265.29	264.85	257.52	277.41	269.17	266.88	262.52	275.27	268.52	266.11	262.36
Year	July				August				September				October				November				December			
1993					259.79	262.42	263.77	258.34	260.05	264.31	259.13	258.44	254.7	257.02	253.06	248.82	249.5	255.74	246.32	246.27	248.16	238.46	238.94	250.75
1994					268.21	263.85	261.04	259.78	263.32	262.32	259.89	258.96	256.42	254.89	251.57	251.44					259.28	248.17	245.77	244.08
1995	273.06	267.91	264.08	261.76	264.82	263.43	258.69	259.15	262.12	261.43	256.77	257.22	256.78	261.11	253.68	254.72								
1996	273.41	265.48	262.78	261.13	270.15	264.08	261.08	259.5	264.32	262.82	257.55	257.65	260.93	259.43	254.41	253.14	252.97	254.45	242.71	250.98	261.51	240.36	254.97	243.7
1997																								
1998	273.69	266.73	263.85	261.08	270.45	265.47	261.92	259.13	266.31	264.72	259.64	255.86	259.84	258.35	253.83	250.17	255.93	259.74	252.01	248.29	251.08	254.01	243.54	247.23
1999									262.18	260.91	258.54	253.63	257.96	260.03	255.54	256.77	255.42	254.37	251.07	248.17	264.2	250.55	261.18	249.75
2000	270.44	267.54	265.59	260.17					263.91	263.47	259.55	255.47	260.87	262.17	258.32	254.91	254.58	255.71	252.65	248.01				
2001	267.86	266.93	265.35	260.37	263.79	264.81	261.56	258.5	262.41	263.26	259.67	254.99	262.36	261.5	262.49	258.93								
2002	270.93	266.9	263.68	261.82	268.57	264.7	259.57	259.09	263.9	262.46	257.89	256.37	265.51	258.36	258.25	254.39								
2003	269.43	267.35	263.46	260.45	263.59	265.66	260.36	259.85					257.09	257.19	257.01	254.35	252.66	250.86	248.08	245.1				
2004	274.54	265.56	262.56	261.05	273.79	263.74	260.47	259.15																

Table 14

Mean Monthly Temperatures (K) per Dataset (FIT Subset)

Year	January				February				March				April				May				June			
	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55
1993		250.92		248.35		253.43		250.23		259.08		254.17		265.36		259.17		269.26		262.29		269.54		262.76
1994	251.83	251	246.23	248.43	255	253.52	246.8	250.31	260.35	259.16	255.55	254.25	269.73	265.44	264.84	259.25	275.77	269.35	267.98	262.37	274.69	269.62	268.3	262.84
1995	253.02	250.63	245.32	248.65	256.2	253.54	245.89	250.63	260.77	259.49	254.65	253.43	260.11	265.21	263.92	258.11	267.3	268.98	267.12	261.99	274.58	269.21	267.49	262.37
1996	250.25	251.94	245.28	249.57	251.21	253.49	244.84	251.26	259.74	259.21	255	253.54	266.57	265.1	263.23	258.96	269.74	268.78	266.12	262.41	272.58	269.4	267.6	262.08
1997	252.67	254.31	245.18	249.78	255.16	254.54	245.69	250.97	264.89	259.6	254.47	253.44	273.1	266.75	263.23	259.02	274.78	269.26	266.57	262.5	276.74	268.71	267.18	261.83
1998	256.46	255.53	249.78	250.1	256.3	255.1	249.6	250.94	265.37	259.69	257.16	253.69	271.25	267.3	265.3	258.89	273.2	269.16	267.32	261.89	274.05	268.53	267.75	262.03
1999	256.13	254.42	252.75	250.15	254.23	254.56	251.42	251.35	263.48	260.36	258.14	253.79	270.3	268.26	267.08	257.9	271.69	269.33	267.76	261.46	272.86	268.56	267.95	262.32
2000	254.53	252.7	253.42	249.9	254.4	255.1	252.8	251.2	263.21	262.01	261.21	254.54	269.23	267.34	267.81	258.93	271.61	268.56	267.89	261.39	272.19	268.96	269	261.91
2001	251.9	252.23	254.09	250.16	255.87	256.52	254.01	251.15	260.19	263.29	260.62	253.33	268.35	267.36	266.61	258.26	271.22	268.89	268	261.95	270.03	269.26	268.76	261.99
2002	255.54	250.48	253.34	248.35	255.54	257.48	253.44	250.21	261.7	264.01	259.25	253.69	270.01	266.55	265.09	258.51	272.26	268.87	267.36	261.83	270.83	269.25	267.53	262.14
2003	256.14	251.3	257.76	248.2	257.19	256.38	256.2	249.84	260.84	260.93	256.91	252.14	268.64	265.84	264.23	257.08	272.51	269.26	267.75	262.24	271.19	269.1	266.72	262.71
2004	256.64	250.61	258.24	249.65	262.92	255.68	257.45	250.03	266.72	260.41	258.42	252.56	274.58	265.26	265.53	257.84	278.19	268.57	269.06	262.53	276.84	268.36	267.85	263.12
Year	July				August				September				October				November				December			
1993		267.18		261.2	266.75	264.56	261.93	259.64	262.61	262.39	260.36	257.73	254.15	259.42	254.65	252.73	248.93	255.57	247.99	247.61	248.74	252.17	247.19	246.89
1994	272.63	267.27	264.99	261.27	270.4	264.63	261.02	259.77	263.81	262.55	259.46	257.7	255.34	259.36	253.75	252.87	250.12	255.83	247.09	247.75	249.94	252.3	246.28	246.67
1995	272.14	266.66	263.86	260.99	267.61	264.34	260.28	259.71	263.7	262.44	258.63	257.32	256.16	258.55	251.96	253.23	253.14	255.03	247.75	248.06	252.66	253.37	249.1	246.71
1996	271.72	266.98	263.32	260.83	267.63	264.15	259.06	259.2	263.95	262.68	257.9	256.85	256.77	259.32	251.51	252.98	255.66	254.35	245	248.44	256.45	253.62	245.72	247.5
1997	273.16	266.98	263.73	260.65	268.58	264.53	259.78	258.97	265.94	262.89	258.88	256.08	259.41	259.2	252.62	252.04	256.14	253.4	246.08	248.4	258.98	253.49	248.42	248.38
1998	270.43	267.5	264.9	260.73	265.98	264.62	259.77	258.34	263.48	263.06	259.27	256.41	257.96	259.92	254.62	252.5	254.64	253.87	246.07	248.78	258.29	252.92	248.66	248.61
1999	269.45	267.28	265.66	260.51	265.52	264.81	260.83	258.34	263.49	262.85	259.94	256.18	257.17	258.74	253.91	252.38	256.79	254.3	248.64	248.8	260.35	253.06	252.6	248.56
2000	267.45	267.32	265.63	260.4	264.06	264.61	260.98	258.32	261.75	262.8	259.78	256.26	255.64	258.53	253.99	252.79	251.05	254.2	249.2	248.84	251.18	252.19	252.72	248.2
2001	267.67	267.26	265.07	260.44	264.4	264.74	260.77	258.94	263.21	262.21	259.16	256.12	256.12	258.71	253.37	251.87	250.78	255.15	249.12	248.45	254.24	250.74	252.16	247.52
2002	267.89	267.32	264.43	260.56	263.61	264.56	259.89	258.98	262.01	262.25	258.36	256.62	255.01	259.6	255.65	252.53	247.67	254.03	249.41	249.05	251.71	249.27	251.6	247.53
2003	269.73	267.03	263.79	260.56	262.13	264.34	259.71	258.94	255.38	261.85	258.2	256.95	256.69	259	255.68	252.47	253.77	253.68	250.45	248.93	249.97	249.06	252.44	248.91
2004	275.29	266.23	264.88	260.87	271.03	263.54	260.84	259.28		261.05	259.4	257.22		258.16	256.74	252.69		252.76	251.58	249.26		248.28	253.87	249.32

Table 15

Mean Monthly Temperatures (K) per Dataset (FITL Subset)

Year	January				February				March				April				May				June			
	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55	ALO-USU	ERA-20C	MERR A-2	JRA-55
1993																								
1994	252.17	250.91	245.98	248.58	255.19	253.75	246.83	250.43	261.26	259.89	256.79	254.82					275.53	268.96	267.73	262.04	273.83	269.16	267.93	262.58
1995	252.58	250.61	245.56	248.36	257.2	255.09	247.37	251.21	260.48	259.48	254.65	253.47	259.55	263.31	261.45	256.23					274.81	269.2	267.58	262.37
1996	250.87	251.98	246.14	249.15	249.65	252.43	243.29	251.03																
1997	252.51	254.35	245.23	249.76	254.86	254.42	245.44	250.93	265.69	260.18	255.21	253.83	272.21	265.05	261.47	257.22								
1998																	273.89	268.9	267.53	262.18	273.89	268.5	267.62	261.91
1999	256.08	254.28	252.37	250.13	254.12	254.59	251.32	251.44	264.55	261.2	259.2	254.09	270.58	268.64	267.56	257.94					273.1	268.6	268.06	262.4
2000	255.77	252.68	254.3	249.63					266.19	264.47	264.68	256.05	268.43	266.69	267.47	257.99					271.98	268.91	268.88	261.86
2001	252.24	252.39	253.94	250.37	256.65	257.76	254.74	251.32	259.42	262.79	259.93	253.03	268.43	267.4	266.73	258.25	270.97	269.2	268.34	262.43	270.19	269.43	268.89	262.37
2002					255.82	257.47	253.13	250.14	261.48	264.19	259.15	253.55	266.61	265.7	263.07	256.2	272.1	268.67	267.25	261.64	270.81	269.35	267.68	262.24
2003	256.57	250.28	257.46	247.88	258.13	257.93	255.24	250.39	260.83	260.93	256.91	252.13	268.81	265.91	264.39	257.12	272.45	269.43	267.74	262.58	271.34	269.25	266.86	262.86
2004					263.24	256.06	257.19	250.03	266.07	260.05	257.77	252.15	275.23	265.67	266.12	258.35	278.29	268.83	269.22	262.94	276.91	268.52	268.06	263.29
Year	July				August				September				October				November				December			
1993					266.75	263.52	261.21	259.05	265.16	263.08	261.06	258.66	251.72	258.23	251.88	250.66	248.51	255.2	247.39	247.14	248.88	251.76	247.34	246.98
1994					269.66	264.29	260.62	259.55	263.57	262.48	259.43	257.63	253.17	258.33	251.28	251.1					250.27	251.66	246.43	246.84
1995	272.76	267.02	264.48	261.11	266.84	264.04	259.97	259.45	264.83	262.98	259.36	257.89	257.85	259.77	254.19	254.6								
1996	270.76	266.28	261.76	260.54	267.45	264.04	259.01	259.09	264.2	262.75	258.02	257	258.54	260.42	253.46	254.01	255.42	254.63	245.21	248.71	256.87	253.44	245.52	247.24
1997																								
1998	271.1	267.77	265.56	261.01	266.54	265.21	260.42	258.79	263.88	263.21	259.25	256.75	257.96	260.02	254.76	252.52	254.54	254.08	246.06	248.85	257.8	252.62	247.72	248.41
1999									261.79	261.86	258.98	255.01	257.16	258.85	254.08	252.47	255.97	254.77	248.54	249.17	258.5	252.94	253.47	248.9
2000	266.8	267.01	264.98	260.17					262.48	263.27	260.24	256.67	259.06	261.04	257.86	254.81	251.96	255.59	249.68	250.18				
2001	267.24	266.98	264.49	260.32	264.46	264.82	260.85	259.01	262.51	261.45	258.46	255.22	260.13	260.27	256.72	253.84								
2002	267.55	267.06	263.92	260.4	264.17	265.06	260.51	259.28	262.1	262.15	258.41	256.52	258.31	261.09	257.87	254.4								
2003	269.59	266.96	263.67	260.51	265.43	265.23	260.82	259.29					257.14	258.01	254.35	251.16	252.54	252.52	249.61	248.55				
2004	273.25	265.15	263.08	259.97	270.23	264.25	261.63	259.56																

Table 16

Standard Deviations of Monthly Average Temperatures (K) per Dataset (REAL Subset)

Year	January				February				March				April				May				June			
	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55
1993		6.19		4.21		5.88		3.30		4.39		4.44		2.88		2.68		1.82		1.02		1.30		1.29
1994	2.25	4.77	6.09	4.94	9.07	5.19	5.06	2.25	3.76	3.58	3.69	2.37		3.12	4.20	1.86	2.66	1.26	1.47	1.45	2.17	1.13	1.51	0.87
1995	9.76	4.85	8.90	5.57	3.47	4.20	3.14	5.04	4.94	2.72	3.30	2.69	4.08	2.52	4.01	1.82		1.21	1.50	3.07	2.22	1.36	1.45	1.06
1996	0.00	7.11	8.05	5.28	4.44	4.64	8.01	2.66		2.44	5.99	3.03		2.70	3.23	3.32		1.27	2.05	2.21		1.37	1.39	1.21
1997	7.46	5.71	6.14	5.52	8.36	5.40	6.49	4.53	4.41	1.99	2.93	6.63	0.73	2.87	3.88	1.58		2.05	2.50	1.26		1.03	1.55	1.47
1998		10.04	8.53	4.82		6.64	6.31	2.63		1.83	5.31	2.57		2.82	1.93	1.46	2.25	1.49	2.11	2.24	1.56	1.52	1.58	1.09
1999	3.99	5.00	8.57	5.09	2.59	7.67	5.47	4.14	3.73	5.36	4.54	3.16	1.93	2.17	2.80	1.74		1.08	2.23	2.65	2.29	1.01	1.13	1.05
2000	0.00	4.65	11.02	6.25		4.82	7.30	11.14	13.23	6.58	5.08	2.37	1.54	1.88	2.22	3.39		1.33	1.01	1.80	2.24	1.15	1.26	0.99
2001	3.43	5.21	4.73	3.85	6.53	9.79	8.84	2.36	3.24	2.44	1.96	7.63	5.61	1.86	3.22	1.38	2.67	2.39	1.96	1.27	1.69	1.96	1.48	0.94
2002		6.69	6.31	3.90	12.64	6.13	5.90	1.85	4.19	2.20	4.31	3.29	0.00	3.33	4.24	2.23	2.37	1.39	1.39	2.51	1.65	1.34	2.04	1.03
2003	9.25	8.70	7.53	5.56	6.70	2.75	4.19	3.44	6.46	2.90	4.43	5.24	4.00	4.85	5.18	1.64	2.67	2.05	1.83	2.74	1.52	1.11	1.18	0.78
2004		3.17	9.75	3.50	14.79	4.22	4.73	6.14	8.24	4.49	3.31	2.90	4.73	4.91	3.77	5.91	2.84	1.28	2.35	1.65	3.46	1.37	1.38	1.14
Year	July				August				September				October				November				December			
1993		1.39		1.38	0.00	1.46	1.94	1.19	1.15	1.24	2.20	1.02	0.00	3.07	4.30	3.28	5.27	4.03	3.82	3.68	2.15	7.24	5.31	4.32
1994		1.25	2.62	1.28	3.74	0.90	1.54	1.38	3.50	1.68	2.08	1.57	3.38	3.35	3.60	2.15		3.82	6.47	4.59	7.14	6.57	6.63	3.81
1995	2.71	1.77	2.18	1.20	2.69	1.28	1.10	1.09	3.13	1.50	1.52	1.03	3.89	2.42	3.00	3.78		4.27	3.21	3.31		5.27	7.97	3.19
1996	3.86	1.46	2.15	1.25	3.29	1.30	1.25	1.48	2.38	1.19	1.52	1.59	4.94	4.13	4.64	2.92	6.33	4.30	4.77	2.00	3.10	12.34	3.65	6.03
1997		1.54	2.87	1.06		1.66	2.73	1.40		1.22	1.33	1.49		3.45	4.60	2.43		5.00	4.46	3.56		7.57	6.29	3.61
1998	2.31	1.16	1.74	0.92	2.02	1.03	1.36	1.28	2.47	1.78	1.92	2.30	4.61	3.14	4.39	2.05	7.38	5.27	5.13	2.16	15.75	4.67	19.01	4.86
1999		1.42	1.28	1.23		1.09	1.53	1.45	1.44	1.92	1.16	1.67	3.82	2.97	4.44	4.05	3.32	3.78	5.67	3.23	8.48	4.40	12.92	3.43
2000	3.31	1.36	1.79	1.09		1.02	1.81	1.23	2.34	1.78	1.76	1.12	0.00	4.88	4.67	1.97	0.00	5.82	3.68	2.77		6.20	9.46	3.81
2001	1.62	1.30	1.67	1.09	1.82	1.67	1.86	1.07	1.23	1.64	2.25	1.51	0.14	2.94	4.35	3.33		2.06	4.59	6.86		9.46	11.35	6.25
2002	3.58	1.37	2.24	1.29	2.92	1.84	1.93	1.43	2.78	1.80	1.87	1.24	0.00	2.97	3.92	2.20		6.47	4.46	3.90		10.64	10.47	2.97
2003	3.05	1.39	2.23	1.12	1.37	1.21	1.49	1.36		1.49	1.16	2.10	3.06	2.85	3.72	2.21	0.23	3.19	6.19	3.56		8.39	8.44	3.29
2004	0.00	1.66	1.96	1.26	0.92	1.15	1.41	1.12		1.29	1.88	1.76		2.52	2.85	3.64		3.24	4.98	5.99		6.52	11.88	8.24

Table 17

Standard Deviations of Monthly Average Temperatures (K) per Dataset (RLID Subset)

Year	January					February					March					April					May					June				
	ALO-USU	ERA-20C	MER RA-2	JRA-55		ALO-USU	ERA-20C	MER RA-2	JRA-55		ALO-USU	ERA-20C	MER RA-2	JRA-55		ALO-USU	ERA-20C	MER RA-2	JRA-55		ALO-USU	ERA-20C	MER RA-2	JRA-55		ALO-USU	ERA-20C	MER RA-2	JRA-55	
1993																														
1994	2.25	2.76	2.81	3.84		9.07	4.21	3.59	2.49		3.76	2.99	4.40	2.61							2.66	2.03	0.67	3.31		2.17	1.03	0.97	0.59	
1995	9.76	7.98	8.97	6.38		3.47	1.00	2.23	7.62		4.94	1.63	4.35	4.09		4.08	1.31	0.84	2.55							2.22	1.33	1.32	1.03	
1996	0.00	0.00	0.00	0.00		4.44	4.41	6.09	1.16																					
1997	7.46	4.22	4.26	5.99		8.36	4.97	7.45	3.92		4.41	1.96	3.03	7.73		0.73	2.11	0.27	0.16											
1998																					2.25	0.74	2.36	0.80		1.56	2.29	1.35	0.93	
1999	3.99	1.23	2.95	12.49		2.59	6.45	0.92	3.07		3.73	5.81	3.46	3.00		1.93	1.94	1.12	0.89							2.29	1.11	0.27	2.26	
2000	0.00	0.00	0.00	0.00							13.23	1.87	1.97	2.38		1.54	2.01	1.13	1.10							2.24	1.17	1.07	1.03	
2001	3.43	3.59	2.38	4.47		6.53	2.25	6.69	2.92		3.24	1.85	0.77	9.32		5.61	1.87	3.46	0.62		2.67	2.56	1.86	1.26		1.69	1.96	1.35	0.62	
2002						12.64	2.75	2.35	1.03		4.19	0.31	0.03	0.28		0.00	0.00	0.00	0.00		2.37	1.89	0.25	2.19		1.65	0.90	2.08	1.56	
2003	9.25	1.54	8.47	4.73		6.70	1.55	2.07	2.17		6.46	3.15	4.38	4.49		4.00	4.84	4.66	1.60		2.67	0.83	1.44	2.74		1.52	1.40	1.14	0.69	
2004						14.79	8.00	3.42	6.41		8.24	2.42	6.54	1.18		4.73	4.75	4.14	6.80		2.84	0.83	1.50	0.92		3.46	1.87	2.11	1.17	
Year	July					August					September					October					November					December				
1993						0.00	0.00	0.00	0.00		1.15	0.59	1.27	0.52		0.00	0.00	0.00	0.00		5.27	2.36	4.92	1.86		2.15	8.33	1.40	3.00	
1994						3.74	0.87	1.39	0.97		3.50	1.31	1.94	1.82		3.38	1.87	2.21	0.86							7.14	7.29	4.91	1.72	
1995	2.71	1.61	1.56	0.80		2.69	1.21	0.95	1.16		3.13	2.12	1.78	1.00		3.89	2.61	3.59	2.13											
1996	3.86	1.41	1.17	1.54		3.29	1.72	1.12	1.07		2.38	0.97	1.42	1.52		4.94	3.66	1.83	2.33		6.33	4.26	4.19	1.37		3.10	16.10	1.77	5.71	
1997																														
1998	2.31	1.31	2.02	0.72		2.02	1.06	1.88	1.08		2.47	2.24	2.98	3.41		4.61	3.02	3.62	2.01		7.38	5.06	4.47	2.54		15.75	2.65	20.62	5.01	
1999											1.44	1.57	0.39	1.05		3.82	2.62	3.28	3.72		3.32	1.97	5.76	4.28		8.48	1.30	9.35	0.85	
2000	3.31	1.20	1.63	1.01							2.34	2.46	2.30	1.21		0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00						
2001	1.62	1.01	1.90	1.09		1.82	1.46	1.81	0.77		1.23	1.61	1.96	1.10		0.14	0.33	2.06	0.62											
2002	3.58	1.56	3.01	1.43		2.92	1.76	2.08	1.69		2.78	1.12	0.86	1.02		0.00	0.00	0.00	0.00											
2003	3.05	1.47	2.42	1.10		1.37	1.21	1.19	0.81							3.06	1.66	2.42	2.36		0.23	0.78	2.05	3.80						
2004	0.00	0.00	0.00	0.00		0.92	1.40	1.02	0.91																					

Table 18

Standard Deviations of Monthly Average Temperatures (K) per Dataset (FIT Subset)

Year	January				February				March				April				May				June			
	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55
1993		0.24		0.45		1.24		0.76		1.98		1.53		1.59		1.27		0.66		0.55		0.43		0.25
1994	1.02	0.24	0.62	0.45	0.95	1.24	1.36	0.76	2.37	1.98	3.45	1.53	2.76	1.59	1.73	1.27	0.66	0.66	0.40	0.55	0.82	0.43	0.36	0.25
1995	1.02	0.26	0.62	0.67	0.95	1.44	1.36	0.55	1.40	1.90	3.43	1.17	0.99	1.43	1.75	1.44	3.08	0.73	0.41	0.71	0.59	0.55	0.39	0.33
1996	0.62	0.25	1.63	0.86	1.72	1.18	1.80	0.27	2.84	1.98	3.53	1.26	1.19	1.41	1.27	1.66	0.99	0.75	0.73	0.34	0.43	0.46	0.41	0.35
1997	0.88	0.13	0.69	0.62	1.56	0.58	1.41	0.24	2.96	2.34	3.36	1.36	1.04	1.54	1.60	1.63	0.73	0.15	0.65	0.36	0.38	0.36	0.47	0.47
1998	1.49	0.18	0.50	0.36	1.63	0.31	0.97	0.32	2.88	2.45	3.17	1.38	0.70	1.57	1.32	1.45	0.66	0.26	0.22	0.35	0.30	0.05	0.24	0.20
1999	2.09	0.14	0.41	0.39	1.35	0.58	0.48	0.41	3.13	2.76	3.41	1.06	0.75	1.42	1.32	1.24	0.45	0.42	0.31	0.75	0.24	0.05	0.26	0.23
2000	1.17	0.11	0.89	0.32	1.58	1.50	1.21	0.58	2.90	2.25	3.18	1.35	1.03	0.80	0.59	1.07	0.55	0.14	0.15	0.44	0.67	0.12	0.31	0.20
2001	0.65	0.59	0.50	0.51	1.28	1.85	0.94	0.22	2.35	1.85	2.58	1.19	2.05	0.64	0.82	1.49	0.27	0.32	0.33	0.55	0.28	0.23	0.22	0.37
2002	0.71	1.31	0.40	0.40	1.72	2.40	0.85	0.71	2.46	1.25	2.28	1.35	1.79	0.56	1.04	1.35	0.14	0.64	0.40	0.55	0.56	0.35	0.45	0.31
2003	2.76	1.46	0.81	0.46	0.90	1.36	0.93	0.47	1.76	1.39	1.67	1.02	2.32	1.40	2.06	1.79	0.28	0.50	0.19	0.99	0.30	0.43	0.52	0.47
2004	2.75	1.35	0.88	0.08	0.92	1.44	0.73	0.29	1.82	1.40	1.61	1.26	2.27	1.37	2.03	1.69	0.29	0.48	0.21	0.90	0.28	0.47	0.60	0.47
Year	July				August				September				October				November				December			
1993		0.85		0.58	0.00	0.66	0.63	0.35	2.50	0.68	0.76	0.95	2.24	1.07	2.47	1.86	0.81	1.13	1.08	0.93	0.63	0.79	0.17	0.37
1994	0.37	0.89	1.49	0.57	1.23	0.54	0.63	0.36	2.50	0.80	0.76	0.95	2.24	1.03	2.47	1.78	0.81	1.03	1.08	1.04	0.63	0.97	0.17	0.36
1995	1.39	0.89	1.55	0.40	1.27	0.50	0.49	0.46	1.73	0.80	1.05	0.91	1.84	1.34	2.45	1.53	0.17	0.65	0.35	1.26	0.50	0.46	0.35	0.54
1996	1.16	0.85	1.88	0.36	0.97	0.64	0.45	0.60	1.44	0.46	0.86	0.83	2.40	1.55	2.70	1.44	1.19	1.02	0.77	1.02	0.96	0.38	0.55	0.49
1997	1.74	0.66	1.52	0.27	0.79	0.72	0.54	0.73	1.23	0.39	0.71	0.98	2.19	1.85	2.78	1.34	0.54	1.12	0.69	0.62	0.61	0.91	1.28	0.50
1998	1.64	0.64	1.51	0.63	0.85	0.78	0.97	0.59	0.97	0.30	0.21	0.76	2.05	1.73	2.87	1.42	0.28	1.37	1.37	0.63	2.24	0.62	2.17	0.41
1999	1.66	0.80	1.69	0.77	0.51	0.55	0.73	0.49	1.17	0.72	0.62	0.87	1.88	1.63	2.66	1.32	1.47	0.80	0.49	0.64	0.99	0.12	1.50	0.37
2000	1.74	0.90	1.89	0.65	0.36	0.51	0.60	0.54	1.30	0.83	0.76	0.75	2.05	1.54	2.43	1.29	0.51	0.90	0.43	0.84	0.23	0.36	1.38	0.45
2001	1.24	0.88	1.85	0.44	0.31	0.64	0.56	0.61	0.99	0.93	0.89	1.10	2.66	1.02	2.22	1.31	0.45	1.18	0.38	0.64	1.45	1.18	1.17	0.08
2002	1.13	0.73	1.33	0.51	1.24	0.84	1.08	0.49	1.31	0.56	0.08	0.97	2.64	1.17	1.80	1.28	1.06	1.95	1.21	0.76	2.84	0.60	2.22	0.15
2003	1.10	0.73	1.18	0.69	3.04	0.81	0.97	0.33	0.64	0.65	0.15	0.97	0.53	1.17	1.54	1.48	1.94	1.81	1.01	0.50	0.72	0.66	1.98	0.31
2004	1.18	0.74	1.20	0.68	0.97	0.81	0.93	0.33		0.65	0.16	1.01		1.20	1.59	1.47		1.81	0.93	0.46		0.60	2.02	0.31

Table 19

Standard Deviations of Monthly Average Temperatures (K) per Dataset (FITL Subset)

Year	January				February				March				April				May				June			
	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55	ALO-USU	ERA-20C	MER RA-2	JRA-55
1993																								
1994	0.29	0.08	0.20	0.12	0.42	0.58	0.71	0.36	2.62	2.11	3.67	1.64					0.44	0.34	0.19	0.28	0.26	0.21	0.24	0.14
1995	1.10	0.13	0.68	0.73	0.41	0.60	0.79	0.23	1.58	2.30	4.13	1.41	0.73	0.42	0.68	0.39					0.26	0.39	0.24	0.25
1996	0.00	0.00	0.00	0.00	0.53	0.37	0.30	0.13																
1997	0.75	0.10	0.63	0.54	1.27	0.46	1.19	0.20	3.79	2.90	4.11	1.72	0.85	0.23	0.25	0.21								
1998																	0.04	0.21	0.12	0.05	0.33	0.03	0.26	0.23
1999	3.55	0.03	0.45	0.75	1.86	0.72	0.63	0.41	2.01	1.94	2.40	0.74	0.08	0.20	0.18	0.18					0.10	0.01	0.06	0.05
2000	0.00	0.00	0.00	0.00					0.50	0.64	0.88	0.46	0.09	0.21	0.21	0.22					0.77	0.12	0.32	0.22
2001	0.86	0.66	0.52	0.37	0.93	2.08	1.26	0.28	2.55	1.97	2.76	1.23	1.69	0.45	0.49	1.15	0.35	0.09	0.18	0.07	0.05	0.00	0.02	0.03
2002					1.78	0.63	0.19	0.18	0.04	0.10	0.18	0.11	0.00	0.00	0.00	0.00	0.22	0.82	0.54	0.71	0.33	0.26	0.33	0.23
2003	2.91	0.75	0.70	0.23	0.44	0.85	0.30	0.35	1.80	1.48	1.69	1.05	1.76	1.08	1.54	1.40	0.24	0.52	0.20	1.00	0.33	0.51	0.62	0.55
2004					0.50	0.94	0.51	0.20	1.00	0.84	0.87	0.73	2.28	1.40	2.03	1.74	0.19	0.21	0.06	0.45	0.34	0.33	0.47	0.27
Year	July				August				September				October				November				December			
1993					0.00	0.00	0.00	0.00	0.68	0.18	0.08	0.18	0.00	0.00	0.00	0.00	0.21	0.35	0.26	0.24	0.13	0.12	0.03	0.08
1994					1.05	0.35	0.30	0.27	2.28	0.73	0.69	0.87	0.91	0.52	1.22	0.86					0.42	0.40	0.02	0.27
1995	0.75	0.42	0.72	0.21	0.56	0.33	0.19	0.39	0.55	0.29	0.27	0.40	1.23	0.75	1.44	0.75								
1996	1.08	0.51	1.01	0.25	1.04	0.70	0.55	0.61	1.42	0.48	0.90	0.88	2.18	1.15	2.12	1.10	1.21	1.16	0.94	1.14	0.32	0.25	0.61	0.18
1997																								
1998	1.38	0.57	1.30	0.55	0.57	0.42	0.65	0.33	0.92	0.28	0.24	0.74	1.81	1.45	2.42	1.24	0.19	1.09	1.23	0.51	2.45	0.46	1.75	0.30
1999									0.49	0.28	0.42	0.26	1.62	1.40	2.30	1.13	1.03	0.72	0.51	0.60	2.21	0.16	1.73	0.58
2000	1.50	0.82	1.68	0.57					0.82	0.49	0.33	0.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
2001	1.35	0.95	2.00	0.48	0.36	0.67	0.63	0.61	0.84	0.59	0.72	0.70	0.35	0.09	0.16	0.10								
2002	1.26	0.88	1.61	0.60	1.25	0.68	0.96	0.36	0.92	0.41	0.02	0.66	0.00	0.00	0.00	0.00								
2003	1.19	0.80	1.28	0.75	1.45	0.41	0.62	0.17					0.09	0.54	0.73	0.53	0.68	0.59	0.15	0.08				
2004	0.00	0.00	0.00	0.00	0.53	0.14	0.21	0.06																

APPENDIX D. ORIGINAL MATLAB CODE

adaptivefit.m:

```
function [tvals,dailymodel,eqn,nudatas] =
adaptivefit(datas,Ncoefficients)
% Performs a day-by-day adaptive fit to a dataset
% Put datas array as time vs response

maxdays = max(datas(1,:)) - min(datas(1,:)) + 1;
tvals = min(datas(1,:)):max(datas(1,:));

% Set flag if data missing within timespan of data
if maxdays > length(datas(1,:))
    sparse = 1;
else
    sparse = 0;
end

% Make sure timespan is long enough for 3-year algorithm
if maxdays < 1095
    beep;
    fprintf(2,'Check urself b4 u wreck urself');
    return
else
    dailymodel = cell(1,maxdays);
end

if sparse == 1
    nudatas(1,:) = (datas(1,:) - 727930) / 365; % Only lidar data is
sparse
    tvals = (tvals - 727930) / 365;
else
    nudatas(1,:) = datas(1,:) / 365;
    tvals = tvals / 365;
end

nudatas(2,:) = datas(2,:);

if sparse == 1

    j = 1;

    while datas(1,j) <= datas(1,1)+1095
        tempdata = nudatas(:,j);
        j = j + 1;
    end

    todaysmod = modelcalc(tempdata,Ncoefficients);
    todayscoeff = todaysmod.Coefficients.Estimate;
```

```
    eqn = todaysmod.Formula.ModelFun;

    for i = 1:548
        dailymodel{i} = todayscoeff;
    end

    for i = 549:maxdays-547
        clear tempdata;
        j = find(datas(1,:)<datas(1,1)+i,1,'last');
        k = find(datas(1,:)<=datas(1,j)+547+i,1,'last');
        tempdata = nudatas(:,j:k);
        todaysmod = modelcalc(tempdata,Ncoefficients);
        todayscoeff = todaysmod.Coefficients.Estimate;
        dailymodel{i} = todayscoeff;
    end

else

    tempdata = nudatas(:,1:1095);
    todaysmod = modelcalc(tempdata,Ncoefficients);
    todayscoeff = todaysmod.Coefficients.Estimate;
    eqn = todaysmod.Formula.ModelFun;

    for i = 1:548
        dailymodel{i} = todayscoeff;
    end

    for i = 549:maxdays-547
        tempdata = nudatas(:,i-547:i+547);
        todaysmod = modelcalc(tempdata,Ncoefficients);
        todayscoeff = todaysmod.Coefficients.Estimate;
        dailymodel{i} = todayscoeff;
    end

end

for i = maxdays-546:maxdays
    dailymodel{i} = todayscoeff;
end

end
```

modelcalc.m:

```
function modelout = modelcalc(datas,Ncoefficients)
% calculates the daily fit model for adaptive fit

warning('off','stats:nlinfit:IllConditionedJacobian');
```

```
par0 = ones(Ncoefficients,1);
% Set par1 if initial guesses desired, otherwise comment line out
% The below worked for CEDAR
%par1 = [260;10;181;1.25;140;1;65;1;14]; par0 = par1(1:Ncoefficients);
par1 = [260;10;181;1.25;140;1;65;1;14;1;10;1;5;1]; par0 =
par1(1:Ncoefficients);

switch Ncoefficients
    case 3
        %predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)/12);
        predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3));
    case 5
        %predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)/12) +
p(4)*sin(2*pi*t + p(5)/6);
        predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)) +
p(4)*sin(4*pi*t + p(5));
    case 7
        %predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)/12) +
p(4)*sin(2*pi*t + p(5)/6) + p(6)*sin(2*pi*t + p(7)/3);
        predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)) +
p(4)*sin(4*pi*t + p(5)) + p(6)*sin(6*pi*t + p(7));
    case 9
        %predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)/12) +
p(4)*sin(2*pi*t + p(5)/6) + p(6)*sin(2*pi*t + p(7)/3) + p(8)*sin(2*pi*t
+ p(9)/1.5);
        predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)) +
p(4)*sin(4*pi*t + p(5)) + p(6)*sin(6*pi*t + p(7)) + p(8)*sin(8*pi*t +
p(9));
    case 10
        predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)) +
p(4)*sin(4*pi*t + p(5)) + p(6)*sin(6*pi*t + p(7)) + p(8)*sin(8*pi*t +
p(9)) + p(10)*t;
    case 12
        predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)) +
p(4)*sin(4*pi*t + p(5)) + p(6)*sin(6*pi*t + p(7)) + p(8)*sin(8*pi*t +
p(9)) + p(10)*sin(12*pi*t + p(11)) + p(12)*t;
    case 14
        predicted = @(p,t) p(1) + p(2)*sin(2*pi*t + p(3)) +
p(4)*sin(4*pi*t + p(5)) + p(6)*sin(6*pi*t + p(7)) + p(8)*sin(8*pi*t +
p(9)) + p(10)*sin(12*pi*t + p(11)) + p(12)*sin(24*pi*t + p(13))+
p(14)*t;
    otherwise
        beep;
        fprintf(2,'Check urself b4 u wreck urself');
        return
end

predictor = datas(1,:);
response = datas(2,:);
modelout = fitnlm(predictor,response,predicted,par0);

end
```

plotafit.m:

```
function [resids,fighandle,fitted] =
plotafit(tvals,dailymodel,eqn,nudatas)
% Plots a dataset with its adaptive fit
% Get required inputs from adaptivefit.m

yvals = zeros(1,length(tvals));
%clear yvals
resids = zeros(1,length(nudatas(1,:)));
temp = 0;

%for i = 1:200
for i = 1:length(tvals)
    yvals(1,i) = eqn(dailymodel{i},tvals(i));
    j = find(nudatas(1,:)==tvals(i));
    if ~isempty(j)
        temp = temp + 1;
        resids(temp) = yvals(1,i) - nudatas(2,j);
    end
end

figure;
fighandle =
plot(nudatas(1,:),nudatas(2,:),'k.',tvals,yvals,'r','LineWidth',1.7,'Ma
rkerSize',12);
%fighandle =
plot(nudatas(1,1:200),nudatas(2,1:200),'k.',tvals(1:200),yvals,'r','Lin
eWidth',1.7,'MarkerSize',12);
xlim([0 12]);
pos = get(gcf, 'Position');
set(gcf, 'Position', [pos(1) pos(2) 14.5*100 5*100]);
xlabel('Year','FontSize',15);
ylabel('Temperature (K)', 'FontSize',15);
title('Adaptive Model','FontSize',15);
set(gca, 'FontSize',15,'LineWidth',1.5);
%legend('Avg. Nightly Temperature','Regression
Fit','Location','SouthWest');
set(gca,'XTickLabel',{'1993','1995','1997','1999','2001','2003',[]});
%set(gca,'YTickLabel',{[],'230','240','250','260','270','280',[]});

end
```

quickmarkoutliers.m:

```
function graphout = quickmarkoutliers(commonout,varargin)
% this function can be used to mark outliers of temp data
% Pass model data as model{i}
if length(commonout(:,1)) ~= length(varargin)+1
```

```
    fprintf(2,'Include each model as arguments.');
    return
end

graphout = commonout;

for i = 2:length(graphout(:,1))

    otherout = table2array(varargin{i-1}.Variables)';
    temp = 0;

    for j = 1:length(graphout(1,:))
        k = find(otherout(1,:)==graphout(1,j));
        graphout(i,j) = otherout(2,k);
    end

end


end
```

coefout.m:

```
function coefout(models)
% Exports coefficients from model fits to a table for easy access

data = zeros(models{1}.NumCoefficients,2*length(models));
offset = 0;

if exist('coefficients.xls') == 2
    delete('coefficients.xls');
end

for i = 1:length(models)
    data(:,i+offset) = models{i}.Coefficients.Estimate;
    data(:,i+offset+1) = models{i}.Coefficients.SE;
    offset = offset + 1;
end

offset = 0;
offset2 = length(data(:,1)) + 1;

for i = 1:length(models)
    data(offset2,i+offset) = models{i}.Rsquared.Adjusted;
    data(offset2+1,i+offset) = models{i}.NumObservations;
    offset = offset + 1;
end
```

```
dataT = array2table(data);
writetable(dataT,'coefficients.xls');

end
```

quickdiff.m:

```
function [avgdiff,stddev] = quickdiff(nudatassparse,nudatasfull)
% Gives the average difference from a sparse array (lidar) to a full
array

    temparray = zeros(1,length(nudatassparse(1,:)));
    offset = 0;

    for i = 1:length(nudatassparse(1,:))
        j = find(nudatasfull(1,:)==nudatassparse(1,i));
        if isempty(j)
            %temparray(1,end) = [];
            %offset = offset + 1;
        else
            temparray(1,i-offset) = nudatassparse(2,i) -
nudatasfull(2,j);
        end
    end

    temparray(temparray==0) = [];
    avgdiff = mean(temparray);
    stddev = std(temparray);

end
```

compareoutliers.m:

```
function commonout = compareoutliers(varargin)
% Used to find common outliers between datasets
% Function arguments should be 2xN matrices containing outliers and
their
% timestamps. These are produced from outliers.m as an output.

commonout = flipud(varargin{1});

for i = 2:length(varargin)

    otherout = flipud(varargin{i});
    temp = 0; temparray = [];
```

```
    for j = 1:length(commonout(1,:))
        k = find(otherout(1,:)==commonout(1,j));
        if isempty(k)
            temp = temp + 1;
            temparray(temp) = j;
        else
            commonout(i+1,j) = otherout(2,k);
        end
    end

    if ~isempty(temparray)
        commonout(:,temparray) = [];
    end

    if isempty(commonout)
        beep;
        fprintf(2,'There are no common outliers between the
datasets.\n');
        return
    end

end

end
```

compositeyear.m:

```
function modelcompositeyear = compositeyear(modelnudatas)
% takes in a 2 x days sized matrix with dates as decimal values in the
% first row and temps in the second row, spits out a single-year
composite
% year average and std dev of 31 days x 7 years.

datenums = modelnudatas(1,:)*365 + 727930; % convert decimal dates to
whole number datenums
modelcompositeyear = zeros(3,365);

% get dates +/- 15 days

for i = 1:365

    modelcompositeyear(1,i) = (i - 1);
    datevalues = zeros(12,31);

    for j = 1:12

        for h = -15:15

            trydate = (((j - 1)*365) + (i - 1)) + h + 727930;
```

```
            % Find the date within a tolerance to avoid errors
            hello = find(abs(datenums - trydate) < 0.1);

            if hello > 0
                datevalues(j,h+16) = modelnudatas(2,hello);
            else
                datevalues(j,h+16) = NaN;
            end

        end

    end

    goodnums = ~isnan(datevalues);
    goodvals = datevalues(goodnums);
    modelcompositeyear(2,i) = mean(goodvals(:));
    modelcompositeyear(3,i) =
3*(std(goodvals(:))/sqrt(length(goodvals))); % 99% confidence error
bars

end

end
```

monthdivide.m:

```
function modelmonthly = monthdivide(modelnudatas)
% takes in nudatas and spits out a cell array 11 x 12 x 2 where each
cell is a
% numerical array of the dates and temps for that month of that year

modelmonthly = cell(12,12,2);
datenums = modelnudatas(1,:)*365 + 727930; % convert decimal dates to
whole number datenums
runningdate = 727930;

for i = 1:12

    for j = 1:12

        switch j
            case 1
                totdays = 31;
            case 2
                if leapyear(1992 + i)
                    totdays = 29;
                else
                    totdays = 28;
                end
```

```
            case 3
                totdays = 31;
            case 4
                totdays = 30;
            case 5
                totdays = 31;
            case 6
                totdays = 30;
            case 7
                totdays = 31;
            case 8
                totdays = 31;
            case 9
                totdays = 30;
            case 10
                totdays = 31;
            case 11
                totdays = 30;
            case 12
                totdays = 31;
        end

        locales = find(datenums >= runningdate & datenums <
runningdate+totdays);
        modelmonthly{i,j,1} = modelnudatas(1,locales);
        modelmonthly{i,j,2} = modelnudatas(2,locales);
        runningdate = runningdate + totdays;
        clear locales totdays

    end

end

end
```

monthlyport.m:

```
function exportables = monthlyport(modelmonthly)
% takes in a model's monthly cell array and exports mean, std dev, and
% total data count to an excel spreadsheet for quick copying

exportables = zeros(12,12,3);

for i = 1:12
    for j = 1:12
        exportables(i,j,1) = mean(modelmonthly{i,j,2});
        exportables(i,j,2) = std(modelmonthly{i,j,2});
        exportables(i,j,3) = length(modelmonthly{i,j,2});
    end
end
```

```
totalout = [exportables(:,:,1);exportables(:,:,2);exportables(:,:,3)];

csvwrite('exported.csv',totalout(:,:,1),0,0);

end
```

throwshade.m:

```
function throwshade(differ,p)
% this function plots a monthly difference plot and then puts hatchings
on
% it to indicate confidence levels

% Solid color = >99%
% Speckles = >95%
% Lines = >=90%
% Hatch = <90%
differ = differ*-1;
figure
h1 = imagesc(differ);
axis equal
xlim([0.5 12.5])
cmap = jet;
colormap(cmap);

% Gets the RGB map
Av = differ(:);
[~,bin] = histc(Av,linspace(-10,15,min(numel(Av),size(cmap,1)))); % The
-10 and 15 in this refer to the caxis values set outside the function
bin(bin==0) = 1;
mapp = permute(cmap,[1 3 2]);
rgb = reshape(mapp(bin,:,:),12,12,3);
clear map Av bin mapp

% Sets NaNs to white
set(h1,'AlphaData',~isnan(differ));

for i = 1:12
    for j = 1:12
        if ~isnan(differ(i,j))

            rgbpix = rgb(i,j,:);
            h2 = patch('XData',[j-0.5 j+0.5 j+0.5 j-0.5],'YData',[i-0.5
i-0.5 i+0.5 i+0.5],'FaceColor',rgbpix,'EdgeColor','none');

            % hatchfill2 is a function by Kesh Ikuma and Neil Tandon
            % found on the MATLAB File Exchange at:
```

```
            %
https://www.mathworks.com/matlabcentral/fileexchange/53593-hatchfill2
            if p(i,j) > 0.1

hatchfill2(h2,'cross','HatchAngle',25,'HatchDensity',100,'HatchColor','
k','HatchLineWidth',1);
            elseif p(i,j) > 0.05

hatchfill2(h2,'cross','HatchAngle',45,'HatchDensity',50,'HatchColor','k
','HatchLineWidth',1);
            elseif p(i,j) > 0.01

hatchfill2(h2,'single','HatchAngle',65,'HatchDensity',70,'HatchColor','
k','HatchLineWidth',1);
            end

        end
    end
end

end
```

Additional assorted code scraps, including those for generating figures:

```
% netcdf is fickle in matlab for some reason
% DO NOT RUN THIS FILE, run each line of code seperately as its own
command
% or it will freeze and crash the computer

% for reading in era20c data

eralat = ncread('ERA20c.nc','latitude');
eralev = ncread('ERA20c.nc','level');
eralon = ncread('ERA20c.nc','longitude');
eratime = ncread('ERA20c.nc','time');

temps = ncread('ERA20c.nc','t',[249 49 1 1],[2 2 Inf Inf]);
% erageo = ncread('ERA20c.nc','z');

lats = eralat(49:50); clear eralat; % targets correct lats 42-41
long = eralon(249:250); clear eralon; % targets correct longs 248-249
temps = ncread('ERA20c.nc','t',[249 49 1 1],[2 2 Inf Inf]);
geos = ncread('ERA20c.nc','z',[249 49 1 1],[2 2 Inf Inf]);

% Size of geos and temps is (2,2,4,4383), corresponding with longitude,
% latitude, vertical level, and time so lats/longs:
%    [ 42/248 41/248 ]
%    [ 42/249 41/249 ]
% with each third index increment corresponding to lower level

% target location of USU-ALO is
```

```
lat = 41.74; lon = 248.19;

% unit correction for ERA ONLY
geos = geos ./ 10;

% Convert geos into geometric height
geoH = (6368600 .* geos) ./ (6368600 - geos); clear geos;

% Collapse lats and longs into target location
% requires some weird grid interpolations due to my low understanding
of
% how to set this up properly
% target location value corresponds with row(lon) 20, column(lat) 27 on
% interpolated grid (where x=1.19 and y=3.26)

[X,Y] = ndgrid(1:2,3:4);
[Xfine,Yfine] = ndgrid(1:0.01:2,3:0.01:4);
localtemps = zeros(4,4383);
localgeoH = zeros(4,4383);

for i = 1:4
    for j = 1:4383
        if i == 1 && j == 1
            F = griddedInterpolant(X,Y,temps(:,:,1,1),'linear');
        else
            F.Values = temps(:,:,i,j);
        end
        finegrid = F(Xfine,Yfine);
        localtemps(i,j) = finegrid(20,27);
    end
end

for i = 1:4
    for j = 1:4383
        if i == 1 && j == 1
            F = griddedInterpolant(X,Y,geoH(:,:,1,1),'linear');
        else
            F.Values = geoH(:,:,i,j);
        end
        finegrid = F(Xfine,Yfine);
        localgeoH(i,j) = finegrid(20,27);
    end
end

clear F finegrid temps geoH Xfine Yfine X Y i j lat lon lats long

% At this point I can interpolate with
% pressure, geoH, and temps, to get the temps at 45km.
% First interpolate the pressure at 45km

level = double(level);

for i = 1:4383
    p45(i) = spline(localgeoH(:,i),level,45000);
```

```
end

% This interpolates temperatures at the 45km pressures
% Make sure 4 temps  are in 4x4383 matrix called localtemps

for i = 1:4383
    t45(i) = spline(level,localtemps(:,i),p45(i));
end

% If we have each day, time variable is irrelevant for ERA data
clear localtemps localgeoH level i p45 time

% format it in the dumb way I decided to before to make other functions
% still work
ERA45kmT(2,:) = t45;
ERA45kmT(1,:) = (0:(length(t45) - 1))';
clear t45



%% Merrraaaa
% MERRA2 data is already in .mat chunks from python read-in, with
anywhere
% with a dimension of 5 length referring to 3-hr time increment (we
want the
% middle one to correspond with midnight
% Additionally, pressure is in 12 chunks, so our spline has more to
work
% with if desired
% Luckily, there is no geographic interpolation with this one

% Just get midnight values
geop = squeeze(merraH(:,3,:));
temp = squeeze(merraT(:,3,:));
clear merraH merraT merratimes

% temp and geop need to be flipped to work with previously established
% geopotential/pressure fitting method

geop = geop';
temp = temp';

% Convert geop into geometric height
geoH = (6368600 .* geop) ./ (6368600 - geop); clear geop;

% Now do it (only 4171 dates)

for i = 1:4171
    p45(i) = spline(geoH(:,i),merraP,45000);
end

% This interpolates temperatures at the 45km pressures
% Make sure 4 temps  are in 4x4383 matrix called localtemps

for i = 1:4171
```

```
    t45(i) = spline(merraP,temp(:,i),p45(i));
end

% If we have each day, time variable is irrelevant for ERA data
clear temp geoH merraP i p45

% format it in the dumb way I decided to before to make other functions
% still work
MER45kmT(2,:) = t45;
MER45kmT(1,:) = (0:(length(t45) - 1))';
clear t45

%% JRA grib files converted in python
% Got the correct file in Excel, need to read in

uiimport

% define pressure array
presh = [1,2,5,10];

% There are 3 hr increments, and each level is in its own variable
% we just want midnight for now
j = 1;
for i = 1:4:17529
    temps(1,j) = t_lvl_1(i);
    j = j + 1;
end
j = 1;
for i = 1:4:17529
    temps(2,j) = t_lvl_2(i);
    j = j + 1;
end
j = 1;
for i = 1:4:17529
    temps(3,j) = t_lvl_5(i);
    j = j + 1;
end
j = 1;
for i = 1:4:17529
    temps(4,j) = t_lvl_10(i);
    j = j + 1;
end
j = 1;
for i = 1:4:17529
    geop(1,j) = z_lvl_1(i);
    j = j + 1;
end
j = 1;
for i = 1:4:17529
    geop(2,j) = z_lvl_2(i);
    j = j + 1;
end
j = 1;
for i = 1:4:17529
```

```
    geop(3,j) = z_lvl_5(i);
    j = j + 1;
end
j = 1;
for i = 1:4:17529
    geop(4,j) = z_lvl_10(i);
    j = j + 1;
end

clear z_lvl_1 z_lvl_2 z_lvl_5 z_lvl_10 t_lvl_1 t_lvl_2 t_lvl_5 t_lvl_10
i j

% Convert geop into geometric height
geoH = (6368600 .* geop) ./ (6368600 - geop); clear geop;

% Now do it (only 4171 dates)

for i = 1:4383
    p45(i) = spline(geoH(:,i),presh,45000);
end

% This interpolates temperatures at the 45km pressures
% Make sure 4 temps  are in 4x4383 matrix called localtemps

for i = 1:4383
    t45(i) = spline(presh,temps(:,i),p45(i));
end

% If we have each day, time variable is irrelevant for ERA data
clear temps geoH presh i p45

% format it in the dumb way I decided to before to make other functions
% still work
JRA45kmT(2,:) = t45;
JRA45kmT(1,:) = (0:(length(t45) - 1))';
clear t45



% MERRA and MERRA 2
mean(MERnudatas(2,:))
mean(MERRA2nudatas(2,:))
MERyvals = zeros(1,length(MERtvals));
for i = 1:length(MERtvals)
MERyvals(1,i) = eqn(MERdailymodel{i},MERtvals(i));
end
MERRA2yvals = zeros(1,length(MERRA2tvals));
for i = 1:length(MERRA2tvals)
MERRA2yvals(1,i) = eqn(MERRA2dailymodel{i},MERRA2tvals(i));
end
mean(MERyvals)
mean(MERRA2yvals)

% JRA
mean(JRAnudatas(2,:))
```

```
JRAyvals = zeros(1,length(JRAtvals));
for i = 1:length(JRAtvals)
JRAyvals(1,i) = eqn(JRAdailymodel{i},JRAtvals(i));
end
mean(JRAyvals)

% lidar
mean(LIDnudatas(2,:))
LIDyvals = zeros(1,length(LIDtvals));
for i = 1:length(LIDtvals)
LIDyvals(1,i) = eqn(LIDdailymodel{i},LIDtvals(i));
end
mean(LIDyvals)

% this gets data just for lidar dates
rightdate = ismember(MERnudatas(1,:),LIDnudatas(1,:));
rightdate(2,:) = rightdate;
MERnumatches = MERnudatas.*rightdate;
MERnumatches(:,all(MERnumatches==0)) = [];
clear rightdate
rightdate = ismember(CPCnudatas(1,:),LIDnudatas(1,:));
rightdate(2,:) = rightdate;
CPCnumatches = CPCnudatas.*rightdate;
CPCnumatches(:,all(CPCnumatches==0)) = [];
clear rightdate
rightdate = ismember(JRAnudatas(1,:),LIDnudatas(1,:));
rightdate(2,:) = rightdate;
JRAnumatches = JRAnudatas.*rightdate;
JRAnumatches(:,all(JRAnumatches==0)) = [];
clear rightdate
rightdate = ismember(ERAnudatas(1,:),LIDnudatas(1,:));
rightdate(2,:) = rightdate;
ERAnumatches = ERAnudatas.*rightdate;
ERAnumatches(:,all(ERAnumatches==0)) = [];
clear rightdate
rightdate = ismember(newERAnudatas(1,:),LIDnudatas(1,:));
rightdate(2,:) = rightdate;
newERAnumatches = newERAnudatas.*rightdate;
newERAnumatches(:,all(newERAnumatches==0)) = [];
clear rightdate

% frustratingly, the above code doesn't seem to work for MERRA2?
% with some inspection, it appears the time data resolution is greater
for
% all other models, something went wrong with the import of the MERRA2
% dates, probably due to data precision (single vs double)
% copying over matching date data from ERA20C model
MERRA2nudatas = double(MERRA2nudatas);
MERRA2nudatas(1,:) = newERAnudatas(1,1:4171);

% this also implies a need to rerun previous code that seperated out
% outliers and such
MERRA245kmT = double(MERRA245kmT);
clear MERRA2hardout
```

```
[MERRA2tvals,MERRA2dailymodel,eqn,MERRA2nudatas] =
adaptivefit(MERRA245kmT,10);
MERRA2resids =
plotafit(MERRA2tvals,MERRA2dailymodel,eqn,MERRA2nudatas);
a = find(abs(MERRA2resids)>10);
MERRA2hardout(1,:) = MERRA2nudatas(1,a);
MERRA2hardout(2,:) = MERRA2nudatas(2,a);
MERRA2hardout(3,:) = MERRA2resids(a);

% redoing previous Merra 2 avg
mean(MERRA2nudatas(2,:))
MERRA2yvals = zeros(1,length(MERRA2tvals));
for i = 1:length(MERRA2tvals)
MERRA2yvals(1,i) = eqn(MERRA2dailymodel{i},MERRA2tvals(i));
end
mean(MERRA2yvals)

% now can run it
rightdate = ismember(MERRA2nudatas(1,:),LIDnudatas(1,:));
rightdate(2,:) = rightdate;
MERRA2numatches = MERRA2nudatas.*rightdate;
MERRA2numatches(:,all(MERRA2numatches==0)) = [];
clear rightdate

% Lets also get datasets for all the adaptive fits on just lidar dates
% make a residual variable for each dataset also that matches lidar
dates,
% because it could be useful
logics = ismember(MERtvals,LIDnudatas(1,:));
MERlidtvals = MERtvals(logics);
MERlidyvals = MERyvals(logics);
MERlidsids = MERresids(logics);
clear logics
logics = ismember(CPCtvals,LIDnudatas(1,:));
CPClidtvals = CPCtvals(logics);
CPClidyvals = CPCyvals(logics);
CPClidsids = CPCresids(logics);
clear logics
logics = ismember(ERAtvals,LIDnudatas(1,:));
ERAlidtvals = ERAtvals(logics);
ERAlidyvals = ERAyvals(logics);
ERAlidsids = ERAresids(logics);
clear logics
logics = ismember(MERRA2tvals,LIDnudatas(1,:));
MERRA2lidtvals = MERRA2tvals(logics);
MERRA2lidyvals = MERRA2yvals(logics);
MERRA2lidsids = MERRA2resids(logics);
clear logics
logics = ismember(newERAtvals,LIDnudatas(1,:));
newERAlidtvals = newERAtvals(logics);
newERAlidyvals = newERAyvals(logics);
newERAlidsids = newERAresids(logics);
clear logics
logics = ismember(JRAtvals,LIDnudatas(1,:));
```

```
JRAlidtvals = JRAtvals(logics);
JRAlidyvals = JRAyvals(logics);
JRAlidsids = JRAresids(logics);
clear logics
logics = ismember(LIDtvals,LIDnudatas(1,:));
LIDlidtvals = LIDtvals(logics);
LIDlidyvals = LIDyvals(logics);
% LIDresids is already this
clear logics

% now can retrieve averages for just lidar dates
% lidar fit
mean(LIDlidyvals)

% CPC
mean(CPClidyvals)

% JRA
mean(JRAlidyvals)

% ERA
mean(ERAlidyvals)
mean(newERAlidyvals)

% MERRA
mean(MERlidyvals)
mean(MERRA2lidyvals)

% all averages get closer to lidar values, but only ERA20C has dramatic
% change of about 1K bringing it around 1.5K lower than actual lidar
data

%% Try making a chart to show residuals as a heat map

for i = 1:12
    stackeddata(i,:) = JRAresids((JRAtvals < i) & (JRAtvals >= i-1));
end
imagesc(stackeddata)

% just realized this is only residuals of model fit to its own model
data
% I want to get residuals for each set of data to the lidar data and
make a
% heat map of THAT, let's organize my variables so this doesn't happen
% again

%% VARIABLE REFERENCE

% 45kmT - variable containing days as integers and corresponding
% temperature data at 45 km from source

% nudatas - just like 45kmT but the days have been converted to decimal
% fractions representing day of the year in the 12 year dataset
```

```
% numatches - nudatas but just lidar dates

% tvals - just the date decimals from nudatas

% lidtvals - tvals but just lidar dates

% tempvals - just the temperatures from nudatas

% lidtempvals - tempvals but just lidar dates

% dailymodel - cell structure containing values for each of the model
fit
% parameters on a daily basis (from adaptive fit)

% eqn - function that applies the fit model from parameters in
dailymodel

% yvals - fitted temperature values from applying dailymodel parameters

% lidyvals - yvals but just lidar dates

% resids - residuals from yvals to tempvals

% lidsids - resids but just lidar dates

% hardout - contains date decimal, actual temperature, and resids value
for
% dates with 10 K +/- residuals

% directresids - residuals from model data to lidar data (just lidar
dates)

% modsids - residuals from model data to lidar fit; keep in mind these
% match with the lidar model fit dates

% lidmodsids - modsids but just lidar dates

% modelsids - residuals from model fits to lidar fit; keep in mind
these
% match with the lidar model fit dates

% lidmodelsids - modelsids but just lidar dates

%% Creating some of these that I didn't have already

ERAtempvals = ERAnudatas(2,:); ERAlidtempvals = ERAnumatches(2,:);
CPCtempvals = CPCnudatas(2,:); CPClidtempvals = CPCnumatches(2,:);
JRAtempvals = JRAnudatas(2,:); JRAlidtempvals = JRAnumatches(2,:);
LIDtempvals = LIDnudatas(2,:);
MERtempvals = MERnudatas(2,:); MERlidtempvals = MERnumatches(2,:);
MERRA2tempvals = MERRA2nudatas(2,:); MERRA2lidtempvals =
MERRA2numatches(2,:);
newERAtempvals = newERAnudatas(2,:); newERAlidtempvals =
newERAnumatches(2,:);
```

```
ERAdirectresids = LIDtempvals - ERAlidtempvals;
ERAlidmodsids = LIDlidyvals - ERAlidtempvals;
ERAlidmodelsids = LIDlidyvals - ERAlidyvals;
JRAdirectresids = LIDtempvals - JRAlidtempvals;
JRAlidmodsids = LIDlidyvals - JRAlidtempvals;
JRAlidmodelsids = LIDlidyvals - JRAlidyvals;
MERdirectresids = LIDtempvals - MERlidtempvals;
MERlidmodsids = LIDlidyvals - MERlidtempvals;
MERlidmodelsids = LIDlidyvals - MERlidyvals;
newERAdirectresids = LIDtempvals - newERAlidtempvals;
newERAlidmodsids = LIDlidyvals - newERAlidtempvals;
newERAlidmodelsids = LIDlidyvals - newERAlidyvals;


% modsids, modelsids a little more complex because of different sized
sets of
% data; also CPC and MERRA2 didn't want to convert directly (less days
than
% lidar set). Problem children dealt with heavily below.

for i = 1:length(LIDnudatas(1,:))
    place = find(CPCnumatches(1,:) == LIDnudatas(1,i));
    if place > 0
        CPCdirectresids(1,place) = LIDtempvals(i) -
CPClidtempvals(place);
    end
end
for i = 1:length(LIDnudatas(1,:))
    place = find(MERRA2numatches(1,:) == LIDnudatas(1,i));
    if place > 0
        MERRA2directresids(1,place) = LIDtempvals(i) -
MERRA2lidtempvals(place);
    end
end

for i = 1:length(LIDlidtvals)
    place = find(CPClidtvals(:) == LIDlidtvals(i));
    if place > 0
        CPClidmodsids(1,place) = LIDlidyvals(i) -
CPClidtempvals(place);
    end
end
for i = 1:length(LIDlidtvals)
    place = find(MERRA2lidtvals(:) == LIDlidtvals(i));
    if place > 0
        MERRA2lidmodsids(1,place) = LIDlidyvals(i) -
MERRA2lidtempvals(place);
    end
end


for i = 1:length(LIDlidtvals)
    place = find(CPClidtvals(:) == LIDlidtvals(i));
```

```
    if place > 0
        CPClidmodelsids(1,place) = LIDlidyvals(i) - CPClidyvals(place);
    end
end
for i = 1:length(LIDlidtvals)
    place = find(MERRA2lidtvals(:) == LIDlidtvals(i));
    if place > 0
        MERRA2lidmodelsids(1,place) = LIDlidyvals(i) -
MERRA2lidyvals(place);
    end
end

%ERAmodsids = LIDyvals - ERAtempvals;
for i = 1:length(ERAtvals)
    place = find(LIDtvals(:) == ERAtvals(i));
    if place > 0
        ERAmodsids(1,place) = LIDyvals(place) - ERAtempvals(i);
    end
end
for i = 1:length(CPCtvals)
    place = find(LIDtvals(:) == CPCtvals(i));
    if place > 0
        CPCmodsids(1,place) = LIDyvals(place) - CPCtempvals(i);
    end
end
for i = 1:length(JRAtvals)
    place = find(LIDtvals(:) == JRAtvals(i));
    if place > 0
        JRAmodsids(1,place) = LIDyvals(place) - JRAtempvals(i);
    end
end
for i = 1:length(MERtvals)
    place = find(LIDtvals(:) == MERtvals(i));
    if place > 0
        MERmodsids(1,place) = LIDyvals(place) - MERtempvals(i);
    end
end
for i = 1:length(MERRA2tvals)
    place = find(LIDtvals(:) == MERRA2tvals(i));
    if place > 0
        MERRA2modsids(1,place) = LIDyvals(place) - MERRA2tempvals(i);
    end
end
for i = 1:length(newERAtvals)
    place = find(LIDtvals(:) == newERAtvals(i));
    if place > 0
        newERAmodsids(1,place) = LIDyvals(place) - newERAtempvals(i);
    end
end


%ERAmodelsids = LIDyvals - ERAyvals;
for i = 1:length(ERAtvals)
    place = find(LIDtvals(:) == ERAtvals(i));
```

```
    if place > 0
        ERAmodsids(1,place) = LIDyvals(place) - ERAyvals(i);
    end
end
for i = 1:length(CPCtvals)
    place = find(LIDtvals(:) == CPCtvals(i));
    if place > 0
        CPCmodsids(1,place) = LIDyvals(place) - CPCyvals(i);
    end
end
for i = 1:length(JRAtvals)
    place = find(LIDtvals(:) == JRAtvals(i));
    if place > 0
        JRAmodsids(1,place) = LIDyvals(place) - JRAyvals(i);
    end
end
for i = 1:length(MERtvals)
    place = find(LIDtvals(:) == MERtvals(i));
    if place > 0
        MERmodsids(1,place) = LIDyvals(place) - MERyvals(i);
    end
end
for i = 1:length(MERRA2tvals)
    place = find(LIDtvals(:) == MERRA2tvals(i));
    if place > 0
        MERRA2modsids(1,place) = LIDyvals(place) - MERRA2yvals(i);
    end
end
for i = 1:length(newERAtvals)
    place = find(LIDtvals(:) == newERAtvals(i));
    if place > 0
        newERAmodsids(1,place) = LIDyvals(place) - newERAyvals(i);
    end
end



%% Just figured out MERRA 2 dates are offset, need to redo all analysis
for that model
% MERRA 2 starts Aug 1st rather than Jan 1st, this is 212 day offset

MERRA245kmT(1,:) = MERRA245kmT(1,:) + 212;

% Rerun all MERRA 2 related analyses

clear MERRA2hardout
[MERRA2tvals,MERRA2dailymodel,eqn,MERRA2nudatas] =
adaptivefit(MERRA245kmT,10);
MERRA2resids =
plotafit(MERRA2tvals,MERRA2dailymodel,eqn,MERRA2nudatas);
a = find(abs(MERRA2resids)>10);
MERRA2hardout(1,:) = MERRA2nudatas(1,a);
MERRA2hardout(2,:) = MERRA2nudatas(2,a);
MERRA2hardout(3,:) = MERRA2resids(a);
mean(MERRA2nudatas(2,:))
```

```
MERRA2yvals = zeros(1,length(MERRA2tvals));
for i = 1:length(MERRA2tvals)
MERRA2yvals(1,i) = eqn(MERRA2dailymodel{i},MERRA2tvals(i));
end
mean(MERRA2yvals)
rightdate = ismember(MERRA2nudatas(1,:),LIDnudatas(1,:));
rightdate(2,:) = rightdate;
MERRA2numatches = MERRA2nudatas.*rightdate;
MERRA2numatches(:,all(MERRA2numatches==0)) = [];
clear rightdate
logics = ismember(MERRA2tvals,LIDnudatas(1,:));
MERRA2lidtvals = MERRA2tvals(logics);
MERRA2lidyvals = MERRA2yvals(logics);
MERRA2lidsids = MERRA2resids(logics);
clear logics
mean(MERRA2lidyvals)
MERRA2tempvals = MERRA2nudatas(2,:); MERRA2lidtempvals =
MERRA2numatches(2,:);
for i = 1:length(LIDlidtvals)
    place = find(MERRA2lidtvals(:) == LIDlidtvals(i));
    if place > 0
        MERRA2lidmodsids(1,place) = LIDlidyvals(i) -
MERRA2lidtempvals(place);
    end
end
for i = 1:length(LIDlidtvals)
    place = find(MERRA2lidtvals(:) == LIDlidtvals(i));
    if place > 0
        MERRA2lidmodelsids(1,place) = LIDlidyvals(i) -
MERRA2lidyvals(place);
    end
end
for i = 1:length(MERRA2tvals)
    place = find(LIDtvals(:) == MERRA2tvals(i));
    if place > 0
        MERRA2modsids(1,place) = LIDyvals(place) - MERRA2tempvals(i);
    end
end
for i = 1:length(MERRA2tvals)
    place = find(LIDtvals(:) == MERRA2tvals(i));
    if place > 0
        MERRA2modsids(1,place) = LIDyvals(place) - MERRA2yvals(i);
    end
end
MERRA2composite = compositeyear(MERRA2nudatas);



% Holy moley I've never found the average temps of actual temps on
lidar
% dates, that might be important

mean(newERAlidtempvals)
mean(MERRA2lidtempvals)
mean(JRAlidtempvals)
```

```
% Should also have nudatas equivalents but made of fitted models
ERAmodel = ([ERAtvals;ERAyvals]);
CPCmodel = ([CPCtvals;CPCyvals]);
MERmodel = ([MERtvals;MERyvals]);
LIDmodel = ([LIDtvals;LIDyvals]);
JRAmodel = ([JRAtvals;JRAyvals]);
MERRA2model = ([MERRA2tvals;MERRA2yvals]);
newERAmodel = ([newERAtvals;newERAyvals]);

%% '96 '00 are leap years, and if I want to do monthly analysis I need
to take that into account

ERAmonthly = monthdivide(ERAnudatas);
CPCmonthly = monthdivide(CPCnudatas);
MERmonthly = monthdivide(MERnudatas);
LIDmonthly = monthdivide(LIDnudatas);
JRAmonthly = monthdivide(JRAnudatas);
MERRA2monthly = monthdivide(MERRA2nudatas);
newERAmonthly = monthdivide(newERAnudatas);
ERAmodelmonthly = monthdivide(ERAmodel);
CPCmodelmonthly = monthdivide(CPCmodel);
MERmodelmonthly = monthdivide(MERmodel);
LIDmodelmonthly = monthdivide(LIDmodel);
JRAmodelmonthly = monthdivide(JRAmodel);
MERRA2modelmonthly = monthdivide(MERRA2model);
newERAmodelmonthly = monthdivide(newERAmodel);

% with current monthdivide code there appear to be a few errors in the
% algorithm that need to be inspected
% every data set has 31 days in year 2 Apr, 27 days in year 11 Feb
% seems to be some kind of rounding error possibly
% I put in some temporary code to test and examine

% Some examination shows deviances at -1.11^-15, so yeah, its
computational
% in nature.
% To fix this, I've simply put a bandaid line of code to repair the
% runningdate variable as it progresses.
% Just rerun the lines of code above and they should work now.

% Compare monthly datas!
% Writing some code that will export results so that I don't have to
% manually find the mean, std dev, and # of data points 1056 times
each!
% This code is in "monthlyport.m'

% run for each monthly set as desired individually
LIDstatsmonthly = monthlyport(LIDmonthly);

%% IT IS TIME TO FIX EVERYTHING I'VE BEEN DOING WRONG UP TO THIS POINT
% I've been using nightly averages for the lidar data even though the
% models are just the data from 600 UT.
```

```
%% Extracting and extrapolating 600 UT lidar data

lidarUT530T(1,:) = lidar45(:,1)';
lidarUT630T(1,:) = lidar45(:,1)';
lidarUT530T(2,:) = lidar45(:,8)';
lidarUT630T(2,:) = lidar45(:,9)';

flag1 = 0;
flag2 = 0;
for i = 1:771
    lidarINST45kmT(1,i) = lidarUT530T(1,i);

    switch lidarUT530T(2,i)
        case 0
            flag1 = 1;
        case NaN
            flag1 = 1;
        otherwise
            value1 = lidarUT530T(2,i);
    end

    switch lidarUT630T(2,i)
        case 0
            flag2 = 1;
        case NaN
            flag2 = 1;
        otherwise
            value2 = lidarUT630T(2,i);
    end

    if flag1 == 1
        if flag2 == 1
            lidarINST45kmT(2,i) = NaN;
        else
            lidarINST45kmT(2,i) = value2;
        end
    else
        if flag2 == 1
            lidarINST45kmT(2,i) = value1;
        else
            lidarINST45kmT(2,i) = mean([value1 value2]);
        end
    end

    flag1 = 0; flag2 = 0;

end

NaNCols = any(isnan(lidarINST45kmT));
lidarINST45kmT = lidarINST45kmT(:,~NaNCols);
clear flag1 flag2 value1 value2 NaNCols
```

```
%% Now redo basic lidar stuff with the instantaneous midnight temps,
including all model comparisons, since those now need to be done
against the instantaneous fit instead of the all-night fit and so on

[lidarINSTtvals,lidarINSTdailymodel,eqn,lidarINSTnudatas] =
adaptivefit(lidarINST45kmT,10);
lidarINSTresids =
plotafit(lidarINSTtvals,lidarINSTdailymodel,eqn,lidarINSTnudatas);
a = find(abs(lidarINSTresids)>10);
lidarINSThardout(1,:) = lidarINSTnudatas(1,a);
lidarINSThardout(2,:) = lidarINSTnudatas(2,a);
lidarINSThardout(3,:) = lidarINSTresids(a);
lidarINSTyvals = zeros(1,length(lidarINSTtvals));
for i = 1:length(lidarINSTtvals)
lidarINSTyvals(1,i) = eqn(lidarINSTdailymodel{i},lidarINSTtvals(i));
end
lidarINSTtempvals = lidarINSTnudatas(2,:);
lidarINSTcomposite = compositeyear(lidarINSTnudatas); % compare with
Josh
lidarINSTmodel = ([lidarINSTtvals;lidarINSTyvals]);
lidarINSTmonthly = monthdivide(lidarINSTnudatas);
lidarINSTmodelmonthly = monthdivide(lidarINSTmodel);

logics = ismember(lidarINSTtvals,lidarINSTnudatas(1,:));
lidarINSTlidtvals = lidarINSTtvals(logics);
lidarINSTlidyvals = lidarINSTyvals(logics);
clear logics

% How well do my previous conclusions hold up, if at all?

mean(lidarINSTtempvals)
mean(lidarINSTyvals)
mean(lidarINSTlidyvals)

% Even warmer! This is good news!
% Export month-by-month for further analysis

lidarINSTstatsmonthly = monthlyport(lidarINSTmonthly);

% And fit model too

lidarINSTstatsmodelmonthly = monthlyport(lidarINSTmodelmonthly);
lidarINSTmodelcomposite = compositeyear(lidarINSTmodel);

% Now the variables that were comparative results with lidar and lidar
fit
% These are: numatches, lidtvals, lidtempvals, lidyvals, lidsids,
% directresids, modsids, lidmodsids, modelsids, and lidmodelsids
% I'm only going to bother with the new 3

% JRA

clear JRAnumatches JRAlidtvals JRAlidtempvals JRAlidyvals JRAlidsids
JRAdirectresids JRAmodsids JRAlidmodsids JRAmodelsids JRAlidmodelsids
```

```
rightdate = ismember(JRAnudatas(1,:),lidarINSTnudatas(1,:));
rightdate(2,:) = rightdate;
JRAnumatches = JRAnudatas.*rightdate;
JRAnumatches(:,all(JRAnumatches==0)) = [];
clear rightdate
logics = ismember(JRAtvals,lidarINSTnudatas(1,:));
JRAlidtvals = JRAtvals(logics);
JRAlidyvals = JRAyvals(logics);
JRAlidsids = JRAresids(logics);
clear logics
JRAlidtempvals = JRAnumatches(2,:);
for i = 1:length(lidarINSTnudatas(1,:))
    place = find(JRAnumatches(1,:) == lidarINSTnudatas(1,i));
    if place > 0
        JRAdirectresids(1,place) = lidarINSTtempvals(i) -
JRAlidtempvals(place);
    end
end
for i = 1:length(lidarINSTlidtvals)
    place = find(JRAlidtvals(:) == lidarINSTlidtvals(i));
    if place > 0
        JRAlidmodsids(1,place) = lidarINSTlidyvals(i) -
JRAlidtempvals(place);
    end
end
for i = 1:length(lidarINSTlidtvals)
    place = find(JRAlidtvals(:) == lidarINSTlidtvals(i));
    if place > 0
        JRAlidmodelsids(1,place) = lidarINSTlidyvals(i) -
JRAlidyvals(place);
    end
end
for i = 1:length(JRAtvals)
    place = find(lidarINSTtvals(:) == JRAtvals(i));
    if place > 0
        JRAmodsids(1,place) = lidarINSTyvals(place) - JRAtempvals(i);
    end
end
for i = 1:length(JRAtvals)
    place = find(lidarINSTtvals(:) == JRAtvals(i));
    if place > 0
        JRAmodelsids(1,place) = lidarINSTyvals(place) - JRAyvals(i);
    end
end

% ERA-20C

clear newERAnumatches newERAlidtvals newERAlidtempvals newERAlidyvals
newERAlidsids newERAdirectresids newERAmodsids newERAlidmodsids
newERAmodelsids newERAlidmodelsids

rightdate = ismember(newERAnudatas(1,:),lidarINSTnudatas(1,:));
rightdate(2,:) = rightdate;
```

```
newERAnumatches = newERAnudatas.*rightdate;
newERAnumatches(:,all(newERAnumatches==0)) = [];
clear rightdate
logics = ismember(newERAtvals,lidarINSTnudatas(1,:));
newERAlidtvals = newERAtvals(logics);
newERAlidyvals = newERAyvals(logics);
newERAlidsids = newERAresids(logics);
clear logics
newERAlidtempvals = newERAnumatches(2,:);
for i = 1:length(lidarINSTnudatas(1,:))
    place = find(newERAnumatches(1,:) == lidarINSTnudatas(1,i));
    if place > 0
        newERAdirectresids(1,place) = lidarINSTtempvals(i) -
newERAlidtempvals(place);
    end
end
for i = 1:length(lidarINSTlidtvals)
    place = find(newERAlidtvals(:) == lidarINSTlidtvals(i));
    if place > 0
        newERAlidmodsids(1,place) = lidarINSTlidyvals(i) -
newERAlidtempvals(place);
    end
end
for i = 1:length(lidarINSTlidtvals)
    place = find(newERAlidtvals(:) == lidarINSTlidtvals(i));
    if place > 0
        newERAlidmodelsids(1,place) = lidarINSTlidyvals(i) -
newERAlidyvals(place);
    end
end
for i = 1:length(newERAtvals)
    place = find(lidarINSTtvals(:) == newERAtvals(i));
    if place > 0
        newERAmodsids(1,place) = lidarINSTyvals(place) -
newERAtempvals(i);
    end
end
for i = 1:length(newERAtvals)
    place = find(lidarINSTtvals(:) == newERAtvals(i));
    if place > 0
        newERAmodelsids(1,place) = lidarINSTyvals(place) -
newERAyvals(i);
    end
end

% MERRA 2

clear MERRA2numatches MERRA2lidtvals MERRA2lidtempvals MERRA2lidyvals
MERRA2lidsids MERRA2directresids MERRA2modsids MERRA2lidmodsids
MERRA2modelsids MERRA2lidmodelsids

rightdate = ismember(MERRA2nudatas(1,:),lidarINSTnudatas(1,:));
rightdate(2,:) = rightdate;
MERRA2numatches = MERRA2nudatas.*rightdate;
```

```
MERRA2numatches(:,all(MERRA2numatches==0)) = [];
clear rightdate
logics = ismember(MERRA2tvals,lidarINSTnudatas(1,:));
MERRA2lidtvals = MERRA2tvals(logics);
MERRA2lidyvals = MERRA2yvals(logics);
MERRA2lidsids = MERRA2resids(logics);
clear logics
MERRA2lidtempvals = MERRA2numatches(2,:);
for i = 1:length(lidarINSTnudatas(1,:))
    place = find(MERRA2numatches(1,:) == lidarINSTnudatas(1,i));
    if place > 0
        MERRA2directresids(1,place) = lidarINSTtempvals(i) -
MERRA2lidtempvals(place);
    end
end
for i = 1:length(lidarINSTlidtvals)
    place = find(MERRA2lidtvals(:) == lidarINSTlidtvals(i));
    if place > 0
        MERRA2lidmodsids(1,place) = lidarINSTlidyvals(i) -
MERRA2lidtempvals(place);
    end
end
for i = 1:length(lidarINSTlidtvals)
    place = find(MERRA2lidtvals(:) == lidarINSTlidtvals(i));
    if place > 0
        MERRA2lidmodelsids(1,place) = lidarINSTlidyvals(i) -
MERRA2lidyvals(place);
    end
end
for i = 1:length(MERRA2tvals)
    place = find(lidarINSTtvals(:) == MERRA2tvals(i));
    if place > 0
        MERRA2modsids(1,place) = lidarINSTyvals(place) -
MERRA2tempvals(i);
    end
end
for i = 1:length(MERRA2tvals)
    place = find(lidarINSTtvals(:) == MERRA2tvals(i));
    if place > 0
        MERRA2modelsids(1,place) = lidarINSTyvals(place) -
MERRA2yvals(i);
    end
end

newERAmodelcomposite = compositeyear(newERAmodel);
MERRA2modelcomposite = compositeyear(MERRA2model);
JRAmodelcomposite = compositeyear(JRAmodel);

% Get new set-length averages (tempvals and yvals haven't changed)

mean(JRAlidyvals)
mean(JRAlidtempvals)

mean(newERAlidyvals)
```

```
mean(newERAlidtempvals)

mean(MERRA2lidyvals)
mean(MERRA2lidtempvals)

%% Let's give the fit a try with EVEN MORE PARAMETERS

% 12 parameter
[try12tvals,try12dailymodel,eqn,try12nudatas] =
adaptivefit(lidarINST45kmT,12);
try12resids = plotafit(try12tvals,try12dailymodel,eqn,try12nudatas);
a = find(abs(try12resids)>10);
try12hardout(1,:) = try12nudatas(1,a);
try12hardout(2,:) = try12nudatas(2,a);
try12hardout(3,:) = try12resids(a);
try12yvals = zeros(1,length(try12tvals));
for i = 1:length(try12tvals)
try12yvals(1,i) = eqn(try12dailymodel{i},try12tvals(i));
end
try12tempvals = try12nudatas(2,:);
try12composite = compositeyear(try12nudatas);
try12model = ([try12tvals;try12yvals]);
try12monthly = monthdivide(try12nudatas);
try12modelmonthly = monthdivide(try12model);
logics = ismember(try12tvals,lidarINSTnudatas(1,:));
try12lidtvals = try12tvals(logics);
try12lidyvals = try12yvals(logics);
clear logics
mean(try12tempvals)
mean(try12yvals)
mean(try12lidyvals)
try12modelcomposite = compositeyear(try12model);
try12hardoutmonthly = monthdivide(try12hardout(1:2,:));
monthlyport(try12hardoutmonthly);

% 14 parameter

[try14tvals,try14dailymodel,eqn,try14nudatas] =
adaptivefit(lidarINST45kmT,14);
try14resids = plotafit(try14tvals,try14dailymodel,eqn,try14nudatas);
a = find(abs(try14resids)>10);
try14hardout(1,:) = try14nudatas(1,a);
try14hardout(2,:) = try14nudatas(2,a);
try14hardout(3,:) = try14resids(a);
try14yvals = zeros(1,length(try14tvals));
for i = 1:length(try14tvals)
try14yvals(1,i) = eqn(try14dailymodel{i},try14tvals(i));
end
try14tempvals = try14nudatas(2,:);
try14composite = compositeyear(try14nudatas);
try14model = ([try14tvals;try14yvals]);
try14monthly = monthdivide(try14nudatas);
try14modelmonthly = monthdivide(try14model);
logics = ismember(try14tvals,lidarINSTnudatas(1,:));
```

```
try14lidtvals = try14tvals(logics);
try14lidyvals = try14yvals(logics);
clear logics
mean(try14tempvals)
mean(try14yvals)
mean(try14lidyvals)
try14modelcomposite = compositeyear(try14model);
try14hardoutmonthly = monthdivide(try14hardout(1:2,:));
monthlyport(try14hardoutmonthly);

% The results of adding more parameters seem to indicate that the
problems
% in the winter with the fit are likely due to the data itself,
probably
% attributable to geophysical variability. There is no noticeable
% improvement in the model gained by adding more parameters than 10.

%% Examine number of 10 K outliers per month by using the hardout
variable along with the monthdivide and monthlyport functions

lidarINSThardoutmonthly = monthdivide(lidarINSThardout(1:2,:));
[~] = monthlyport(lidarINSThardoutmonthly);
newERAhardoutmonthly = monthdivide(newERAhardout(1:2,:));
[~] = monthlyport(newERAhardoutmonthly);
MERRA2hardoutmonthly = monthdivide(MERRA2hardout(1:2,:));
[~] = monthlyport(MERRA2hardoutmonthly);
JRAhardoutmonthly = monthdivide(JRAhardout(1:2,:));
[~] = monthlyport(JRAhardoutmonthly);

% Did the above with all possible try12 and try14 iterations as well to
see
% how they changed, VERY LITTLE CHANGE

% In this process I discovered with some cross-checking that my
monthdivide
% and monthlyport functions STILL aren't working as intended. I have
now
% fixed them again and cross-checking shows more sensible data now.
This
% also got rid of the problems with rounding errors, so the bandaid
code
% has been removed. I also updated the compositeyear code to use a
similar
% method for scanning through the data.

% I should look at the differences in averages on just lidar days in
the
% models and fits too

lidarINSTlidmodmonth =
monthdivide([lidarINSTlidtvals;lidarINSTlidyvals]);
lidarINSTstatslidmodmonth = monthlyport(lidarINSTlidmodmonth);
newERAlidmonthly = monthdivide(newERAnumatches);
newERAstatslidmonthly = monthlyport(newERAlidmonthly);
```

```
newERAlidmodmonth = monthdivide([newERAlidtvals;newERAlidyvals]);
newERAstatslidmodmonth = monthlyport(newERAlidmodmonth);
MERRA2lidmonthly = monthdivide(MERRA2numatches);
MERRA2statslidmonthly = monthlyport(MERRA2lidmonthly);
MERRA2lidmodmonth = monthdivide([MERRA2lidtvals;MERRA2lidyvals]);
MERRA2statslidmodmonth = monthlyport(MERRA2lidmodmonth);
JRAlidmonthly = monthdivide(JRAnumatches);
JRAstatslidmonthly = monthlyport(JRAlidmonthly);
JRAlidmodmonth = monthdivide([JRAlidtvals;JRAlidyvals]);
JRAstatslidmodmonth = monthlyport(JRAlidmodmonth);


% Find average amplitudes phases of the model components (annual,
semiannual etc)
extract_cell = @(C, k) cellfun(@(c)c(k), C) ; % to extract components
from cell arrays
% Output mean paramater values all to window
for i = 1:10
    i
    mean(extract_cell(lidarINSTdailymodel,i))
    mean(extract_cell(newERAdailymodel,i))
    mean(extract_cell(MERRA2dailymodel,i))
    mean(extract_cell(JRAdailymodel,i))
end
% Do the same for std deviations
for i = 1:10
    i
    std(extract_cell(lidarINSTdailymodel,i))
    std(extract_cell(newERAdailymodel,i))
    std(extract_cell(MERRA2dailymodel,i))
    std(extract_cell(JRAdailymodel,i))
end
clear i



% Do some simple linear fits to examine total trend and see how they
match
% with the linear parameter in the fit method
polyfit(lidarINSTlidtvals,lidarINSTtempvals,1) % output is trend/yr and
intercept
polyfit(newERAtvals,newERAtempvals,1)
polyfit(MERRA2tvals,MERRA2tempvals,1)
polyfit(JRAtvals,JRAtempvals,1)

% Get some specific data for figure
jun2001JRAdates = JRAmonthly{8,6,1};
jun2001JRAtemps = JRAmonthly{8,6,2};
jun2001JRAmoddates = JRAmodelmonthly{8,6,1};
jun2001JRAmodtemps = JRAmodelmonthly{8,6,2};
jun2001JRAliddates = JRAlidmonthly{8,6,1};
jun2001JRAlidtemps = JRAlidmonthly{8,6,2};

% Plot figures
% OLDTHING
```

```
that = 0:1/365:12;
predicted = modelb{1}.Formula.ModelFun;
yhat = predicted(modelb{1}.Coefficients.Estimate,that);
figure
plot(t{1},y{1},'g.',that,yhat,'k','LineWidth',1.5,'MarkerSize',14);
xlim([0 12]);
pos = get(gcf, 'Position');
set(gcf, 'Position', [pos(1) pos(2) 14.5*100 5*100]);
xlabel('Year','FontSize',15);
ylabel('Temperature (K)', 'FontSize',15);
title('Figure 2 - ALO-USU Data and Regression Line','FontSize',15);
set(gca, 'FontSize',15,'LineWidth',1.5);
legend('Avg. Nightly Temperature','Regression
Fit','Location','SouthWest');
set(gca,'XTickLabel',{'1993','1995','1997','1999','2001','2003',[]});
set(gca,'YTickLabel',{[],'230','240','250','260','270','280',[]});

% NEWTHINGS
[~,a] =
plotafit(lidarINSTtvals,lidarINSTdailymodel,eqn,lidarINSTnudatas);
set(gca,'YTickLabel',{[],'230','240','250','260','270','280',[]});
set(gca,'XTickLabel',{'1993','1995','1997','1999','2001','2003','2005'}
);
title('Rayleigh-scatter Lidar (ALO-USU) Temperatures at 45 km');
hold on;
h =
plot(lidarINSThardout(1,:),lidarINSThardout(2,:),'b*','MarkerSize',8);
q(1) = a(1); q(2) = h; q(3) = a(2);
legend(q,'Model Temperature','> 10 K Outlier','Adaptive
Fit','Location','SouthWest');

[~,a] = plotafit(newERAtvals,newERAdailymodel,eqn,newERAnudatas);
set(gca,'YTickLabel',{[],'220','230','240','250','260','270','280',[]})
;
set(gca,'XTickLabel',{'1993','1995','1997','1999','2001','2003','2005'}
);
title('ERA-20C (ECMWF) Temperatures at 45 km');
hold on;
h = plot(newERAhardout(1,:),newERAhardout(2,:),'b*','MarkerSize',8);
q(1) = a(1); q(2) = h; q(3) = a(2);
legend(q,'Model Temperature','> 10 K Outlier','Adaptive
Fit','Location','SouthWest');

[~,a] = plotafit(MERRA2tvals,MERRA2dailymodel,eqn,MERRA2nudatas);
set(gca,'YTickLabel',{[],'220','230','240','250','260','270','280',[]})
;
set(gca,'XTickLabel',{'1993','1995','1997','1999','2001','2003','2005'}
);
title('MERRA-2 (NASA) Temperatures at 45 km');
hold on;
h = plot(MERRA2hardout(1,:),MERRA2hardout(2,:),'b*','MarkerSize',8);
q(1) = a(1); q(2) = h; q(3) = a(2);
legend(q,'Model Temperature','> 10 K Outlier','Adaptive
Fit','Location','SouthWest');
```

```
[~,a] = plotafit(JRAtvals,JRAdailymodel,eqn,JRAnudatas);
set(gca,'YTickLabel',{[],'220','240','260','280',[]});
set(gca,'XTickLabel',{'1993','1995','1997','1999','2001','2003','2005'}
);
title('JRA-55 (JMA) Temperatures at 45 km');
hold on;
h = plot(ERAhardout(1,:),ERAhardout(2,:),'b*','MarkerSize',8);
q(1) = a(1); q(2) = h; q(3) = a(2);
legend(q,'Model Temperature','> 10 K Outlier','Adaptive
Fit','Location','SouthWest');

% WILMAAAAA
figure
lidarINSTpd = plotPDFtLC(lidarINSTresids,'ALO-USU');
xlim([-30 30])
ylim([0 .18])

figure
newERApd = plotPDFtLC(newERAresids,'ERA-20C');
xlim([-30 30])
ylim([0 .18])

figure
MERRA2pd = plotPDFtLC(MERRA2resids,'MERRA-2');
xlim([-30 30])
ylim([0 .18])

figure
JRApd = plotPDFtLC(JRAresids,'JRA-55');
xlim([-30 30])
ylim([0 .18])

% PORKBARREL
[try12tvals,try12dailymodel,eqn12,try12nudatas] =
adaptivefit(MERRA245kmT,12); %generate data
[try14tvals,try14dailymodel,eqn14,try14nudatas] =
adaptivefit(MERRA245kmT,14); %generate data

[~,a] = plotafit(try12tvals,try12dailymodel,eqn12,try12nudatas);
set(gca,'YTickLabel',{[],'220','230','240','250','260','270','280',[]})
;
set(gca,'XTickLabel',{'1993','1995','1997','1999','2001','2003','2005'}
);
title('MERRA-2 (NASA) Temperatures at 45 km');
hold on;
h = plot(MERRA2hardout(1,:),MERRA2hardout(2,:),'b*','MarkerSize',8);
q(1) = a(1); q(2) = h; q(3) = a(2);
legend(q,'Model Temperature','> 10 K Outlier','Adaptive Fit (12-
Parameter)','Location','SouthWest');

[~,a] = plotafit(try14tvals,try14dailymodel,eqn14,try14nudatas);
set(gca,'YTickLabel',{[],'220','230','240','250','260','270','280',[]})
;
```

```
set(gca,'XTickLabel',{'1993','1995','1997','1999','2001','2003','2005'}
);
title('MERRA-2 (NASA) Temperatures at 45 km');
hold on;
h = plot(MERRA2hardout(1,:),MERRA2hardout(2,:),'b*','MarkerSize',8);
q(1) = a(1); q(2) = h; q(3) = a(2);
legend(q,'Model Temperature','> 10 K Outlier','Adaptive Fit (14-
Parameter)','Location','SouthWest');

% POTATOHEAD see which type is better
%type 1
figure
bar(twelve,2,'LineWidth',1)
title('Total Outliers by Month (12-Parameter)')
xlim([0 13]);
pos = get(gcf, 'Position');
set(gcf, 'Position', [pos(1) pos(2) 10*100 4*100]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'})
legend('ALO-USU','ERA-20C','MERRA-2','JRA','Location','north')

figure
bar(fourt,2,'LineWidth',1)
title('Total Outliers by Month (14-Parameter)')
xlim([0 13]);
pos = get(gcf, 'Position');
set(gcf, 'Position', [pos(1) pos(2) 10*100 4*100]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'})
legend('ALO-USU','ERA-20C','MERRA-2','JRA','Location','north')

%type 2
figure
bar(outliercomplidar,1.6,'LineWidth',1)
title('ALO-USU Total Outliers by Month')
xlim([0 13]);
pos = get(gcf, 'Position');
set(gcf, 'Position', [pos(1) pos(2) 10*100 4*100]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'})
legend('10-Parameter','12-Parameter','14-Parameter','Location','north')
ylabel('# of Outliers')

figure
bar(outliercompera,1.6,'LineWidth',1)
title('ERA-20C Total Outliers by Month')
xlim([0 13]);
pos = get(gcf, 'Position');
set(gcf, 'Position', [pos(1) pos(2) 10*100 4*100]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'})
legend('10-Parameter','12-Parameter','14-Parameter','Location','north')
ylabel('# of Outliers')
```

```
figure
bar(outliercompmerra,1.6,'LineWidth',1)
title('MERRA-2 Total Outliers by Month')
xlim([0 13]);
pos = get(gcf, 'Position');
set(gcf, 'Position', [pos(1) pos(2) 10*100 4*100]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'})
legend('10-Parameter','12-Parameter','14-Parameter','Location','north')
ylabel('# of Outliers')

figure
bar(outliercompjra,1.6,'LineWidth',1)
title('JRA-55 Total Outliers by Month')
xlim([0 13]);
pos = get(gcf, 'Position');
set(gcf, 'Position', [pos(1) pos(2) 10*100 4*100]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'})
legend('10-Parameter','12-Parameter','14-Parameter','Location','north')
ylabel('# of Outliers')

% PUFFERFISH
figure
h1 = plot(newERAcomposite(1,:),lidarINSTcomposite(2,:),'Color',[.224
.898 .235],'LineWidth',2); hold on;
plot(newERAcomposite(1,:),lidarINSTcomposite(2,:)-
lidarINSTcomposite(3,:),'--','Color',[.224 .898 .235]);
plot(newERAcomposite(1,:),lidarINSTcomposite(2,:)+lidarINSTcomposite(3,
:),'--','Color',[.224 .898 .235]);
h2 = plot(newERAcomposite(1,:),newERAcomposite(2,:),'k','LineWidth',2);
plot(newERAcomposite(1,:),newERAcomposite(2,:)-newERAcomposite(3,:),'--
','Color','k');
plot(newERAcomposite(1,:),newERAcomposite(2,:)+newERAcomposite(3,:),'--
','Color','k');
h3 = plot(newERAcomposite(1,:),MERRA2composite(2,:),'b','LineWidth',2);
plot(newERAcomposite(1,:),MERRA2composite(2,:)-MERRA2composite(3,:),'--
','Color','b');
plot(newERAcomposite(1,:),MERRA2composite(2,:)+MERRA2composite(3,:),'--
','Color','b');
h4 = plot(newERAcomposite(1,:),JRAcomposite(2,:),'r','LineWidth',2);
plot(newERAcomposite(1,:),JRAcomposite(2,:)-JRAcomposite(3,:),'--
','Color','r');
plot(newERAcomposite(1,:),JRAcomposite(2,:)+JRAcomposite(3,:),'--
','Color','r');
title('Composite years (REAL datasets), 3\sigma error bars')
ylabel('Temperature (K)')
xlabel('Month of Year')
xlim([-0.5 364.5]);
set(gca,'XTick',[-0.5,31,59,90,120,151,181,212,243,273,304,334]);
set(gca,'XTickLabel',{'         Jan','         Feb','         Mar','
Apr','         May','         Jun','         Jul','         Aug','
Sep','         Oct','         Nov','         Dec'});
legend([h1,h2,h3,h4],'ALO-USU','ERA-20C','MERRA-2','JRA')
```

```
figure
h1 =
plot(newERAcomposite(1,:),lidarINSTmodelcomposite(2,:),'Color',[.224
.898 .235],'LineWidth',2); hold on;
plot(newERAcomposite(1,:),lidarINSTmodelcomposite(2,:)-
lidarINSTmodelcomposite(3,:),'--','Color',[.224 .898 .235]);
plot(newERAcomposite(1,:),lidarINSTmodelcomposite(2,:)+lidarINSTmodelco
mposite(3,:),'--','Color',[.224 .898 .235]);
h2 =
plot(newERAcomposite(1,:),newERAmodelcomposite(2,:),'k','LineWidth',2);
plot(newERAcomposite(1,:),newERAmodelcomposite(2,:)-
newERAmodelcomposite(3,:),'--','Color','k');
plot(newERAcomposite(1,:),newERAmodelcomposite(2,:)+newERAmodelcomposit
e(3,:),'--','Color','k');
h3 =
plot(newERAcomposite(1,:),MERRA2modelcomposite(2,:),'b','LineWidth',2);
plot(newERAcomposite(1,:),MERRA2modelcomposite(2,:)-
MERRA2modelcomposite(3,:),'--','Color','b');
plot(newERAcomposite(1,:),MERRA2modelcomposite(2,:)+MERRA2modelcomposit
e(3,:),'--','Color','b');
h4 =
plot(newERAcomposite(1,:),JRAmodelcomposite(2,:),'r','LineWidth',2);
plot(newERAcomposite(1,:),JRAmodelcomposite(2,:)-
JRAmodelcomposite(3,:),'--','Color','r');
plot(newERAcomposite(1,:),JRAmodelcomposite(2,:)+JRAmodelcomposite(3,:)
,'--','Color','r');
title('Composite years (FIT datasets), 3\sigma error bars')
ylabel('Temperature (K)')
xlabel('Month of Year')
xlim([-0.5 364.5]);
set(gca,'XTick',[-0.5,31,59,90,120,151,181,212,243,273,304,334]);
set(gca,'XTickLabel',{'         Jan','         Feb','         Mar','
Apr','         May','         Jun','         Jul','         Aug','
Sep','         Oct','         Nov','         Dec'});
legend([h1,h2,h3,h4],'ALO-USU','ERA-20C','MERRA-2','JRA')



% GLAZED
figure
trakday = round(lidarINSTtvals * 365 + 1);
yr = 0;
for i = 1:length(lidarINSTtvals)
    j = trakday(i) - 365 * yr;
    if j == 366
        j = 1;
        yr = yr + 1;
    end
    stackeddata(yr+1,j) = newERAmodsids(i);
end
stackeddata(1,1:242) = NaN; stackeddata(12,226:365) = NaN;
h1 = imagesc(stackeddata);
colormap(jet);
```

```
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
set(h1,'AlphaData',~isnan(stackeddata));
title('LID-FIT - ERA-REAL Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,32,60,91,121,152,182,213,244,274,305,335]);
set(gca,'XTickLabel',{'         Jan','         Feb','         Mar','
Apr','         May','         Jun','         Jul','         Aug','
Sep','         Oct','         Nov','         Dec'});
ylabel(h2,'Deviation (K)');
clear trakday stackeddata

figure
trakday = round(lidarINSTtvals * 365 + 1);
yr = 0;
for i = 1:length(lidarINSTtvals)
    j = trakday(i) - 365 * yr;
    if j == 366
        j = 1;
        yr = yr + 1;
    end
    stackeddata(yr+1,j) = MERRA2modsids(i);
end
stackeddata(1,1:242) = NaN; stackeddata(12,226:365) = NaN;
h1 = imagesc(stackeddata);
colormap(jet);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
set(h1,'AlphaData',~isnan(stackeddata));
title('LID-FIT - MER-REAL Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,32,60,91,121,152,182,213,244,274,305,335]);
set(gca,'XTickLabel',{'         Jan','         Feb','         Mar','
Apr','         May','         Jun','         Jul','         Aug','
Sep','         Oct','         Nov','         Dec'});
ylabel(h2,'Deviation (K)');
clear trakday stackeddata

figure
trakday = round(lidarINSTtvals * 365 + 1);
yr = 0;
for i = 1:length(lidarINSTtvals)
    j = trakday(i) - 365 * yr;
    if j == 366
        j = 1;
        yr = yr + 1;
    end
    stackeddata(yr+1,j) = JRAmodsids(i);
```

```
end
stackeddata(1,1:242) = NaN; stackeddata(12,226:365) = NaN;
h1 = imagesc(stackeddata);
colormap(jet);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
set(h1,'AlphaData',~isnan(stackeddata));
title('LID-FIT - JRA-REAL Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,32,60,91,121,152,182,213,244,274,305,335]);
set(gca,'XTickLabel',{'         Jan','         Feb','         Mar','
Apr','         May','         Jun','         Jul','         Aug','
Sep','         Oct','         Nov','         Dec'});
ylabel(h2,'Deviation (K)');
clear trakday stackeddata


% POWDERED
figure
trakday = round(lidarINSTtvals * 365 + 1);
yr = 0;
for i = 1:length(lidarINSTtvals)
    j = trakday(i) - 365 * yr;
    if j == 366
        j = 1;
        yr = yr + 1;
    end
    stackeddata(yr+1,j) = newERAmodelsids(i);
end
stackeddata(1,1:242) = NaN; stackeddata(12,226:365) = NaN;
h1 = imagesc(stackeddata);
colormap(jet);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
set(h1,'AlphaData',~isnan(stackeddata));
title('LID-FIT - ERA-FIT Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,32,60,91,121,152,182,213,244,274,305,335]);
set(gca,'XTickLabel',{'         Jan','         Feb','         Mar','
Apr','         May','         Jun','         Jul','         Aug','
Sep','         Oct','         Nov','         Dec'});
ylabel(h2,'Deviation (K)');
clear trakday stackeddata

figure
trakday = round(lidarINSTtvals * 365 + 1);
yr = 0;
for i = 1:length(lidarINSTtvals)
```

```
    j = trakday(i) - 365 * yr;
    if j == 366
        j = 1;
        yr = yr + 1;
    end
    stackeddata(yr+1,j) = MERRA2modelsids(i);
end
stackeddata(1,1:242) = NaN; stackeddata(12,226:365) = NaN;
h1 = imagesc(stackeddata);
colormap(jet);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
set(h1,'AlphaData',~isnan(stackeddata));
title('LID-FIT - MER-FIT Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,32,60,91,121,152,182,213,244,274,305,335]);
set(gca,'XTickLabel',{'         Jan','         Feb','         Mar','
Apr','         May','         Jun','         Jul','         Aug','
Sep','         Oct','         Nov','         Dec'});
ylabel(h2,'Deviation (K)');
clear trakday stackeddata

figure
trakday = round(lidarINSTtvals * 365 + 1);
yr = 0;
for i = 1:length(lidarINSTtvals)
    j = trakday(i) - 365 * yr;
    if j == 366
        j = 1;
        yr = yr + 1;
    end
    stackeddata(yr+1,j) = JRAmodelsids(i);
end
stackeddata(1,1:242) = NaN; stackeddata(12,226:365) = NaN;
h1 = imagesc(stackeddata);
colormap(jet);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
set(h1,'AlphaData',~isnan(stackeddata));
title('LID-FIT - JRA-FIT Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,32,60,91,121,152,182,213,244,274,305,335]);
set(gca,'XTickLabel',{'         Jan','         Feb','         Mar','
Apr','         May','         Jun','         Jul','         Aug','
Sep','         Oct','         Nov','         Dec'});
ylabel(h2,'Deviation (K)');
clear trakday stackeddata
```

```
% OMAZING
throwshade(newERAdiff,newERAp);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
title('LID-REAL - ERA-REAL Monthly Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Deviation (K)');

throwshade(MERRA2diff,MERRA2p);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
title('LID-REAL - MER-REAL Monthly Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Deviation (K)');

throwshade(JRAdiff,JRAp);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
title('LID-REAL - JRA-REAL Monthly Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Deviation (K)');

% ACKMAZING
throwshade(newERAliddiff,newERAplid);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
title('LID-RLID - ERA-RLID Monthly Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Deviation (K)');
```

```
throwshade(MERRA2liddiff,MERRA2plid);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
title('LID-RLID - MER-RLID Monthly Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Deviation (K)');

throwshade(JRAliddiff,JRAplid);
h2 = colorbar;
caxis([-10 15]);
h2.TickLabels = {'<-10','-5','0','5','10','>15'};
title('LID-RLID - JRA-RLID Monthly Temperature Difference')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Deviation (K)');

% BURRITO
figure
h1 = imagesc(lidarINSTstatsmonthly(:,:,1));
colormap(jet);
axis equal
xlim([0.5 12.5])
set(h1,'AlphaData',~isnan(lidarINSTstatsmonthly(:,:,1)));
h2 = colorbar;
caxis([245 285]);
title('LID-REAL Monthly Means')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Average Temperature (K)');

figure
h1 = imagesc(newERAstatsmonthly(:,:,1));
colormap(jet);
axis equal
xlim([0.5 12.5])
set(h1,'AlphaData',~isnan(newERAstatsmonthly(:,:,1)));
h2 = colorbar;
caxis([245 285]);
title('ERA-REAL Monthly Means')
ylabel('Year')
```

```
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Average Temperature (K)');

figure
h1 = imagesc(MERRA2statsmonthly(:,:,1));
colormap(jet);
axis equal
xlim([0.5 12.5])
set(h1,'AlphaData',~isnan(MERRA2statsmonthly(:,:,1)));
h2 = colorbar;
caxis([245 285]);
title('MER-REAL Monthly Means')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Average Temperature (K)');

figure
h1 = imagesc(JRAstatsmonthly(:,:,1));
colormap(jet);
axis equal
xlim([0.5 12.5])
set(h1,'AlphaData',~isnan(JRAstatsmonthly(:,:,1)));
h2 = colorbar;
caxis([245 285]);
title('JRA-REAL Monthly Means')
ylabel('Year')
xlabel('Month')
set(gca,'YTickLabel',{'1994','1996','1998','2000','2002','2004'});
set(gca,'XTick',[1,2,3,4,5,6,7,8,9,10,11,12]);
set(gca,'XTickLabel',{'Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','
Sep','Oct','Nov','Dec'});
ylabel(h2,'Average Temperature (K)');

% ROSETTA
figure
h1 = plot(jun2001JRAdates,jun2001JRAtemps,'r.','MarkerSize',20);
hold on
plot(jun2001JRAmoddates,jun2001JRAmodtemps,'k-','LineWidth',2);
plot(jun2001JRAliddates,jun2001JRAlidtemps,'.','Color',[.224 .898
.235],'MarkerSize',20);
xlim([7.4192 7.4986])
set(gca,'XTick',[7.4192,7.4324,7.4456,7.4588,7.472,7.4852]);
set(gca,'XTickLabel',{'1','6','11','16','21','26'})
xlabel('Day of June 2001')
ylabel('Temperature (K)')
title('Guide to Monthly Means (JRA-55 Example)')
```